# CURSO TALLER: “R APLICADO A LOS PROYECTOS DE INVESTIGACIÓN”

# PRESENTACIÓN

Curso dirigido a personal de salud que se encuentre laborando en el Instituto Nacional de Salud del Niño San Borja (INSNSB), que cuente con entrenamiento básico en estadística aplicada a ciencias de la salud, estadística o ciencia de datos. Su propósito es capacitar y brindar herramientas en análisis bioestadístico de datos para preguntas de investigación explicativas. Se revisarán conceptos relacionados al manejo de datos, análisis descriptivo e inferencial y modelado de regresión. El entrenamiento se realizará de manera virtual empleando la plataforma Zoom. Al finalizar el entrenamiento, el alumno estará en la capacidad de llevar a cabo un análisis exploratorio explicativo (factores asociados) o confirmatorio (efectividad clínica comparativa o evaluación de impacto) para desenlaces numéricos continuos, y binarios mediante un enfoque de modelado de regresión.

# OBJETIVOS

**Objetivo General:**

Desarrollar capacidades de análisis estadístico de datos para investigación en profesionales de la salud del Instituto *nacional de Salud del Niño San Borja.*

**Objetivos Específicos:**

Al finalizar las unidades, el personal de salud participante del curso será capaz de:

* Usar R y RStudio para el manejo y limpieza de base de datos de estudios de investigación en salud.
* Realizar un análisis descriptivo e inferencial con un paquete estadístico de datos para la investigación en salud.
* Responder preguntas explicativas mediante modelado de regresión para desenlaces numéricos y categóricos binarios.
* Generar tablas y reporte reproducibles de estos mediante el uso de markdown en R/RStudio.

# ESTRUCTURA

* 1. **PERFIL DE INGRESO**

Personal de salud que se encuentre laborando en el Instituto Nacional de Salud del Niño San Borja (INSNSB), que cuente con entrenamiento básico en estadística aplicada a ciencias de la salud, estadística o ciencia de datos, y que deseen utilizar R/RStudio para análisis de datos aplicados a investigación en salud.

* 1. **PERFIL DE EGRESO**

Personal de salud del Instituto Nacional de Salud del Niño con competencias para realizar análisis estadístico de datos para investigación en salud a nivel descriptivo, inferencial y de modelado explicativo en R y RStudio, así como generar los reportes reproducibles en estos.

* 1. **PROGRAMACIÓN DEL CURSO**

|  |  |  |
| --- | --- | --- |
| **SESIÓN** | **FECHA Y HORA** | **CONTENIDOS** |
| **MÓDULO 1: MANEJO Y PRE-PROCESAMIENTO DE DATOS CON R** | | |
| **1** | 03/10/2022 | Presentación de curso y recursos informáticos. Verificación de instalación de R, Rstudio y Rtools. Recursos informáticos del curso y recursos de aprendizaje en línea. Introducción a R y Rstudio. Operaciones usuales. Instalación de paquetes. Administración de entorno. Configuración de proyecto y carpetas. Instalación de paquetes. Operaciones básicas: aritméticas, asignar (<-) y seleccionar ($). Operar booelanos: | e & (atajos con ASCII). |
| 14:00 – 16:00 |
| **2** | 05/10/2022 | Estrategias gráficas de inspección/exploración de datos con ggplot(). Operador asignar y pipe (%>%). Manejo de datos con dplyr 1, funciones select(), filter(), y arrange(). |
| 14:00 – 16:00 |
| **3** | 07/10/2022 | Importar datos de Excel, Stata, SPSS, CSV. Manejo de datos con dplyr 2: mutate(), rename() y recode() . Uso de helpers y otros atajos.  Inspección Inicial de Datos vs. Análisis Exploratorio de Datos.  Estrategias analíticas de inspección/exploración de datos con glimpse(), skim(), describe(). |
| 14:00 – 16:00 |
| **4** | 10/10/2022 | Manejo de datos con dplyr 2, case\_when() para categorización de variables. Etiquetado variables con paquete labelled(). Intro a control de calidad: duplicados, datos perdidos y valores extremos univariados. |
| 14:00 – 16:00 |
| 16/10/2022 | | Trabajo asincrónico. Desarrollo de ‘Problem Set 1’. |
| 22:00 | |
| **MÓDULO 2: ANÁLISIS DESCRIPTIVO E INFERENCIAL BIVARIADO** | | |
| **5** | 12/10/2022 | Inspección inicial de datos 2 y análisis descriptivo con summarise(), count(), group\_by(). Creación automatizada de Tabla 1 reproducible con paquete gtsummary(). |
| 14:00 – 16:00 |
| **6** | 14/10/2022 | Revisión de inferencia estadística: Estimación de intervalos de confianza y pruebas de hipótesis paramétricas y no paramétricas para un grupo, dos grupos y tres grupos con paquete rstatix. Creación automatizada de Tabla 1 reproducible con paquete gtsummary(). |
| 14:00 – 16:00 |
| **7** | 17/10/2022 | Inferencia estadística no paramétrica 2: ¿Qué hacer cuando no se cumplen los supuestos? Uso de paquete infer para pruebas de hipótesis con remuestreo y uso de varianzas robustas. |
| 14:00-16:00 |
| 23/10/2022 | | Trabajo asincrónico. Desarrollo de ‘Problem Set 2’. |
| 22:00 | |
| **MÓDULO 2: INTRODUCCION A LA REGRESION PARA DATOS NUMÉRICOS Y BINARIOS** | | |
| **8** | 19/10/2022 | Regresión para datos numéricos: Regresión lineal, de Poisson y Binomial Negativa con R. Evaluación de supuestos, interpretación de resultados y reporte reproducible con gt\_summary(). Caso aplicado de análisis en efectividad clínica comparativa. |
| 14:00 – 16:00 |
| **9** | 21/10/2022 | Regresión para datos binarios para estimar OR y RR/RP: Regresión Logística Binaria, Log-Binomial y Poisson con R. Evaluación de supuestos, interpretación y reporte reproducible con gt\_summary(). Casos aplicado de análisis en efectividad clínica comparativa. |
| 14:00 – 16:00 |
| **10** | 24/10/2022 | Estrategias de regresión en modelado explicativo 1: Confusión y ajuste. Uso de DAG para seleccionar variables. Caso aplicado de análisis en efectividad clínica comparativa. |
| 14:00 – 16:00 |
| **11** | 26/10/2022 | Estrategias de regresión en modelado explicativo 2: Lidiando con incumplimiento de supuestos (no linealidad y no i.i.d) con R. Caso aplicado de análisis. Casos aplicados de análisis en efectividad clínica comparativa. |
| 14:00 – 16:00 |
| **12** | 28/10/2022 | Estrategias de regresión en modelado explicativo 3: Modificación de efecto y análisis de subgrupos en R. Recomendaciones para robustecer las inferencias cuando se analizan múltiples grupos. Caso aplicado de análisis en efectividad clínica comparativa. |
| 14:00 – 16:00 |
| 30/10/2022 | | Trabajo asincrónico. Desarrollo de ‘Problem Set 3’. |
| 22:00 | |

|  |  |  |
| --- | --- | --- |
| xxx | Horas Cronológicas | Horas Académicas |
| Teoría | 12 | 16 |
| Práctica | 12 | 16 |
| Trabajo independiente\* | 12 | 16 |

*\* Resolución de problema set.*

# ESTRATEGIAS DE FORMACION

Este curso taller consta de 2 partes; teórica que es utilizando el método expositivo dialogado y la parte práctica es a través de talleres de código en vivo en los que los alumnos siguen, paso a paso, las instrucciones que el profesor muestra en clase para el análisis de datos. Los recursos a utilizar son los siguientes:

* Computadora personal
* Aplicativo para conferencias de vídeo y reuniones en línea
* Página web del curso
* Bibliografía recomendada

**CLASES TEORICAS.**

Los alumnos recibirán las clases teóricas de manera virtual. Inicialmente, el docente dictará las sesiones académicas, y los participantes brindarán comentarios respecto al tema. Posteriormente, se entregarán las diapositivas de las sesiones académicas por correo electrónico y a través de una carpeta en Google Drive que serán leídas por los participantes. Las ponencias serán grabadas y se almacenarán en un repositorio académico virtual que será compartido a los participantes. Las clases teóricas se desarrollarán de manera fluida con los talleres (por eso, taller teórico-práctico). En conjunto, las teorías deberían durar entre 30 minutos a 1 hora.

**CLASES PRÁCTICAS: TALLERES.**

Los alumnos recibirán las clases prácticas de manera virtual. Previamente, se hará entrega de conjunto de datos de estudios reales publicados en repositorios abiertos o conjunto de datos “juguete” especialmente seleccionados para las clases por su valor didáctico. Durante las clases prácticas, guiados por el docente a cargo, se realizará el análisis de los datos aplicando los métodos enseñados durante la sesión. Los talleres serán grabados y se almacenarán en un repositorio académico virtual que será compartido a los participantes. Los conjuntos de datos y el código de análisis de datos estará disponible para los estudiantes. Adicionalmente, se dispondrá de 15 minutos adicionales por cada clase práctica para poder hacer preguntas directas al docente a cargo, para solucionar cualquier duda o consulta, o si es que algún alumno lo solicita, hacer una revisión específica de algún aspecto del análisis de datos. Las clases teóricas se desarrollarán de manera fluida con los talleres (por eso, taller teórico-práctico). En conjunto, los talleres deberían durar entre 1 hora a 1 hora y media.

**Conjunto de datos para analizar en casos aplicados:** Siempre que sea posible, trabajaremos con conjunto de datos reales u obtenidos mediante simulación de estudios reales. También haremos uso del conjunto de datos del paquete “medicaldata” de R (). Complementaremos con conjunto de datos “juguete” por motivos didácticos:

* **Módulo 1:**
* Conjunto de datos real del Censo Nacional Penitenciario – Preguntas relacionadas a Salud de los presos. Working Paper titulado: “Self-reported HIV in male Peruvian inmates: results of the 2016 prison census”. Datos y código disponibles en: <https://github.com/culquichicon/2016-Peruvian-Prison-Census>. Artículo preliminar en español disponible en: MINJUS. El Fenómeno Criminológico y la Política Criminal en las Regiones. I Convocatoria Nacional de Artículos. Página 33. Enlace: <https://indagaweb.minjus.gob.pe/wp-content/uploads/2019/12/Libro-I-Convocatoria-Nacional-de-Art%C3%ADculos-compilaci%C3%B2n.pdf>
* Conjunto de datos simulados del estudio: Valdivia M, Soto-Becerra P, Laguna-Barraza R, et al. Effect of a natural supplement containing glucosinolates, phytosterols and citrus flavonoids on body weight and metabolic parameters in a menopausal murine model induced by bilateral ovariectomy. Gynecological Endocrinology 36; 2020. doi: <https://doi.org/10.1080/09513590.2020.1821639>
* Conjunto de datos simulados del estudio: Villar-López M, Soto-Becerra P, Choque R, et al. Safety and tolerability of a natural supplement containing glucosinolates, phytosterols and citrus flavonoids in adult women: a randomized phase I, placebo-controlled, multi-arm, double-blinded clinical trial. Gynecological Endocrinology 37; 2021. Doi: 10.1080/09513590.2021.1960965
* Conjunto de datos reales del estudio: Elmunzer JB, Scheiman JM, Lehman GA, et al. A Randomized Trial of Rectal Indomethacin to Prevent Post-ERCP Pancreatitis. NEJM 2012; 366. doi: 10.1056/NEJMoa1111103. Dataset: “indo\_rct” disponible en: <https://higgi13425.github.io/medicaldata/>
* **Módulo 2:**
* Conjunto de datos simulados del estudio: Villar-López M, Soto-Becerra P, Choque R, et al. Safety and tolerability of a natural supplement containing glucosinolates, phytosterols and citrus flavonoids in adult women: a randomized phase I, placebo-controlled, multi-arm, double-blinded clinical trial. Gynecological Endocrinology 2021; 37. Doi: 10.1080/09513590.2021.1960965
* Conjunto de datos juguetes del libro “Vach W. Regression Models as a Tool in Medical Research. CRC Press – Taylor & Francis Group. First Edition 2013.”. Datasets: “hiaa.dta” y “hb.dta”. Disponibles en: <https://www.uniklinik-freiburg.de/imbi/stud-le/regression-models-as-a-tool-in-medical-research/data-sets.html>
* Conjunto de datos simulados del estudio: Vargas Hererara, Ferández-Navarro, Cabezudo-Pillpe, et al. “Immunogenicity and reactogenicity of a third dose of BNT162b2 vaccine for COVID-19 after a primary regimen with BBIBP-CorV or BNT162b2 vaccines in Lima, Peru.”. Disponible en: <https://www.medrxiv.org/content/10.1101/2022.05.01.22274548v1>
* **Módulo 3:**
* Conjunto de datos reales del estudio: Ramirez-Ramirez R, Soto-Becerra P. Dependencia funcional y diabetes mellitus tipo 2 en adultos mayores afiliados al Seguro Social de Salud del Perú: análisis de la ENSSA-2015. Acta Médica 2020; 37. doi: <http://dx.doi.org/10.35663/amp.2020.374.1075>
* Conjunto de datos reales del estudio: Elmunzer JB, Scheiman JM, Lehman GA, et al. A Randomized Trial of Rectal Indomethacin to Prevent Post-ERCP Pancreatitis. NEJM 2012; 366. doi: 10.1056/NEJMoa1111103. Dataset: “indo\_rct” disponible en: <https://higgi13425.github.io/medicaldata/>
* Conjunto de datos juguetes del libro “Vach W. Regression Models as a Tool in Medical Research. CRC Press – Taylor & Francis Group. First Edition 2013.”. Datasets: “allergy2.dta” y “backpain.dta”. Disponibles en: <https://www.uniklinik-freiburg.de/imbi/stud-le/regression-models-as-a-tool-in-medical-research/data-sets.html>
* Conjunto de datos simulados del estudio: Vargas Hererara, Ferández-Navarro, Cabezudo-Pillpe, et al. “Immunogenicity and reactogenicity of a third dose of BNT162b2 vaccine for COVID-19 after a primary regimen with BBIBP-CorV or BNT162b2 vaccines in Lima, Peru.”. Disponible en: <https://www.medrxiv.org/content/10.1101/2022.05.01.22274548v1>

**PROBLEM SETS.**

Durante todo el curso habrá tres ‘problem sets’ (trad. conjunto de problemas) que corresponderán a un conjunto de datos con una pregunta de investigación y objetivo de análisis pre-establecido. Los alumnos tendrán entre 7 a 10 días para realizar los análisis solicitados y presentarlos como tarea calificada. Asimismo, se compartirá un solucionario al final del cierre de cada entrega. El trabajo deberá realizarse de manera grupal en grupos de entre 4 a 6 participantes o, excepcionalmente, de manera individual. Los grupos deberán confirmarse por correo electrónico a más tardar el 05 de octubre de 2022.

# EVALUACION DE LOS APRENDIZAJES

Para efectos de evaluar a los participantes se tomarán en cuenta los siguientes criterios:

* Problem sets: 90% (30% cada uno)
* Asistencia: 10%

Para efectos de evaluar la asistencia, los participantes deberán haber asistido a mínimo 9/12 sesiones académicas. Para aprobar el entrenamiento, los participantes deberán cumplir con dos requisitos:

* Cumplir con el requisito de asistencia del entrenamiento (9/12 sesiones académicas).
* Aprobar los problema set (calificación mayor o igual a 14.0 en promedio final)

Se otorgará certificado como asistente a quienes asistan mínimo al 75% de clases (equivalente a 9 de las 12), y un certificado como aprobado a quienes adicionalmente aprueben la evaluación final.

La nota mínima aprobatoria para que el participante apruebe el entrenamiento es de 14.00 (trece) en el sistema vigesimal.

# PLANA DOCENTE

* Percy Soto Becerra, M.D.

Investigador RENACYT con publicaciones en revistas científicas indizadas a SCOPUS, Medline y Web of Science. Médico cirujano por la Universidad Nacional Mayor de San Marcos (UNMSM), egresado de la Maestría en Ciencias en Investigación Epidemiológica de la Universidad Peruana Cayetano Heredia (UPCH) y maestrando de la maestría de Estadística de la Universidad Nacional Agraria La Molina (UNALM). Cuenta con más de tres años de experiencia en la formulación, implementación, y gestión de proyectos de investigación científica en salud. Posee una sólida formación en métodos cuantitativos para la investigación epidemiológica y redacción científica, asimismo, ha realizado entrenamiento avanzado y cuenta con experiencia en bioestadística y ciencia de datos. Ha obtenido financiamiento internacional como investigador principal junior; así como participar en investigaciones en salud con la aplicación de métodos avanzados en análisis de datos. Ha sido becario del Centro Internacional Fogarty de los Institutos Nacionales de Salud de los Estados Unidos mediante el proyecto D43 TW007393 Peru Epidemiology Research Training Consortium, el cual le permitió recibir entrenamiento en investigación biomédica. Ha colaborado como investigador en el Centro de Investigaciones Económicas y Sociales de la Universidad San Ignacio de Loyola (USIL), el Instituto de Evaluación de Tecnologías en Salud e Investigación (IETSI) del Seguro Social (EsSalud), el Centro Nacional de Salud Ocupacional (CENSOPAS) y el Centro Nacional de Salud Pública (CNSP) del Instituto Nacioanl de Salud (INS) y el Centro de Salud Global de la UPCH. Actualmente y es investigador asociado a la Universidad Continental. Interesado en aplicar métodos cuantitativos modernos para inferencia causal en epidemiología (efectividad clínica comparativa y evaluación de impacto de programas de salud), investigación clínica pronóstica y diagnóstica con especial énfasis en aplicaciones a medicina de precisión.

Google Scholar: <https://scholar.google.com/citations?user=lYRCQzYAAAAJ&hl=en>

Github: <https://github.com/psotob91>

E-mail: [percys1991@gmail.com](mailto:percys1991@gmail.com)

# DURACIÓN

El curso taller se llevará a cabo en 12 fechas, con una duración de 03 horas teóricas y 05 horas prácticas.

# CERTIFICACIÓN

Se entregará un certificado por el Instituto Nacional de Salud del Niño San Borja, con el auspicio y creditaje académico de una institución universitaria. La certificación se dará a los participantes que hayan culminado y aprobado satisfactoriamente el curso, nota mínima aprobatoria 14. Además de haber cumplido con el 75% de asistencia al curso. La asistencia es obligatoria y con estricta puntualidad.

# BENEFICIOS.

* Institucionales: La imagen institucional del INSNSB mejora al resaltarse su preocupación por promover la investigación biomédica en el Instituto a través del desarrollo de competencias en estadística de sus investigadores.
* Participante: La calidad de los trabajos de investigación mejora debido a las competencias de análisis estadísticos ganados por los participantes del curso.
* Paciente: Las investigaciones, de mejor calidad por las competencias brindadas en el curso, permiten soportar decisiones o políticas basadas en evidencia que mejoran la atención de los pacientes.

# MONITOREO Y EVALUACIÓN DEL PROGRAMA

La presente actividad académica será monitoreada por el Área de Recursos Humanos y su evaluación se realizará a través de una encuesta de satisfacción de la actividad académica a los participantes.

# INFORMACIÓN GENERAL

Cronograma Académico: 3, 5, 7, 10, 12, 14, 17, 19, 21, 24, 26 y 28 de octubre de 2022.

Vacantes: 20 participantes

Horario : 2-4 pm

Lugar : Instituto Nacional de Salud del Niño San Borja – Plataforma virtual

**Organizadora:**

Lic. Jorge Zafra Remigio

Cel. 2300600 (anexo 3051)

Correo: [jzafra@insnsb.gob.pe](mailto:jzafra@insnsb.gob.pe)

# BIBLIOGRAFÍA RECOMENDADA

* Batra, Neale, et al. The Epidemiologist R Handbook. 2021. DOI 10.5281/zenodo.4752646. Capítulos 17. Enlace: <https://epirhandbook.com/en/index.html>
* Cannell B, Livingston M. R for Epidemiology. Capítulo IV. Enlace: <https://www.r4epi.com/>
* Driscoll P, Lecky FArticle 5. An introduction to estimation—2: from z to tEmergency Medicine Journal 2001;18:65-70. <https://emj.bmj.com/content/18/1/65.citation-tools>
* Driscoll P, Lecky FArticle 6. An introduction to hypothesis testing. Parametric comparison of two groups—1Emergency Medicine Journal 2001;18:124-130. <https://emj.bmj.com/content/18/2/124.citation-tools>
* Greenland, S., Senn, S.J., Rothman, K.J. et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 31, 337–350 (2016). <https://doi.org/10.1007/s10654-016-0149-3>
* Altman, N., Krzywinski, M. P values and the search for significance. Nat Methods 14, 3–4 (2017). <https://doi.org/10.1038/nmeth.4120>
* Are confidence intervals better termed “uncertainty intervals”? BMJ 2019; 366 doi: <https://doi.org/10.1136/bmj.l5381> (Published 10 September 2019)
* Valentin Amrhein, David Trafimow & Sander Greenland (2019) Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don’t Expect Replication, The American Statistician, 73:sup1, 262-270, DOI: 10.1080/00031305.2018.1543137
* Hawkins AT, Samuels LR. Use of Confidence Intervals in Interpreting Nonstatistically Significant Results. JAMA. 2021;326(20):2068–2069. doi:10.1001/jama.2021.16172
* Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. marzo de 2019;567(7748):305-7. <https://www.nature.com/articles/d41586-019-00857-9>
* Kim TK. Understanding one-way ANOVA using conceptual figures. Korean J Anesthesiol. 2017;70(1):22-26. doi:10.4097/kjae.2017.70.1.22
* Weissgerber TL, Garcia-Valencia O, Garovic VD, Milic NM, Winham SJ. Why we need to report more than 'Data were Analyzed by t-tests or ANOVA'. Elife. 2018;7:e36163. Published 2018 Dec 21. doi:10.7554/eLife.36163
* Lachin JM. Nonparametric Statistical Analysis. JAMA. 2020;323(20):2080–2081. doi:10.1001/jama.2020.5874
* Cao J, Zhang S. Multiple Comparison Procedures. JAMA. 2014;312(5):543–544. doi:10.1001/jama.2014.9440
* Bewick V, Cheek L, Ball J. Statistics review 9: one-way analysis of variance. Crit Care. 2004 Apr;8(2):130-6. doi: 10.1186/cc2836. Epub 2004 Mar 1. PMID: 15025774; PMCID: PMC420045.
* Bewick V, Cheek L, Ball J. Statistics review 10: further nonparametric methods. Crit Care. 2004 Jun;8(3):196-9. doi: 10.1186/cc2857. Epub 2004 Apr 16. PMID: 15153238; PMCID: PMC468904.
* Hernán M, et al. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. Chance 2019, 32.
* Shmueli G. To Explain or to Predict? Statistical Science 2010; 25. Doi: 10.1214/10-STS330. <https://www.stat.berkeley.edu/~aldous/157/Papers/shmueli.pdf>
* Schooling, C.M., Jones, H.E. Clarifying questions about “risk factors”: predictors versus explanation. Emerg Themes Epidemiol 15, 10 (2018). <https://doi.org/10.1186/s12982-018-0080-z>
* Ramspek CL, Steyerberg EW, Riley RD, Rosendaal FR, Dekkers OM, Dekker FW, van Diepen M. Prediction or causality? A scoping review of their conflation within current observational research. Eur J Epidemiol. 2021 Sep;36(9):889-898. doi: 10.1007/s10654-021-00794-w. Epub 2021 Aug 15. PMID: 34392488; PMCID: PMC8502741.
* Wallisch C, Bach P, Hafermann L, Klein N, Sauerbrei W, et al. (2022) Review of guidance papers on regression modeling in statistical series of medical journals. PLOS ONE 17(1): e0262918. https://doi.org/10.1371/journal.pone.0262918
* Altman, N., Krzywinski, M. Simple linear regression. Nat Methods 12, 999–1000 (2015). <https://doi.org/10.1038/nmeth.3627>
* Krzywinski, M., Altman, N. Multiple linear regression. Nat Methods 12, 1103–1104 (2015). <https://doi.org/10.1038/nmeth.3665>
* Sauerbrei, W., Perperoglou, A., Schmid, M. et al. State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues. Diagn Progn Res 4, 3 (2020). https://doi.org/10.1186/s41512-020-00074-
* Bland JM, Altman DG. Statistics notes. The odds ratio. BMJ. 2000;320(7247):1468. doi:10.1136/bmj.320.7247.1468
* Lever, J., Krzywinski, M. & Altman, N. Logistic regression. Nat Methods 13, 541–542 (2016). <https://doi.org/10.1038/nmeth.3904>
* Harris JKPrimer on binary logistic regression. Family Medicine and Community Health 2021;9:e001290. doi: 10.1136/fmch-2021-001290
* Sedgwick P. Logistic regression. BMJ. 12 de julio de 2013;347:f4488.
* Tamhane AR, Westfall AO, Burkholder GA, Cutter GR. Prevalence odds ratio versus prevalence ratio: choice comes with consequences [published correction appears in Stat Med. 2017 Oct 15;36(23 ):3760]. Stat Med. 2016;35(30):5730-5735. doi:10.1002/sim.7059
* Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004 Apr 1;159(7):702-6. doi: 10.1093/aje/kwh090. PMID: 15033648.
* Cummings P. The Relative Merits of Risk Ratios and Odds Ratios. Arch Pediatr Adolesc Med. 4 de mayo de 2009;163(5):438-45.