
Participación en las XLVI Jornadas Nacionales de Estadística

Los días 17 al 20 de octubre de 2022, en la hermosa ciudad de San Pedro de Atacama, la Sociedad Chilena de Estadística (SOCHE) realizó una nueva versión de las tradicionales Jornadas Nacionales de Estadística. Este evento volvió a la presencialidad después de más de dos años de la etapa más dura de la pandemia de COVID-19.

En esta ocasión el Programa de Bioestadística y el Magíster en Bioestadística tuvieron una activa participación, con 10 trabajos presentados tanto por estudiantes, docentes y graduadas/os. A continuación les mostramos los resúmenes de los trabajos presentados, los que hemos ordenado cronológicamente según el programa del encuentro.

Evaluación de las dimensiones tradicionales de la segregación residencial por nivel educacional en Chile

Autoras/es: Sandra Flores Alvarado, Tamara Doberti Herrera, Mauricio Fuentes Alburquenque

Expositora: Sandra Flores Alvarado, académica del Programa de Bioestadística, docente del Magíster en Bioestadística

Sesión temática “Estadística y Sociedad”

Introducción: La segregación residencial se ha propuesto como parte del camino causal de diferentes resultados de salud y se ha estudiado principalmente ligada a la raza. En 1988, Massey y Denton propusieron 5 dimensiones teóricas de la segregación residencial: desigualdad, exposición, agrupamiento, concentración y centralización, que confirmaron utilizando 20 índices de segregación construidos a partir de las proporciones de razas observadas en el censo de los Estados Unidos en la década de 1980. Estos índices aún son ampliamente utilizados, sin embargo, en Latinoamérica y especialmente en Chile, la segregación se encuentra determinada por niveles socioeconómicos y educacionales, no por raza. **Objetivo:** Este estudio pretendió evaluar si las dimensiones tradicionales de la segregación residencial se mantienen cuando se utilizan variables de segregación y poblaciones diferentes a las utilizadas por Massey y Denton, tales como la segregación por nivel educacional en la población chilena. **Metodología:** Replicamos la evaluación realizada por Massey y Denton en 247 comunas chilenas utilizando la proporción de población mayor de 25 años con educación universitaria completa como variable de segregación. Para los 20 índices de segregación residencial, construimos una matriz de correlaciones y llevamos a cabo un análisis factorial de ejes principales y comparamos nuestros resultados con las dimensiones obtenidas por Massey y Denton. **Resultados:** No hay una alta intercorrelación entre los índices asociados con las 5 dimensiones teóricas excepto por la dimensión de desigualdad y de modo parcial, para la dimensión de exposición. Esto sugiere que los índices de la primera dimensión son capaces de medir de manera efectiva el constructo en datos chilenos, mientras que esto sólo ocurre parcialmente para la segunda dimensión y no ocurre en las otras tres. Esto se reafirma a través del análisis factorial exploratorio, en que no es posible verificar de manera empírica las mismas 5 dimensiones. Para la población chilena, el primer factor predomina en la explicación de la varianza y se asocia tanto a la dimensión de desigualdad como a las otras dimensiones teóricas de manera parcial. La segunda y tercera dimensión también se asocian parcialmente a las dimensiones correspondientes a exposición y concentración, sin embargo, explican proporciones de la varianza inferiores a las observadas originalmente por Massey y Denton y se asocian a una menor cantidad de índices en cada caso. **Conclusiones:** Este estudio demuestra la necesidad de caracterizar localmente no solo los índices de segregación residencial, sino también las dimensiones teóricas a las que se refieren. Futuros estudios deberán profundizar las implicancias teóricas y prácticas de los factores observados en el contexto chileno y latinoamericano, considerando las diferencias históricas y culturales entre estos territorios y los Estados Unidos. Estos análisis permitirán mejorar las interpretaciones de los efectos que tiene la segregación residencial sobre la salud, y de esta forma, entender de mejor manera los caminos causales a través de los que ocurren estos efectos.

Factores causales asociados a polivictimización de violencia intrafamiliar en mujeres participantes de la ENVIF-VCM año 2020 a través de grafos acíclicos dirigidos y modelos de ecuaciones estructurales

Autoras: Ana Karen Espinoza Ortiz, Sandra Flores Alvarado, Francisca Crispi Galleguillos

Expositora: Ana Karen Espinoza Ortiz, graduada del Magíster en Bioestadística

Sesión temática “Estadística y Sociedad”

Introducción: La violencia de género es un problema de salud pública y uno de los tipos de violación a los derechos humanos más frecuentes. Según datos entregados el año 2021 por la Organización Mundial de la Salud

736 millones de mujeres han sido violentadas globalmente. En Chile, la prevalencia de victimización por violencia intrafamiliar se ha estimado entre 1 y 51,9% para las mujeres, con consecuencias relevantes en la salud de las víctimas. Se hace imperativo analizar este fenómeno, comprender sus causas, permitiendo implementar políticas públicas que permitan reducir su prevalencia y consecuencias. El principal objetivo de esta investigación es proponer un modelo causal que explique la ocurrencia de polivictimización de violencia intrafamiliar en mujeres residentes en Chile. **Métodos:** Se exploran los datos provenientes de la encuesta de violencia contra la mujer en el contexto intrafamiliar (ENVIF-VCM) 2020, la que presenta una muestra representativa de mujeres mayores de 15 años en Chile. Se describen características sociodemográficas de la muestra, estimando prevalencias de los tipos de violencia y polivictimización. Se construye un grafo acíclico dirigido (DAG) a partir de la bibliografía relacionada a polivictimización y se modelan las vías causales encontradas en el DAG mediante modelos de ecuaciones estructurales (SEM). **Resultados:** La muestra está compuesta por 7.735 mujeres mayores de 15 años, con un promedio de edad de 46,8 años. El 95,6% es de nacionalidad chilena. El 48% tiene un nivel socioeconómico bajo, un 48% posee educación media completa y el 40,2% se encontraba trabajando al momento de realizar la encuesta. A nivel nacional se describen prevalencias en los últimos 12 meses y en la vida de 3,34% y 15,52% de violencia física; 18,64% y 37,2% de violencia psicológica; 2,96% y 9,88% de violencia sexual; y 20,02% y 40,81% de polivictimización. Destacan las altas prevalencias de violencia en Coquimbo (25,3%, 6,58% y 27,7% para violencia psicológica, sexual y polivictimización) y Tarapacá (5,53% para violencia física), al analizar los últimos 12 meses. Cuando el fenómeno es medido en la vida, la región del Biobío destaca con un 20,07% para violencia física y 48,56% para violencia psicológica; Coquimbo con un 14,12% para violencia sexual y Arica y Parinacota para polivictimización con un 50,7%. Se estableció como principal factor de exposición en el DAG la Internalización de cultura de violencia (ICV). Se modeló mediante SEM encontrando asociación entre ICV y el bajo nivel socioeconómico ($\beta = -0,089$; $p < 0,001$), edades avanzadas ($\beta = 0,005$; $p < 0,001$) y el bajo nivel educacional ($\beta = -0,153$; $p < 0,001$). Los bajos niveles de red de apoyo aumentan los niveles de polivictimización cuando interactúan con el consumo problemático de alcohol y problemas psicológicos de la mujer ($\beta = 0,236$; $p < 0,001$). No se evidencia relación directa entre ICV y polivictimización ($\beta = -0,054$; $p = 0,351$). **Conclusiones:** Se establece que existen factores sociodemográficos y contextuales asociados con la presencia de polivictimización de violencia intrafamiliar. Para mejorar el estudio de esta asociación, se sugiere incorporar otros elementos teóricos al diseño de la ENVIF-VCM, tales como preguntas que mejoren la especificación y análisis del constructo cultura de violencia, la inclusión de población rural y características del agresor.

Análisis de datos funcionales aplicado a espectros infrarrojos para la evaluación de la formación de hidroxiapatita en distintos biomateriales para la regeneración ósea

Autores/as: Amaru Agüero Jiménez, Melissa Lezana Zúñiga, Bárbara Castillo Villalobos, Cristián Covarrubias Gallardo, Felipe Medina Marín, Andrea Canals Cifuentes

Expositor: Amaru Agüero Jiménez, tesista del Magíster en Bioestadística

Introducción: El ensayo en fluido corporal simulado (SBF) es utilizado para medir la capacidad de formación de hidroxiapatita (HA) en superficies de biomateriales para la regeneración ósea (ISO23317:2014). Ésta puede ser analizada por Espectrometría Infrarroja con Transformada de Fourier (FTIR), la cual proporciona un espectro de transmitancia de las bandas asociadas a grupos funcionales de los biomateriales, permitiendo su identificación. En la formación de HA, se observa una disminución de la transmitancia en las bandas vibratorias: 567-610 cm^{-1} y 960-1100 cm^{-1} , correspondiente a distintos modos vibratorios del fosfato (PO_4) de la HA. El análisis de estas bandas es cualitativo y no permite cuantificar la variabilidad ni las diferencias estadísticas en la formación de HA entre distintos biomateriales. El análisis de datos funcionales (FDA) es una rama de la estadística que estudia curvas indexadas a un parámetro continuo y provee las metodologías adecuadas para el análisis estadístico de los espectros FTIR. El FDA convierte los datos observados en datos funcionales $X_i(t)$ utilizando funciones bases tales que:

$$X_i(t) \approx \sum_{l=1}^K c_{il} \phi_l(t), \quad i = 1, \dots, n; \quad 1 \leq i \leq K,$$

donde c_{il} es un escalar de la i -ésima curva y ϕ_l son una colección de K funciones base. En este trabajo se utilizó FDA para describir la formación de HA en los espectros FTIR de los biomateriales: Zeolita (SAPO-5), Zeolite-Like-Metal-Organic-Framework (zMOF), nano-bioactive-glass (nBG), nano-mesoporous-bioactive-glass (nMBG) y copper-nano-bioactive-glass (Cu-nBG), expuestos a 3, 7 y 14 días en SBF. **Métodos:** Para el FDA se utilizó R (v4.20) y RStudio (v2022.07.1+554) (paquetes `fda` y `fda.usc`). Se consideró un $n = 30$, para cada combinación de biomaterial y tiempo. Las bandas 567-610 cm^{-1} y 960-1100 cm^{-1} se analizaron indepen-

dientemente. Se ocupó el método de validación cruzada generalizada (GCV) para obtener un número óptimo de K bases funcionales de Fourier. Para las funciones de transmitancias $X_i(t)$ de los espectros FTIR se graficaron: un boxplot funcional (Q_1 , mediana, Q_3); la media funcional $\bar{X}(t) = \frac{1}{n} \sum_{i=1}^n X_i(t)$ con intervalos de confianza (IC), $\bar{X}(t) \pm Z_\alpha \frac{\sigma(t)}{\sqrt{n}}$, donde $Z \sim G(0, \rho)$, $\rho(t, s) = \frac{\gamma(t, s)}{\sigma(t)\sigma(s)}$, con $1 - \alpha = 0,95$; y la covarianza $\hat{\gamma}(t, s) = \frac{1}{n-1} [X_i(t) - \bar{X}_i(t)][X_i(s) - \bar{X}_i(s)]$, en donde $t, s \in T(a, b)$, con $T_a = (567, 610)$ y $T_b = (960, 1100)$. **Resultados:** Mediante GCV se obtuvo $K = 6$ y $K = 12$ funciones bases, para las bandas 567-610 cm^{-1} y 960-1100 cm^{-1} respectivamente. En todos los biomateriales (excepto SAPO-5) hay una disminución de la media funcional $X(t)$ (i.e. un aumento en la formación de HA) a partir del día 3 en SBF. La variabilidad en la formación de HA, cambia en el tiempo y depende de cada biomaterial, disminuyendo a mayor tiempo en SBF. La banda 567-610 cm^{-1} , en todos los casos, presentó IC más precisos, menor covarianza y rango intercuartílico. En una búsqueda exhaustiva, no se encontró evidencia de FDA aplicada en espectros FTIR para estudiar la formación de HA en biomateriales. Dada la naturaleza de los datos FTIR (que ya contienen la transformada de Fourier), al utilizar el GCV, se obtuvo un número reducido de K bases funcionales. **Conclusiones:** El FDA permitió describir de forma cuantitativa y detallada la formación de HA, esto ayudará a seleccionar el mejor biomaterial para la regeneración ósea. Esta experiencia nos revela que el FDA podría utilizarse para el análisis estadístico de otras señales biológicas que solo se analizan cualitativamente.

Propiedades de dos estimadores frecuentistas para la estimación de parámetros en muestras finitas de distribuciones “pseudo-exponenciales” bivariantes

Autores: Felipe Medina Marín, Milan Stehlik

Expositor: Felipe Medina Marín, académico del Programa de Bioestadística, docente del Magíster en Bioestadística

Antecedentes: Una distribución “pseudo-exponencial” bivalente (BPE) es un modelo tal que: i) X y Y están negativamente correlacionadas; ii) sus distribuciones marginales son exponenciales; y iii) una de sus distribuciones condicionales al evento de exceso de la otra, i.e., $Y|X > x$, también es exponencial. Este modelo mezcla las aproximaciones de modelos condicionalmente especificados y los modelos pseudo-exponenciales, y puede ser aplicado en áreas tan distintas como lo son salud pública, hidrometeorología, y contaminación acuática. En términos de su función de sobrevivencia bivalente, el modelo BPE puede escribirse simplemente como

$$S_{X,Y}(x, y) = P(X > x, Y > y) = \exp(-\theta_1 x - \beta(x)y)$$

donde el soporte de (X, Y) es $(0, \infty)^2$, $\theta_1 > 0$, y β es una función positiva que depende de $\theta = (\theta_1, \theta_2, \theta_3)$, y cumple con que $0 \leq \beta'(x) \leq \theta_1 \beta(x)$ para todo $x > 0$, entre otros requerimientos. Para facilitar la interpretación de θ , es posible parametrizar $\beta(x)$ de forma que θ_1 y $\beta(0) = \theta_2$ sean las tasas de marginales de X y Y , respectivamente, y θ_3 en el intervalo cerrado entre 0 y 1 esté asociado a la tasa de $Y|X > x$ y la asociación entre X y Y . **Objetivo:** Estudiar las propiedades en muestras finitas de los estimadores basados en los métodos de los momentos generalizados (GMME) y de la máxima verosimilitud (MLE) en la estimación de los parámetros de modelos pseudo-exponencial bivariantes. **Metodología:** Estudiamos cuatro modelos BPE, determinamos sus GMME y MLE, y definimos dos algoritmos de simulación basados en el método de composición. En **R** programamos funciones para simular datos y estimar parámetros, y usamos los paquetes `simulator` y `looplot` para realizar y analizar los resultados de un estudio de simulación que fue planificado bajo las directrices ADEMP. **Resultados.** Como se espera de un buen estimador, GMME y MLE mostraron sesgos que se reducían para la estimación de todos los parámetros a medida que aumentaba el tamaño muestral, aunque con algunos patrones complejos de magnitudes y signos dependientes de θ . En general los sesgos para estimar θ_1 y θ_2 fueron positivos, con magnitudes proporcionales al estimando (MLE) o relacionadas con todo el vector θ (GMME). Por otro lado, el sesgo estandarizado en la estimación de θ_3 , independientemente del valor de θ_1 y θ_2 , fue mayor a medida que θ_3 se acercaba a 0 o a 1, lo que podría deberse a que el sesgo de ambos estimadores disminuía más lento que sus errores estándares a medida que el tamaño muestral aumenta. Al estudiar la matriz de varianza-covarianza de los estimadores se observó que, a medida que aumentó el tamaño muestral, disminuyeron tanto las varianzas como las covarianzas, siendo estas últimas en la mayoría de los casos negativas. **Conclusiones:** En general, recomendamos usar MLE por sobre GMME cuando se desean estimar los parámetros de un modelo BPE, ya que MLE presentó menos problemas de convergencia y patrones de comportamiento más sencillos para los efectos de los parámetros y tamaños muestrales sobre sus propiedades estadísticas.

Modelo de tiempo al evento para estimar el riesgo de readmisión a tratamiento de rehabilitación por consumo de sustancias en Chile

Autor: José Antonio Ruiz-Tagle Maturana

Expositor: José Antonio Ruiz-Tagle Maturana, graduado del Magíster en Bioestadística

Introducción: Los trastornos por uso de sustancias pueden ser entendidos como una condición crónica, por ende, es esperable que los pacientes presenten varios intentos de rehabilitación en su trayectoria, sin embargo, bajo una perspectiva biomédica y de eficiencia de los sistemas sanitarios, las readmisiones suelen verse como un evento negativo que evidencian el sobre uso de los servicios de salud y la falta de sostenibilidad de los logros terapéuticos. Esta investigación busca identificar los factores que se relacionan con el riesgo a ser readmitido en tratamientos de rehabilitación en Chile. **Materiales y método:** Utilizamos una base de datos que incluye información de los centros con convenio SENDA y contempla todas las admisiones del programa de población adulta en el período 2010-2019. Se estima que el Sistema de Información de Tratamientos (SISTRAT) de SENDA contiene entre el 80-85 % de todos los programas de tratamiento que se llevan a cabo en Chile ($n = 109.756$). Como los pacientes pueden experimentar múltiples eventos durante el seguimiento (readmisiones), los modelos tiempo al evento más comunes como el modelo de riesgos proporcionales de Cox no son la mejor opción, esto debido a que solo se incluye la información hasta el primer evento. Para utilizar toda la información disponible, ajustamos un modelo Prentice, Williams, and Petersen gap time (PWP-GT). Este “reinicia” el tiempo inicial a 0 cada vez que se experimenta el evento. Adicionalmente, descuenta el tiempo en que los individuos no están en riesgo de experimentar el evento, lo que resulta conveniente dado que los pacientes no están en riesgo de ser readmitidos mientras se encuentran en tratamiento. Los modelos PWP utilizan *grouped jackknife* para estimar la matriz de varianza-covarianza y tienen diferentes riesgos basales entre eventos. Esto significa que las personas no están en riesgo de experimentar el $k^{\text{ésimo}}$ evento hasta que experimenten el $(k - 1)^{\text{ésimo}}$ evento. Adicionalmente, controlamos por efectos que varían en el tiempo estimando *Average Hazard Ratios*. **Resultados:** Encontramos que alcanzar una alta terapéutica reduce el riesgo de readmisión en 11 % (AHR[IC95 %] = 0,89 [0,85-0,92]). También, las personas que consumen pasta base como sustancia principal tienen un 46 % más riesgo de readmisión en comparación a los consumidores de alcohol (AHR[IC95 %] = 1,46 [1,41-1,52]). Por último, los pacientes con un compromiso biopsicosocial severo presentan 43 % más riesgo de readmisión que los pacientes con un compromiso leve (AHR[IC95 %] = 1,43 [1,35-1,51]). **Conclusiones:** Los programas de tratamiento de rehabilitación en población general tienen un impacto positivo en los pacientes, disminuyendo el riesgo de readmisión. También, encontramos que algunas características basales como el tipo de sustancia consumida con mayor frecuencia y la severidad del compromiso biopsicosocial predisponen a ciertos individuos a ser readmitidos. Es importante recalcar la necesidad de seguimiento post tratamiento de los pacientes, de esta manera, se podría mejorar la toma de decisiones, adoptando estrategias que busquen aumentar el tiempo transcurrido entre el alta y la recaída, a la vez que disminuye el tiempo entre la recaída y la readmisión a tratamiento.

Estimación ante la presencia de datos perdidos: orientaciones para su enseñanza

Autor: Óscar Rivera Cortez

Expositor: Óscar Rivera Cortez, estudiante del Magíster en Bioestadística

Introducción: La presencia de datos perdidos es un problema habitual en estudios de diversas áreas científicas, que utilizan como insumo una muestra. La ausencia de una o varias observaciones condiciona el cómo se llevan a cabo los análisis estadísticos, dado que los métodos habituales suponen bases de datos “rectangulares” (completas). Por otro lado, la credibilidad de un estudio (y sus resultados) podría verse reducida si el número de casos perdidos es muy alto. El trabajo realizado es una propuesta didáctica, donde nos aproximamos de forma básica e intuitiva al problema, empezando por métodos de imputación sencillos que permiten, de forma sistemática, reemplazar los valores perdidos por una estimación obtenida a partir del resto de datos completos. **Metodología:** Estos métodos podrían levantar ciertas suspicacias obvias, como por ejemplo: ¿cuándo es correcto imputar un dato perdido?, ¿cómo podríamos comparar la efectividad o desempeño de los distintos métodos de imputación?, ¿qué porcentaje de datos perdidos consideramos aceptable?, entre otras. A partir de lo que se encuentra en la literatura, clasificamos los datos perdidos según las circunstancias en que éstos se pierden. La principal diferencia tiene relación con la intervención del azar o no. **Resultados:** Ilustramos sólo en el primer caso, cuando los datos se pierden de forma aleatoria. Concretamente, se proponen breves ejercicios de simulación, usando el software RStudio, para comparar distintos métodos de imputación para variables numéricas y categóricas. Las simulaciones nos permiten trabajar con una base de datos completa, a partir de la cual calculamos estimaciones “completas”. El mecanismo de pérdida de datos, aleatorio, se puede replicar un gran

número de veces con las funciones base de **R**. Así obtenemos un gran número de bases incompletas, sobre las cuales podemos probar distintos métodos de imputación. El contar con estimaciones “completas” e “imputadas” nos permite calcular diferencias, que se usaron para comparar el desempeño entre los diferentes métodos de imputación propuestos. **Conclusiones:** Si bien existe *software* que permite llevar a cabo métodos de imputación de forma automatizada, se hizo un esfuerzo por generar un código simple que evidencie de forma clara en qué consiste cada uno de los métodos estudiados, con un énfasis en mostrar también cómo comparamos la efectividad de cada uno. Esta aproximación a la solución del problema tiene relevancia, dado que utiliza conceptos sencillos como los de estimador, estimación, varianza o iteración. Esto hace que los conocimientos previos al abordaje del contenido nuevo es, discutiblemente, mínimo. Además resulta sencillo “escalar” la metodología. Es decir, generar propuestas alternativas a los métodos de imputación propuestos, que sean más complejas. Finalmente, la habilidad de simular es transversal, por lo que puede ser una herramienta para resolver o ilustrar una variedad de problemas de inferencia y otras áreas asociadas.

Problema de Monty Hall: aversión a la pérdida y remordimiento adelantado

Autor: Mauricio Fuentes Alburquenque

Expositor: Mauricio Fuentes Alburquenque, académico del Programa de Bioestadística, docente del Magíster en Bioestadística

Introducción: El problema de Monty Hall es un dilema, a estas alturas clásico, de probabilidades y se caracteriza por ser contraintuitivo, incluso para personas con amplios conocimientos en la materia. Se trata de un concurso en donde se presentan tres puertas a un participante, de las cuales sólo una contiene un premio, y él o ella debe elegir una de las puertas. Luego el animador, llamado Monty Hall (famoso animador estadounidense), que sabe dónde está el premio, abre una de las otras dos puertas que está vacía y pregunta al participante si desea cambiar su elección o mantenerse con la puerta ya escogida. Como ya se ha demostrado, cambiarse de puerta tiene una probabilidad de $2/3$ de ganar el premio, contra $1/3$ en caso de mantener la elección original.

Objetivo: Investigar cómo, enfrentadas a este problema, las personas evalúan las probabilidades de ganar y si la decisión que tomarían es consistente. **Método:** Se planteó el problema de Monty Hall a 49 estudiantes de postgrado en salud pública, antes y después de haber visto contenidos de probabilidades en el curso de Bioestadística del semestre en curso en ese momento (año 2017). Los participantes debieron responder dos preguntas en cada ocasión, la primera sobre la alternativa que creían más probable de ganar y la segunda sobre la decisión que tomarían. **Resultados:** En la primera ocasión un 11 % señaló que cambiarse de puerta tenía más probabilidades de ganar y un 32 % respondió que daba lo mismo cambiarse o no, a la vez que el 91 % decidiría no cambiar de puerta. De los 15 sujetos que dijeron que daba lo mismo qué elección tomar, 13 manifestaron que no cambiarían su elección. En la segunda ocasión un 8 % señaló que era más conveniente cambiarse y un 86 % que ambas alternativas tenían igual probabilidad de ganar, diferencia estadísticamente significativa con respecto a la primera ocasión. A su vez, el 92 % indicó que no cambiaría la puerta seleccionada, proporción que no difiere significativamente de la obtenida la primera vez. En esta oportunidad los 42 participantes que dijeron que ambas elecciones tienen igual probabilidad de ganar respondieron que se quedarían con la puerta seleccionada. **Conclusiones:** Las personas acertaron en mayor proporción a la alternativa con más posibilidades de ganar cuando no habían visto los contenidos de probabilidades, lo que podría indicar que este dilema es más contraintuitivo cuando se tienen más conocimientos de dicha materia. En cuanto a la decisión, en ambas ocasiones más del 90 % de los participantes señaló que decidiría no cambiar la opción inicial, lo que se observó prácticamente en el total de los sujetos que dijeron que las probabilidades de ganar eran iguales manteniendo la puerta o cambiándola. Esto apoyaría lo encontrado en investigaciones respecto a la influencia de sesgos de *statu quo* y la aversión a la pérdida, en donde las personas darían más valor (negativo) a la posibilidad de no ganar el premio por haberse cambiado que por no haberlo hecho, posiblemente adelantando un sentimiento de remordimiento que preferirían evitar.

Ubicación geoespacial de jóvenes entre 15 a 29 años expuestos a ciberacoso y que declaran tener síntomas asociados a la depresión en Chile, años 2020 - 2021

Autora: María Eugenia García Fernández

Expositora: María Eugenia García Fernández, estudiante del Magíster en Bioestadística

Introducción: La finalidad de esta investigación es estudiar la distribución geográfica de los jóvenes en Chile que declaran haber sido acosados virtualmente y además informan síntomas asociados a la depresión para determinar lo siguiente: ¿Tiene un joven en Chile mayor o menor riesgo a sufrir ciberacoso y/o estar deprimido, por la

región geográfica donde vive? **Materiales y métodos:** El estudio fue transversal, utilizando datos auto informados. Se recolectaron tres muestras independientes para replicación y comparación, en agosto 2020, diciembre 2020 y diciembre 2021. La metodología de muestreo emplea inteligencia artificial para identificar a los participantes en base a información disponible públicamente, proporcionada voluntariamente por los participantes en estudios previos. La estrategia de muestreo fue probabilística y estratificada con asignación proporcional al tamaño de cada estrato. La selección de la muestra se realizó en dos fases. La probabilidad de ser un usuario de Internet se consideró en primer lugar y en segundo lugar la probabilidad de ser víctima del ciberacoso. Los tamaños de muestra respectivos fueron 2.056 , 2.370 y 3.818 jóvenes entre 15 y 29 años. Las muestras se ajustaron activamente ya que las personas aceptaban o declinaban libremente la invitación a participar. Para determinar la ubicación geográfica de la muestra, se le pregunta al encuestado en qué región vive, y esta información se cruza con la ubicación, como la señal GPS del dispositivo, y cierta información de conexión, como la conexión Wi-Fi o IP (la dirección del protocolo de Internet), nos permite conocer su ubicación (con permiso del usuario). Para evaluar los niveles de depresión, utilizamos el Cuestionario de Salud del Paciente-9 o PHQ-9. **Resultados y Conclusiones:** Una de las principales contribuciones de este estudio es que nos permitió ver que en todas las regiones de Chile existen altos porcentajes de ciberacoso en jóvenes entre 15 y 29 años, superiores al 30%. Por lo que indistintamente de la región en Chile donde vivan los jóvenes reportan altos niveles de ciberacoso, indicando así que este fenómeno corresponde más a la exposición de ambientes digitales y no necesariamente zonas geográficas. Al analizar los resultados como Metaestudio y calcular el promedio móvil de los estudios nacionales para los años 2020 y 2021, se observa que las regiones que presentan los porcentajes más altos de exposición al ciberacoso son: O'Higgins (59%), Atacama (58%) y Ñuble (53%). La ubicación geográfica de esta población ha sido insumo para campañas comunicacionales de salud pública e iniciativas y/o actividades de prevención, información y sensibilización orientadas principalmente a segmentos identificados como más vulnerables o expuestos a ciberacoso.

Diagnóstico del método RDS (*Respondent-Driven Sampling*) en una muestra de migrantes de la Región Metropolitana

Autoras/es: Mauricio Fuentes Alburquenque, Julieta Belmar Prieto, Valeria Stuardo Ávila

Expositor: Mauricio Fuentes Alburquenque, académico del Programa de Bioestadística, docente del Magíster en Bioestadística

Introducción: El RDS (*Respondent-Driven Sampling*) es un método de muestreo para poblaciones ocultas o de difícil acceso. La recolección de la muestra se realiza a través de un reclutamiento tipo “bola de nieve”, donde los propios participantes son quienes reclutan a personas de la misma población objetivo, y donde se consideran y registran características de la red social de los individuos. Dichas características son consideradas posteriormente en la obtención de los estimadores de interés para la población objetivo. Desde su publicación original por Heckathorn (1997), el método RDS ha tenido una importante evolución, especialmente en los estimadores desarrollados, los que han ido incorporando y corrigiendo las distintas fuentes de sesgo presentes. Estos estimadores tienen una serie de supuestos que en muchos casos pueden ser poco realistas y que es necesario evaluar en la medida que sea posible. **Objetivo:** Realizar un diagnóstico de los supuestos del método RDS en un estudio de población migrante de Venezuela en la Región Metropolitana, cuya muestra corresponde al año 2022, usando lo propuesto por Gile et al (2015). **Método:** Se evalúa el supuesto de muestreo con reemplazo a través de las respuestas sobre los contactos conocidos que ya han participado en el estudio, y el posible efecto local de población finita comparando los estimadores VH (Volz-Heckathorn) y SS (successive sampling) de Gile. Para evaluar la dependencia de las semillas en las estimaciones se usa la evolución de los estimadores para cada semilla mediante gráficos de “cuello de botella” y de “todos los puntos” (*all points plot*). Se analizan las respuestas para evaluar el supuesto de reciprocidad, así como el tamaño de la red (grado) y la ventana de tiempo para reclutar a otros participantes. **Resultados:** El equilibrio muestral se logra para la mayoría de las características que definieron la selección de las semillas aunque existe un grado de influencia de éstas. Los supuestos del método RDS se cumplen en gran medida, con algunos resultados que indican algún grado de “embotellamiento”, particularmente en la cadena generada por una de las tres semillas. **Conclusión:** El diagnóstico permite obtener información respecto a las limitaciones y sesgos presentes en la muestra de este estudio, que es fundamental para sus conclusiones y para abordarlas en futuros estudios que utilicen muestreo RDS.

Evaluación del sistema de vigilancia centinela de IRAG para la detección de tendencias de COVID-19

Autoras/es: Sandra Flores Alvarado, María Fernanda Olivares, Christian García, Cristóbal Cuadrado

Expositora: Sandra Flores Alvarado, académica del Programa de Bioestadística, docente del Magíster en Bioestadística

Introducción: La vigilancia epidemiológica consiste en la recolección, análisis, interpretación y difusión permanente de información de salud para la prevención, planificación, implementación y evaluación de acciones de salud pública y ha sido considerada como la mejor estrategia en la prevención de epidemias al permitir identificar brotes y patógenos. Los métodos de vigilancia se enfocan en ser prácticos, uniformes y rápidos, por sobre su precisión o completitud, con el propósito de observar cambios de tendencia y tomar medidas de control. Entre ellos, la vigilancia centinela se basa en la selección de grupos de la población para representar su experiencia particular, asumiéndolos como representativos de la población a la que pertenecen, pese a que el método de selección no se encuentra enfocado en la validez estadística de la muestra. **Objetivos:** El objetivo de este trabajo consistió en evaluar la capacidad de los centros centinela de enfermedades respiratorias para detectar cambios de tendencia y magnitud de COVID-19 comparando las tendencias poblacionales descritas a partir de información censal de casos de COVID-19 con las tendencias estimadas a partir de vigilancia centinela de enfermedades respiratorias. **Metodología:** Se trabajó con las bases de datos de vigilancia universal de COVID-19 y vigilancia centinela IRAG entre marzo de 2020 y noviembre de 2021 provistas por el Ministerio de Salud. Se obtuvieron conteos de casos notificados a nivel de semana, estratificados por sexo, edad y unidad geográfica. Se analizaron descriptivamente para caracterizar las proporciones de individuos por sexo y edad, las proporciones de casos e incidencias acumuladas por unidad geográfica y la incidencia semanal de casos por sexo, edad y unidad geográfica. Las series se comparan escaladas y centradas utilizando medidas de distancias elásticas que permiten establecer similitudes entre sus formas, pese a diferencias temporales. **Resultados:** A nivel nacional se observa que la curva de incidencia de centros centinela IRAG se aproxima a la curva de incidencia de la vigilancia universal de COVID-19, por lo que estos centros permiten detectar cambios de tendencia y magnitud a este nivel. A nivel de macrozona, la curva de incidencia de IRAG se aproxima a la curva universal en todas las macrozonas excepto en la macrozona centro sur. A nivel de región, la curva de incidencia de IRAG se aproxima a la curva universal en Tarapacá, Antofagasta, Valparaíso, Metropolitana, Biobío, La Araucanía y Los Lagos. Las mejores aproximaciones se dan en la región Metropolitana o zona Centro y región del Biobío o zona Sur, mientras que las peores aproximaciones a nivel de macrozona se encuentran en la zona Centro-Sur y a nivel de región del establecimiento se encuentran en Magallanes y no hay información para Arica y Parinacota, Atacama, Coquimbo, O'Higgins, Maule, Ñuble, Los Ríos y Aysén. **Conclusiones:** Los centros centinela existentes informan adecuadamente respecto de la zona/región en que se sitúan. Los centros centinelas actuales permiten obtener información confiable de las dinámicas de COVID-19 a nivel nacional, sin embargo, no permiten realizar vigilancia a nivel de macrozonas o regiones ya que no hay información suficiente para todas ellas.