

DEPÓSITO LEGAL ZU2020000153

ISSN 0041-8811

E-ISSN 2665-0428

Revista de la Universidad del Zulia

Fundada en 1947
por el Dr. Jesús Enrique Lossada



Ciencias
Exactas,
Naturales
y de la Salud

77
ANIVERSARIO

Año 15 N° 43
Mayo - Agosto 2024
Tercera Época
Maracaibo-Venezuela

Estimación de intervalos de confianza Bootstrap de proporciones de factores asociados a inhumaciones por causa de COVID-19

Diego Iván Santillán-Espinoza*
José Luis Pérez-Rojas **
Fabián Roberto Allauca-Pancho ***
Nora Tahirí Mejía-Cabezas ****
Juan Carlos Santillán-Lima *****

RESUMEN

El objetivo del presente trabajo consiste en estimar los intervalos de confianza Bootstrap de proporciones de factores asociados a inhumaciones por causa de COVID-19, en el cementerio general de Riobamba - Ecuador, período marzo 2020 – abril 2021. El método Bootstrap consiste en realizar re muestreos con repetición, es decir, obteniendo muestras mediante algún procedimiento aleatorio que utilice la muestra original. De esta manera, se han identificado los grupos vulnerables para inhumaciones a causa de COVID-19, por sexo, edad y edad de acuerdo a sexo. Con respecto al análisis de vulnerabilidad de los grupos, en cuanto a sexo, el grupo más vulnerable es el sexo masculino. A su vez, dentro de intervalos de edades determinadas mediante los ciclos de vida, se encontró que el grupo más vulnerable es el conocido como adultos mayores, que son las personas que tienen una edad superior a 60 años, tanto para los sexos masculino y femenino.

PALABRAS CLAVE: Estadística, Bootstrap, COVID-19, Ecuador.

*Docente en la Escuela Superior Politécnica del Chimborazo – Ecuador. ORCID: <https://orcid.org/0000-0002-4213-1936>. E-mail: ivan.santillan@espoch.edu.ec

**Docente en la Escuela Superior Politécnica del Chimborazo – Ecuador. ORCID: <https://orcid.org/0000-0002-8958-5556>. E-mail: jose.perezl@espoch.edu.ec

***Docente en la Escuela Superior Politécnica del Chimborazo – Ecuador. ORCID: <https://orcid.org/0000-0001-7668-3053> E-mail: fabian.allauca@espoch.edu.ec

****Docente en la Escuela Superior Politécnica del Chimborazo – Ecuador. ORCID: <https://orcid.org/0000-0002-0308-5412>. E-mail: nora.mejia@espoch.edu.ec

*****Investigador Doctoral Independiente. Candidato a Doctor en Ciencias Informáticas. Universidad Nacional de La Plata, Facultad de Ciencias Informáticas, La Plata, Argentina. ORCID: <https://orcid.org/0000-0001-5812-7766>. E-mail: juancarlos.santillanl@info.unlp.edu.ar

Estimation of Bootstrap Confidence Intervals of Proportions of Factors Associated with Burials Due to COVID-19

ABSTRACT

The objective of this work is to estimate the Bootstrap confidence intervals of proportions of factors associated with burials due to COVID-19, in the general cemetery of Riobamba - Ecuador, period March 2020 - April 2021. The Bootstrap method consists of carrying out re sampling with repetition, that is, obtaining samples through some random procedure that uses the original sample. In this way, vulnerable groups for burials due to COVID-19 have been identified, by sex, age and age according to sex. Regarding the vulnerability analysis of the groups, in terms of sex, the most vulnerable group is the male sex. In turn, within age intervals determined through life cycles, it was found that the most vulnerable group is known as older adults, which are people who are over 60 years old, for both the male and female sexes.

KEYWORDS: Statistics, Bootstrap, COVID-19, Ecuador.

Introducción

La presente investigación hace referencia a la estimación de intervalos de confianza bootstrap de proporciones de factores asociados a inhumaciones por causa de COVID-19, en el cementerio general de Riobamba- Ecuador, periodo marzo 2020 – abril 2021. bootstrap puede definirse como un modelo matemático que nos sirve para procesar datos y analizar los grupos vulnerables más relevantes para las inhumaciones en el cementerio general de Riobamba - Ecuador.

Existen escasos estudios que utilicen modelos estadísticos aplicados al reconocimiento de factores asociados a inhumaciones por causa de COVID-19, y más aún nula bibliografía de estudios que aplican la estimación de intervalos de confianza bootstrap de proporciones a dicha enfermedad; dentro de las aplicaciones de modelos estadísticos aplicados al COVID-19 encontramos los análisis de proyecciones de propagación de la enfermedad. Como ejemplo, los autores (Manrique et al, 2020) en su investigación titulada “Modelo SIR de la pandemia de Covid-19 en Colombia” concluyen que “los resultados de este estudio son relevantes y de interés, ya que los modelos logísticos y SIR

realizados a medida que avanza la epidemia son claves para evaluar la efectividad de las políticas y medidas adoptadas para contener el contagio de la enfermedad y garantizar una atención hospitalaria adecuada”

Por otra parte, tenemos el estudio realizado por (Cirera et al, 2021) titulado “Defunciones por COVID-19: no están todas las que son y no son todas las que están”, cuyo propósito es mejorar la producción de las estadísticas de las defunciones y de los sistemas de información sanitaria. Se afirma que es necesario un acuerdo institucional (INE, Organización Médica Colegial, Ministerio y Consejerías de Sanidad) que alinee recursos y garantías para el procesamiento telemático de las defunciones por causas.

Así también encontramos el estudio realizado por (Llaro Sánchez et al, 2020) titulado “Características clínico-epidemiológicas y análisis de sobrevida en fallecidos por COVID-19 atendidos en establecimientos de la Red Sabogal-Callao 2020”, en el cual los autores concluyen que los pacientes fallecidos fueron, principalmente, adultos mayores de sexo masculino con enfermedades preexistentes como hipertensión arterial y obesidad y clasificados con infección COVID-19 moderada a severa al ingreso hospitalario. El mayor tiempo de sobrevida se observó entre los que estuvieron en ventilación mecánica.

Otra investigación en la que se aplica la estadística al análisis de fallecidos es la realizada por (Sangrador et al, 2021) titulada “Impacto de la COVID-19 en la mortalidad de la comunidad autónoma de Castilla y León”, en la que concluyen que “aunque el aumento de la mortalidad observado no puede ser totalmente atribuido a la enfermedad, es la mejor estimación que tenemos del impacto real en muertes directamente o indirectamente relacionadas con ella. El número de muertes declaradas solo alcanza dos terceras partes del aumento de la mortalidad observado.”

En el presente trabajo de investigación nos planteamos estimar los intervalos de confianza BOOTSTRAP de proporciones de factores asociados a inhumaciones por causa de COVID-19, en el cementerio general de Riobamba - Ecuador, periodo marzo 2020 – abril 2021.

1. Métodos de re muestreo

En Estadística el re muestreo en su acepción más general consiste en generar muestras o muestrear elementos de manera repetitiva, los cuales se pueden dividir en métodos de re muestreo paramétricos y no paramétricos. El re muestreo paramétrico es aquel que genera

muestras a partir de la muestra con mejor ajuste al modelo. Éste es usado comúnmente cuando el conjunto de datos que se quiere muestrear contiene pocos elementos. El re muestreo no paramétrico es aquel en el que se asume que la mayoría de los elementos en el conjunto de datos proviene de la misma distribución, por lo que el conjunto es apto para realizar re muestreo con reemplazo, con la finalidad de preservar la función de densidad de probabilidad del conjunto de datos. El re muestreo paramétrico se aplicará en la presente investigación por lo que a continuación se mencionan algunos métodos para realizarlo.

Para llevar a cabo el re muestreo existe una variedad de métodos. El método de re muestreo que vaya a ser empleado depende de la tarea a realizar, entre las tareas más comunes para las que se usa el re muestreo se encuentran:

Estimación de la precisión de una muestra estadística (medianas, varianzas, percentiles) mediante el uso sistemático de subconjuntos de datos disponibles (jackknife) o seleccionando aleatoriamente con reemplazo subconjuntos de datos (bootstrap).

Validación de modelos mediante re muestreo aleatorio (bootstrap, validación cruzada)
Intercambiar elementos de subconjuntos de datos para pruebas de significancia estadística (pruebas de permutación, pruebas exactas o pruebas de aleatoriedad).

Los métodos de re muestreo empleados para la estimación de la precisión de una muestra estadística de manera general seleccionan conjuntos de datos de forma sistemática o aleatoria.

Algunos de estos métodos de re muestreo realizan tareas adicionales como estimación de un valor de desviación envuelto en el mismo proceso de re muestreo con el objetivo de mejorar en las estimaciones estadísticas. Los métodos de re muestreo más conocidos usados para la estimación de la precisión de una muestra estadística se presentan a continuación:

1.1. Jackknife

En algunos trabajos la manera de re muestrear de este método es denominada sub muestreo, las características principales son:

- (1) el tamaño de las muestras es más pequeño que el número de elementos en el conjunto de datos; y
- (2) el re muestreo es realizado sin reemplazo.

Acorde a las características anteriores, la idea básica detrás de Jackknife reside en volver a calcular sistemáticamente los estimadores estadísticos, omitiendo una o más observaciones a la vez del conjunto de datos. Posteriormente para cada nuevo subconjunto de datos se calcula una aproximación al estimador estadístico y entonces se obtiene el promedio de esos cálculos.

1.2. Bootstrap

El bootstrap es un procedimiento estadístico que sirve para aproximar la distribución en el muestreo (normalmente) de un estadístico. Para ello procede mediante re muestreo, es decir, obteniendo muestras mediante algún procedimiento aleatorio que utilice la muestra original.

Su ventaja principal es que no requiere hipótesis sobre el mecanismo generador de los datos. Sí las requiere, aunque suelen ser más relajadas, para obtener propiedades asintóticas del mismo. Por otra parte, su implementación en ordenador suele ser sencilla, en comparación con otros métodos. Su principal inconveniente es la necesidad de computación intensiva, debido a la fuerza bruta del método de Monte Carlo. Con la capacidad computacional actual, esta mayor carga computacional del bootstrap no suele ser un problema hoy en día. En raras ocasiones el bootstrap no necesita del uso de técnicas de Monte Carlo.

La metodología bootstrap debe su nombre y su formulación original a (Bradley Efron 1979). Constituye la línea más desarrollada, tanto desde el punto de vista teórico como aplicado, de una variedad de técnicas para la inferencia estadística denominadas genéricamente “métodos de re muestreo” (Simon, 1969), entre las que se encuentran la permutación estocástica, el jackknife (Quenouille, 1956; Tukey, 1958) y la validación cruzada (Mosier, 1951). Son básicamente técnicas de simulación que reutilizan los datos observados para constituir un universo del cual extraer repetidas muestras. El requerimiento de gran potencia computacional común a todas ellas ha llevado a denominarlas técnicas de “computación intensiva”.

Al construirse empíricamente la distribución del estimador sobre la base de todas las características de la distribución original de los datos, incluyendo aquellos factores considerados como contaminantes (colas pesadas, outliers, etc.), el enfoque bootstrap está especialmente indicado en los casos en que los datos no siguen una distribución normal;

hecho, que es común a la mayor parte de las medidas utilizadas habitualmente en las ciencias del comportamiento (Micceri, 1989). El procedimiento bootstrap es útil para la descripción de la distribución muestral de aquellos estimadores con propiedades muestrales desconocidas o difícilmente obtenibles por medios analíticos.

La idea subyacente al bootstrap es simple: Los datos muestrales son tratados como si constituyesen los datos de toda la población, es decir se utilizan como el universo del que se extraerán muestras con reemplazamiento. Para cada re muestreo se calculará el valor del estimador bootstrap que se utilizará para estimar la variabilidad muestral. Tal y como los estudios teóricos han demostrado, este enfoque proporciona una buena aproximación de la distribución de los estimadores (Diaconis y Efron, 1983; Efron, 1981) lo cual permitirá describir algunas de sus propiedades muestrales, así como el cálculo de intervalos de confianza y la realización de contrastes de hipótesis.

Sea $\theta(F)$ un parametro dependiente de una desconocida funcion de distribución F y, por otro lado, considérense las variables aleatorias X_1, X_2, \dots, X_n tales que,

$$X_1, X_2, \dots, X_n \text{ i.i.d } F$$

o sea, son variables aleatorias con función de distribución idéntica e independiente. Por otro lado, representaremos mediante el conjunto $\{x_1, x_2, \dots, x_n\}$ una muestra correspondiente a extracciones aleatorias sobre las referidas variables. Nótese que los valores correspondientes a la muestra nos permiten obtener la distribución empírica \hat{F}_n , que constituye la estimación no paramétrica de máxima verosimilitud de la función de distribución F . En apoyo de esta estimación, podemos referirnos al Teorema de Glivenko-Cantelli, que establece una convergencia casi segura, cuando $n \rightarrow \infty$ entre las distribuciones F y \hat{F}_n aunque debe notarse que es una convergencia asintótica.

$$\sup | \hat{F}_n(x) - F(x) | \rightarrow 0$$

Por tanto, es factible establecer $F = \hat{F}_n$, significando que la función de distribución de la variable aleatoria de interés se estima a partir de la distribución empírica, constituyendo el aspecto fundamental del denominado Bootstrap no paramétrico.

Bootstrap se utiliza para evaluar las mediciones de precisión definidas en términos de sesgo, varianza, intervalos de confianza, error de predicción o algún otro. Esta técnica permite la estimación de la distribución muestral de casi cualquier estimador estadístico utilizando métodos de re muestreo aleatorios. Bootstrap trata el conjunto de datos original

como un suplente para la población y lo muestrea repetidamente con reemplazo, calculando la estimación deseada cada vez, finalmente el valor estimado es la media de las estimaciones obtenidas.

Los métodos de re muestreo empleados en pruebas estadísticas de manera general seleccionan conjuntos de forma aleatoria, algunos de estos métodos de re muestreo realizan tareas adicionales como unir y separar conjuntos de datos provenientes de distintos grupos para ser útiles a ciertas pruebas estadísticas.

Algunas tareas o aplicaciones estadísticas que hacen uso o adaptan algún método de re muestreo se presentan a continuación:

-Calcular intervalos de confianza de un parámetro poblacional: Se generan muestras de las que se calcula el valor del parámetro y los valores que cubren el intervalo de confianza deseado.

-Calcular significancia estadística para la diferencia entre dos poblaciones: El método de bootstrap contrasta la hipótesis de que ambas muestras proceden de la misma población (distribución) mediante el estudio de las diferencias debidas al re muestreo aleatorio. Se aplica a estudios en los que no ha habido una asignación aleatoria a los grupos. Aquí se juntan las observaciones de ambas muestras, se emplea bootstrap sobre el conjunto de observaciones para generar una nueva muestra del mismo tamaño, se separan las observaciones de la muestra generada en dos grupos de igual tamaño a los originales y se calcula la diferencia del estadístico entre ambas. El proceso se repite múltiples veces generando así la distribución de las diferencias esperadas debido al re muestreo aleatorio. El valor de probabilidad de dos colas se calcula como la proporción de muestras en las que el valor absoluto de la diferencia calculada es mayor o igual al valor absoluto de la diferencia observada.

-Calcular intervalos de confianza para la diferencia entre dos poblaciones: Se considera como hipótesis nula que las observaciones que proceden de dos poblaciones distintas. Se emplea bootstrap con las observaciones de cada muestra (sin mezclarlas) para generar dos nuevas muestras independientes y se calcula la diferencia del estadístico. Este proceso se repite múltiples veces, generando así la distribución que se obtendría si se obtuviesen cada vez dos muestras, cada una de su respectiva población y se calcula la diferencia. Se asume que la distribución resultante está centrada en la verdadera diferencia entre las poblaciones.

2. Intervalos de confianza

En un proceso de inferencia estadística, la estimación puntual $\hat{\theta}$ del parámetro θ no presenta la medida de los componentes del error cuadrático medio: la precisión del estimador, que está asociado con la variabilidad o dispersión del estimador, y la exactitud de la estimación, que está asociado con el sesgo.

Los intervalos de confianza incorporan las deficiencias de la estimación puntual, donde los intervalos de confianza Bootstrap los podemos obtener utilizando diferentes criterios o estrategias, que para nuestros requerimientos, como sugiere Efron y Tibshirani (1993), utilizaremos el método percentil y el método pivotal, cuya forma básica de un intervalo de confianza Bootstrap.

Una estimación más completa de los parámetros mencionados, se denomina estimación por intervalo, ella consiste en ofrecer no ya un número como en la estimación puntual, sino un intervalo, acerca del cual se tiene cierto grado de certidumbre (o se deposita cierta confianza) que contenga al parámetro.

Así, en lugar de decir que el promedio con que egresan quienes terminan una carrera universitaria “debe ser cercano a 6.50”, se construirá un intervalo, que dirá, por ejemplo, “hay una confianza del 95% en que el intervalo [6.10; 6.90] contiene al promedio con que se termina esa carrera”.

De manera equivalente, en lugar de “entre quienes egresan hay alrededor del 70% de mujeres”, se afirmará algo como “con una confianza del 95%, el intervalo [68; 72] % contiene a la proporción de mujeres sobre el total quienes egresan”. O que “con una confianza del 95%, el ingreso salarial promedio de los trabajadores del aglomerado Gran Córdoba está entre 16710.516710.5 y 17010.517010.5. Y también que en la población mayor de 18 años residente en países de América Latina, hay una confianza de 90% que la proporción de quienes creen que se puede confiar en la mayoría de la gente, esté en el intervalo [14.20, 14.49].

Vemos entonces que esta forma de estimar ofrece dos números, los límites de un intervalo, del que esperamos contenga al parámetro que estimamos. Decimos “esperamos que se contenga” porque no hay certeza absoluta de que se encuentre allí, hay una confianza que en estos ejemplos hemos fijado en el 95% o en 90%, y veremos que puede elegirse.

El Bootstrap fue introducido por Efron (1979) y básicamente es un método de simulación computacional intensivo, que consiste en la aplicación de un proceso de re muestreo para estimar la distribución muestral de un estimador basándose tan solo en una muestra disponible, bajo el supuesto de que ésta es un buen representante de la población (Efron, 1979). Se ha demostrado que el método funciona bien en muchas situaciones y ha sido aceptado por la comunidad estadística, incluso se considera que, en algunos casos, es mejor que la tradicional aproximación normal.

Básicamente, el proceso para construir un intervalo de confianza Bootstrap es:

1. De la única muestra disponible se toman datos aleatoriamente con reposición para generar una re muestra del mismo tamaño que la original. Todos los datos de la muestra tienen la misma probabilidad de ser elegidos y un dato puede aparecer varias veces en una re-muestra.
2. El proceso descrito en el paso 1 se debe repetir una gran cantidad de veces, calculando el valor del estadístico de interés en cada una de las re-muestras.
3. Se construye la distribución de los valores del estadístico calculados en el paso 2, ésta se conoce como la distribución Bootstrap.
4. Si se quiere construir un intervalo con un 95% de confianza se deben calcular los percentiles 2,5 y 97,5 de la distribución Bootstrap.

3. Estimación de la proporción

Cuando trabajamos con variables cualitativas (nominales u ordinales) no es posible calcular la media ni la desviación estándar sino solo considerar la proporción de casos que hay en una categoría que elegimos. La proporción es la frecuencia relativa de la categoría que se elige, la cantidad de casos en esa categoría dividida para el tamaño de la muestra. Cuando se trata de variables con solo dos categorías (dicotómicas) puede elegirse cualquiera de ellas. Por ejemplo, si es el resultado de un examen y las categorías son aprobado – no aprobado, podemos interesarnos por la proporción de cualquiera de ellas, ya que la otra es el complemento (lo que le falta para llegar a uno). Si una es 0.70 (70%), la otra no puede sino ser 0.30 (30%). Es diferente si la variable tiene más de dos categorías, por ejemplo si se trata de la intención de voto para las elecciones presidenciales. Allí es usual que haya más de dos partidos que aspiran a la presidencia, por lo que, conocer la

proporción de uno de ellos no nos dice mucho sobre la de cada uno de los otros: si hay cinco partidos y uno se lleva el 40%, solo sabemos que el 60% restante se reparte entre los otros cuatro, pero no sabemos cuánto le corresponde a cada uno. A estos casos los trataremos como si fueran dicotómicos: una categoría será el partido que nos interesa y la otra categoría estará formada por todos los demás. Así, si un partido tiene una proporción de su favor, solo nos interesa que tiene una proporción de 0.60 que no está a su favor y no nos preocupamos por saber cómo se reparte ese 60% en los demás partidos. Tratamos una categoría frente a todas las demás. De este modo es que puede definirse la proporción de personas que usa anticonceptivos orales, frente a quienes usan todos los demás métodos; o la proporción de quienes promocionaron una asignatura frente a regulares y libres; o la proporción de quienes nacieron en Argentina entre el estudiantado de origen extranjero que hay en España, sin interesarnos por el modo en que se distribuye la proporción entre las demás nacionalidades. Lo que hacemos con este procedimiento es simplemente llamar la atención sobre una categoría y confrontarla con el resto indiscriminado. La categoría elegida o categoría de referencia, se llama éxito, su frecuencia absoluta, cantidad de éxitos y su frecuencia relativa proporción de éxitos.

Por este procedimiento trataremos siempre con dos grupos, uno formado por los casos que son de nuestro interés y el otro por los demás casos.

3.1. Intervalos de confianza de la proporción

Una variable dicotómica, se puede resumir como el hecho de padecer o no cierto acontecimiento adverso (AA), definido mediante la proporción P de pacientes que lo han experimentado. La proporción P de la muestra estima la probabilidad poblacional de que un nuevo paciente de las mismas características presente dicho AA.

Si el tamaño muestral lo justifica, es cómodo recurrir a la aproximación a la Normal de la distribución del estimador, se aceptan como condiciones de aplicación de la aproximación Normal que el tamaño muestral sea grande y las probabilidades de p y q ; donde $q = (1-p)$ no extremas:

$$p^*n \geq 5 \text{ y } (1-p)^*n \geq 5 \text{ o } q^*n \geq 5 \quad (1)$$

El objetivo es construir un I.C. para la proporción de elementos p de una población que poseen determinada característica de interés, a partir de una M.A.S. (Muestra aleatoria simple) de elementos de la población.

Dada la muestra:

$$\vec{X} = (X_1, X_2, \dots, X_n)$$

Donde \vec{X} es una variable cualitativa, X_i tiene un valor de escala e_1, e_2, \dots, e_k

Para cada $i = 1, 2, \dots, n$ arrojar

$U_i \sim U(0,1)$ y hacer $X_i^* = X_{[nU_i]+1}$

Obtener

$$p_j^* = \frac{1}{n} \sum Y_i^* \quad i = 1, 2, \dots, n ; j = 1, 2, \dots, k \quad (2)$$

$$Y_i^* = \begin{cases} 1 & \text{si } X_i^* = e_j \\ 0 & \text{si } X_i^* \neq e_j \end{cases}$$

Calcular el estadístico bootstrap:

$$R_j^* = \frac{(p_j^* - p_j)}{\sqrt{\frac{p_j^*(1-p_j^*)}{n}}} \quad j = 1, 2, \dots, k \quad (3)$$

Repetir B veces los pasos 2, 3 y 4 para obtener las réplicas bootstrap

$$R_j^{*(1)}, \dots, R_j^{*(B)} \quad j = 1, 2, \dots, k$$

Ordenar de forma creciente los valores del estadístico bootstrap:

$$R_{j(b)}^*, \quad j = 1, 2, \dots, k ; b = 1, 2, \dots, B$$

Calcular los puntos críticos, inferior y superior del nivel de significancia α :

$$\begin{aligned} \text{p.c.inf} &= \{R_{j(b)}^*\}_{[B \frac{\alpha}{2}]} \\ \text{p.c.sup} &= \{R_{j(b)}^*\}_{[B (1-\frac{\alpha}{2})]} \end{aligned} \quad (4)$$

Calcular los límites inferior y superior del intervalo de confianza de proporciones de escala nominal u ordinal de una variable cualitativa, con el nivel de significancia α :

$$\begin{aligned} \text{lim.infj} &= p.c.\text{inf} \sqrt{\frac{p_j(1-p_j)}{n}} + p_j \\ \text{lim.supj} &= p.c.\text{sup} \sqrt{\frac{p_j(1-p_j)}{n}} + p_j \end{aligned} \quad (5)$$

Donde p_j es la proporción muestral de la j -ésima escala de la variable cualitativa \vec{X} , $[x]$ función parte entera de x , $U(0,1)$ es la distribución uniforme en el intervalo $(0, 1)$. Por tanto el intervalo de confianza correspondiente a la proporción de la escala j -ésima esta dada por:

$$IC_j = (\text{lim.infj}, \text{lim.supj}) \text{ para } j = 1, 2, \dots, k$$

3.2. La calidad de las estimaciones por intervalo

Intuitivamente, una estimación es de mejor calidad si es “ajustada”, es decir si el intervalo es pequeño. Por ejemplo, si estimamos la edad de una persona entre 28 y 30 años, tenemos una estimación de mejor calidad que si decimos que tiene entre 20 y 40 años. Eso es porque el primer intervalo es más pequeño, los límites están más cerca. La primera estimación nos da más información que la segunda, porque delimita el valor al que estima entre números más cercanos.

En las estimaciones que hemos hecho hasta aquí, de la media y de la proporción y desde él sumamos y restamos la misma cantidad para obtener los límites del intervalo.

Esa cantidad que sumamos y restamos determina la amplitud del intervalo: cuanto más grande sea, tanto mayor será el intervalo, es decir, tanto mayor será la distancia entre los límites inferior y superior. Esa cantidad se denomina error de estimación. Las estimaciones están siempre acompañadas de un error, es un componente intrínseco al proceso. No es error en el sentido de equivocación o de falla, sino de imprecisión, una imprecisión que no puede evitarse, que no puede hacerse igual a cero. Se calcula como la distancia que hay desde el centro del intervalo hasta cualquiera de los límites. En el ejemplo anterior, sobre la estimación intuitiva de la edad de alguien, el centro del primer intervalo es 29, por lo que el error es 1 año, por eso se puede también escribir como 29 ± 1 . El segundo intervalo de este ejemplo tiene centro en 30 y el error es de 10 años, lo escribimos 30 ± 10 . Independientemente que el centro de los intervalos difiera levemente,

este segundo intervalo tiene un mayor error de estimación. Esto es equivalente a decir que tiene menos precisión.

Se llama error de estimación a la distancia que hay entre el estimador puntual y cualquiera de los límites del intervalo. Cuanto mayor es el error de estimación menor es su precisión.

En la estimación del promedio con que se termina una carrera universitaria, escribimos 6.50 ± 0.08 al estimar al 95% de confianza y 6.50 ± 0.10 cuando la confianza se pasó al 99%. Allí estábamos escribiendo el intervalo como el estimador más/menos el error de estimación. En el primer caso el error de estimación es de 0.08 y en el segundo de 0.10, por eso decimos que la primera estimación es más precisa, tiene menos error. Con la estimación de los ingresos salariales sucedió lo mismo, al 95% el error 674.80 fue y al 99% subió a 884.80.

Del mismo modo, al estimar, con una confianza del 95%, la proporción de quienes aprobaron el parcial escribimos 0.920 ± 0.032 , el error de estimación es en este caso de 0.032 (ó 3.2%). La estimación de la proporción de gente que cree que se puede confiar en los demás, hecha con el mismo nivel de confianza, arrojó un error de estimación de 0.0056, 0.56%.

4. Detección de elementos atípicos

Un elemento atípico (outlier en inglés) se define como un elemento cuyo valor cae fuera de los límites que encierran la mayoría de los valores correspondientes de un conjunto de datos. Los elementos que caen dentro de los valores que encierran la mayoría de los elementos son conocidos como elementos no atípicos o elementos normales (inliers en inglés).

La detección de elementos atípicos es un desafío debido a las siguientes razones:

En la mayoría de las ocasiones el límite entre los elementos no atípicos y atípicos es impreciso. Por lo tanto, un elemento atípico que se encuentra cerca del límite puede ser atípico o no. Algunos escenarios hacen que los elementos atípicos aparezcan intrínsecos, por lo que definir un comportamiento normal es difícil. En algunas ocasiones un elemento parece ser un valor atípico cuando realmente no lo es.

En conjuntos de datos altamente correlacionados, para realizar una buena detección de elementos atípicos se tienen que tomar consideraciones adicionales, lo cual vuelve este proceso más complejo.

En muchos dominios el comportamiento normal sigue evolucionando y una noción actual de comportamiento normal podría no ser suficientemente representativa en el futuro. Debido a estas dificultades, el problema de detección de elementos atípicos, en su forma más general, no es fácil de resolver. De hecho, la mayoría de los métodos de detección de elementos atípicos existentes resuelven un problema específico.

Un aspecto importante de un método de detección de elementos atípicos es el tipo de elementos atípicos que se pueden detectar. Los elementos atípicos se pueden clasificar en las siguientes tres categorías:

1. Elemento atípico puntual: Es un elemento individual que es anómalo con respecto al resto de los elementos. Este es el tipo más simple de elemento atípico y es el foco de la mayoría de las investigaciones sobre detección de elementos atípicos.

2.Elemento atípico condicional. Es un elemento que es anómalo en un contexto específico (pero no de otro modo). También se le denomina elemento atípico contextual. La noción de un contexto es inducida por la estructura del conjunto de datos y tiene que ser especificado como parte del problema.

3.Elemento atípico colectivo. Es una colección de elementos relacionados anómalos con respecto al conjunto de datos completo. Los elementos individuales en un elemento atípico colectivo pueden no ser elementos atípicos por sí mismos, pero su ocurrencia juntos como una colección es anómala.

5. Metodología

El análisis estadístico inferencial, se realizará con el lenguaje de programación R versión 4.2.1, mediante su entorno de desarrollo integrado RStudio versión 2022.07.1.

La presente investigación utiliza un muestreo no probabilístico, por conveniencia; la muestra se compone de datos secundarios, es decir datos recolectados y almacenados por el cementerio general de Riobamba, los cuales fueron facilitados para la presente investigación gracias a la colaboración del Gobierno Autónomo Descentralizado Municipal del Cantón Riobamba, mediante carta de auspicio emitida.

Estos datos contienen las siguientes características, indicando que los mismos son determinados de acuerdo al acta de defunción emitida por la entidad o profesional autorizado.

Tabla 1. Datos Inhumaciones Cementerio General de Riobamba.

#	Nombre del fallecido (a)	SEXO	Cédula de identidad del fallecido	Edad	Fecha de fallecimiento	Fecha de inhumación	Lugar fallecimiento	Causa básica de la muerte	Causa
00	Paltan Chacha, Brigida Rosario	Femenino	0601088735	65	02/04/2020	04/04/2020	Riobamba	Hipertensión secundaria no especificada	Otra causa
000	AucanCela Pinduisaca Alberto	Masculino	0601294192	62	01/04/2020	03/04/2020	Riobamba	Insuficiencia respiratoria aguda - por covid	Covid

Realizado por: Santillán et al, 2024

Previo al análisis estadístico, se realizará un Análisis Exploratorio de Datos, tanto para filtrar individuos según los criterios requeridos en esta investigación, como para identificar datos atípicos, faltantes o datos con algún tipo de error en su valor o formato.

La muestra son los datos que tenemos de inhumaciones, tanto como sexo de la persona, edad, si el fallecimiento fue por COVID u otra causa ajena al COVID.

Lo primero que realizamos es depurar la muestra de datos que no estén completos, como edad, o si no se conoce porque falleció, debido a que el sexo de la persona se lo puede colocar por el nombre de los datos obtenidos.

Se describirán las variables categóricas: “CAUSA”, “SEXO” y la variable numérica edad de los fallecidos clasificados de acuerdo al “CICLO DE VIDA” mediante proporciones, porcentajes.

Finalmente se estimarán las proporciones de los factores asociados a inhumaciones por causa de COVID-19, mediante la metodología de intervalos de confianza Bootstrap, de tal forma que se identifiquen sectores vulnerables ante la pandemia.

5.1. Algoritmo para aplicar el Método BOOTSTRAP para calcular intervalos de confianza para las proporciones de escala nominal u ordinal de una variable cualitativa

a. Dada la muestra:

$$\vec{X} = (X_1, X_1, \dots, X_n)$$

Donde \vec{X} es una variable cualitativa, X_i tiene un valor de escala e_1, e_2, \dots, e_k

b. Para cada $i = 1, 2, \dots, n$ arrojar

$$U_i \sim U(0,1) \text{ y hacer } X_i^* = X_{[nU_i]+1}$$

Para cada variable cualitativa se va a dar el valor de escala, de acuerdo a lo deseado determinar y las características necesarias:

1. PROPORCIONES INHUMACIONES POR COVID – OTRAS CAUSAS

- Inhumaciones por COVID.
- Inhumaciones por OTRA CAUSA

2. PROPORCIONES MASCULINO – FEMENINO INHUMACIONES POR COVID

De la muestra seleccionamos únicamente los datos de inhumaciones por COVID-19, de los cuales damos la escala de la siguiente manera:

- Masculino
- Femenino

3. PROPORCIONES CICLOS DE VIDA - INHUMACIONES POR COVID

Para los datos de inhumaciones por COVID-19, determinamos los rangos de edades de acuerdo a las etapas de la vida:

- INFANCIA 0 – 6 años.
- NIÑEZ 7 – 12 años.
- ADOLESCENCIA 13 – 20 años.
- JUVENTUD 21 – 25 años.
- ADULTOS 26 – 60 años.
- ADULTO MAYOR desde 61 años.

4. PROPORCIONES CICLOS DE VIDA – MASCULINO INHUMACIONES POR COVID

De la misma manera que en la proporción de CICLOS DE VIDA - INHUMACIONES POR COVID, aplicamos la misma metodología en su totalidad para los datos de Inhumaciones de sexo MASCULINO de acuerdo a sus edades.

5. PROPORCIONES CICLOS DE VIDA – FEMENINO INHUMACIONES POR COVID

De la misma manera que en la proporción de CICLOS DE VIDA - INHUMACIONES POR COVID, aplicamos la misma metodología en su totalidad para los datos de Inhumaciones de sexo FEMENINO de acuerdo a sus edades.

Para cada una de las proporciones en cada caso aplicamos las siguientes ecuaciones, las cuales nos indican si se pueden aplicar el método empleado de acuerdo al número de datos de cada una de ellas:

$$p \cdot n \geq 5 \text{ y } (1-p) \cdot n \geq 5 \text{ o } q \cdot n \geq 5 \quad (6)$$

Donde p y q son las proporciones de cada una de las variables en análisis, si no cumple con estas condiciones, se pueden considerar como datos atípicos de acuerdo al número de datos, con esto se verifica que mediante el re muestreo no se puede llegar a la aproximación a la Normal de la distribución del estimador estudiado.

c. Obtener

$$p_j^* = \frac{1}{n} \sum Y_i^* \quad i = 1, 2, \dots, n ; j = 1, 2, \dots, k \quad (7)$$

$$Y_i^* = \begin{cases} 1 & \text{si } X_i^* = e_j \\ 0 & \text{si } X_i^* \neq e_j \end{cases}$$

Calculamos el p_j^* que es el valor de la proporción para cada una de las características a encontrarse, en el primer cálculo si utilizamos la muestra original tendremos que $p_j^* = p_j$, donde p_j se convierte el en dato de proporción observado.

d. Calcular el estadístico bootstrap:

$$R_j^* = \frac{(p_j^* - p_j)}{\sqrt{\frac{p_j^*(1-p_j^*)}{n}}} \quad j = 1, 2, \dots, k \quad (8)$$

Con el dato obtenido en el paso anterior obtenemos el R_j^* aplicando la ecuación para cada una de las escalas dadas, como se calculó en el paso anterior, en el primer cálculo si utilizamos la muestra original tendremos que $r_j^* = r_j$, donde r_j se convierte estadístico bootstrap observado, teniendo un valor igual a 0.

e. Repetir B veces los pasos b., c. y d. para obtener las réplicas bootstrap

$$R_j^{*(1)}, \dots, R_j^{*(B)} \quad j = 1, 2, \dots, k$$

En este paso realizamos el re muestreo que consiste en realizar B muestras con los datos de cada una de las características o análisis que realizamos en el paso b., en nuestro caso para cada una realizamos un re muestreo de 100000 muestras aparte de los datos originales entregados por el Cementerio General de la ciudad de Riobamba, con los cuales de cada uno calculamos p_j^* y R_j^* .

f. Ordenar de forma creciente los valores del estadístico bootstrap:

$$R_{j(b)}^*, \quad j = 1, 2, \dots, k ; \quad b = 1, 2, \dots, B$$

Teniendo de acuerdo al planteamiento realizado 100000 valores ordenados de forma creciente.

g. Calcular los puntos críticos, inferior y superior del nivel de significancia α :

$$\begin{aligned} \text{p.c.inf} &= \{R_{j(b)}^*\}_{[B \frac{\alpha}{2}]} \\ \text{p.c.sup} &= \{R_{j(b)}^*\}_{[B (1 - \frac{\alpha}{2})]} \end{aligned} \quad (9)$$

Para $\alpha=0.05$ tendremos de los 100000 valores de $R_{j(b)}^*$, los puntos críticos ubicados en las siguientes posiciones:

$$p.c.inf = \{R_{j(b)}^*\}_{[B \frac{\alpha}{2}]} \quad \text{el valor } R_{j(b)}^* \text{ ubicado en la posición } 2500=R_{j(2500)}^*$$

$$p.c.sup = \{R_{j(b)}^*\}_{[B (1-\frac{\alpha}{2})]} \quad \text{el valor } R_{j(b)}^* \text{ ubicado en la posición } 97500=R_{j(97500)}^*$$

h. Calcular los límites inferior y superior del intervalo de confianza de proporciones de escala nominal u ordinal de una variable cualitativa, con el nivel de significancia α :

$$\begin{aligned} \lim.inf_j &= p.c.inf \sqrt{\frac{p_j (1-p_j)}{n}} + p_j \\ \lim.sup_j &= p.c.sup \sqrt{\frac{p_j (1-p_j)}{n}} + p_j \end{aligned} \quad (10)$$

Aplicando las ecuaciones calculamos los límites del intervalo de confianza de proporciones de escala nominal u ordinal con los cuales podemos determinar su influencia o peso dentro del análisis.

Donde p_j es la proporción muestral de la j -ésima escala de la variable cualitativa \vec{X} , $[x]$ función parte entera de x , $U(0,1)$ es la distribución uniforme en el intervalo $(0, 1)$. Por tanto, el intervalo de confianza correspondiente a la proporción de la escala j -ésima está dada por:

$$IC_j = (\lim.inf_j, \lim.sup_j) \text{ para } j = 1, 2, \dots, k$$

5.2. Función en R como aporte para determinar los Intervalos de Confianza Bootstrap para proporciones

Como aporte para el presente método, facilitando su aplicación, se creó una función en el programa estadístico R, con la cual se aplicará el algoritmo y metodología para

encontrar los Intervalos de confianza Bootstrap para proporciones, como objetivo de esta investigación.

De esta manera se puede aplicar a cada uno de las variables cualitativas analizadas para determinar los Intervalos de confianza Bootstrap para proporciones como son los casos de:

Covid con Otras causas.

Sexo masculino o femenino dentro de inhumaciones por Covid.

Intervalos de edades o ciclos de vida dentro de sexo masculino para Covid

Intervalos de edades o ciclos de vida dentro de sexo femenino para Covid.

6. Resultados y discusión

Para poder realizar los cálculos de los Intervalos de confianza BOOTSTRAP debemos considerar los datos necesarios para cada una de las variables cualitativas y analizar los mismos de acuerdo a la ecuación 1, la cual nos indica si cumple con el mínimo de datos para poder aplicar el método de re muestreo BOOTSTRAP.

6.1. Proporciones inhumaciones por COVID – Otras causas

$$n=1129$$

$$COVID= 419$$

$$Otras causas=710$$

$$p = \frac{COVID}{n}$$

$$p = \frac{419}{1129}$$

$$p * n \geq 5$$

$$\frac{419}{1129} * 1129$$

$$\geq 5$$

$$419 \geq 5$$

CUMPLE

$$q = \frac{Otras causas}{n}$$

$$q = \frac{710}{1129}$$

$$q * n \geq 5$$

$$\frac{710}{1129} * 1129$$

$$\geq 5$$

$$710 \geq 5$$

CUMPLE

Los dos valores de la condición cumplen, con esto se puede aplicar el método BOOTSTRAP para proporciones, por lo cual se puede aplicar el algoritmo y la función dentro del programa estadístico R, con el cual se determinan los límites superior e inferior para determinar los Intervalos de confianza, para los que se aplicaran un nivel de significancia $\alpha=0.05$, teniendo las siguientes gráficas y datos de los intervalos generados para cada una de las variables cualitativas analizadas.

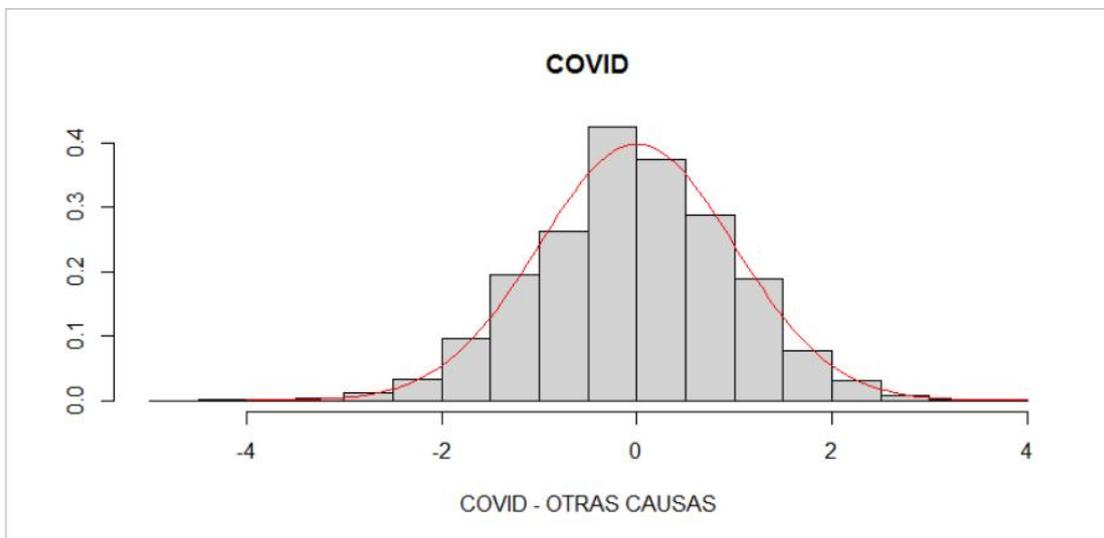


Figura 1. Proporciones Inhumaciones por COVID-19.

Realizado por: Santillán et al 2024.

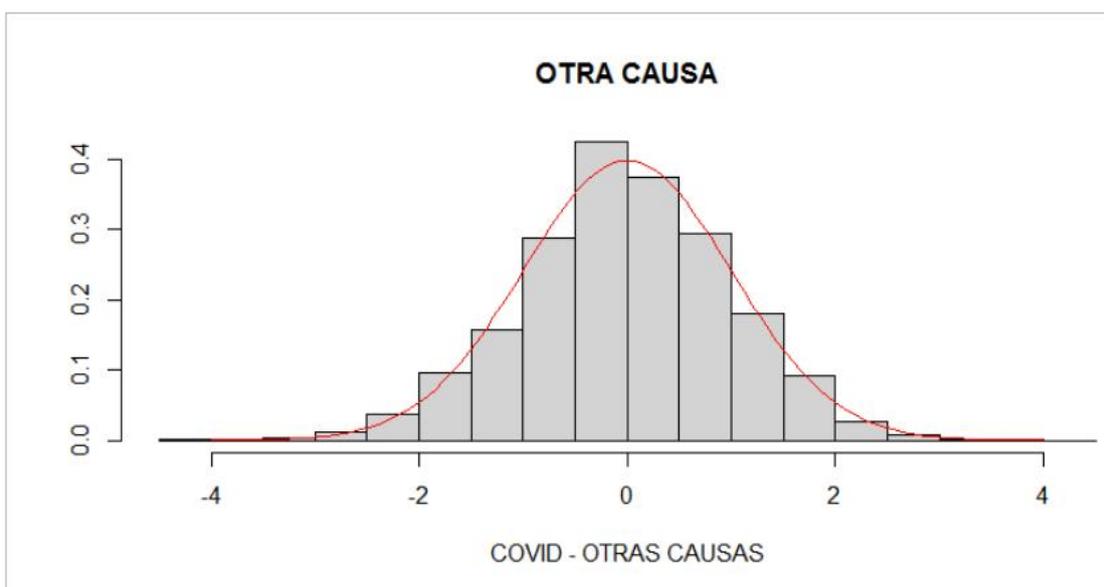


Figura 2. Proporciones Inhumaciones por OTRAS CAUSAS.

Realizado por: Santillán et al 2024.

Intervalos de confianza de proporciones con nivel de significancia $\alpha=0.05$

IC ($\alpha=0.05$)= [34.16; 39.84] % a causa de COVID-19.

IC ($\alpha=0.05$)= [59.98; 65.66] % por otras causas.

El resultado indica que hay una confianza del 95% que el intervalo [34.16; 39.84] % contenga a la proporción de Inhumaciones a causa de COVID-19, y [59.98; 65.66] % contenga a la proporción de Inhumaciones por otras causas.

Tabla 2. Intervalos de confianza COVID – Otra causa.

	COVID	OTRA CAUSA
2.5%	0.3416327	0.5998703
97.5%	0.3984445	0.6566009

Realizado por: Santillán et al 2024

6.2. Proporciones masculino – femenino inhumaciones por COVID

n=418

COVID MASCULINO = 283

COVID FEMENINO = 135

$$p = \frac{\text{COVID MASC.}}{n}$$

$$p = \frac{283}{418}$$

$$p * n \geq 5$$

$$\frac{283}{418} * 418 \geq 5$$

$$283 \geq 5$$

CUMPLE

$$q = \frac{\text{COVID FEMEN.}}{n}$$

$$q = \frac{135}{418}$$

$$q * n \geq 5$$

$$\frac{135}{418} * 418 \geq 5$$

$$135 \geq 5$$

CUMPLE

Los dos valores de la condición cumplen, con esto se puede aplicar el método BOOTSTRAP para proporciones, por lo cual se puede aplicar el algoritmo y la función dentro del programa estadístico R, con el cual se determinan los límites superior e inferior para determinar los Intervalos de confianza, para los que se aplicaran un nivel de significancia $\alpha=0.05$, teniendo las siguientes gráficas y datos de los intervalos generados para cada una de las variables cualitativas analizadas.

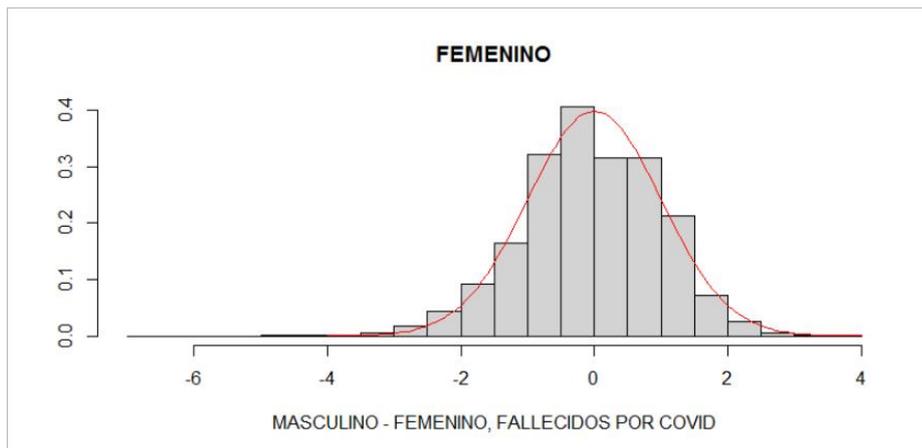


Figura 3. Proporciones Inhumaciones por COVID-19 - Femenino.
 Realizado por: Santillán et al 2024.

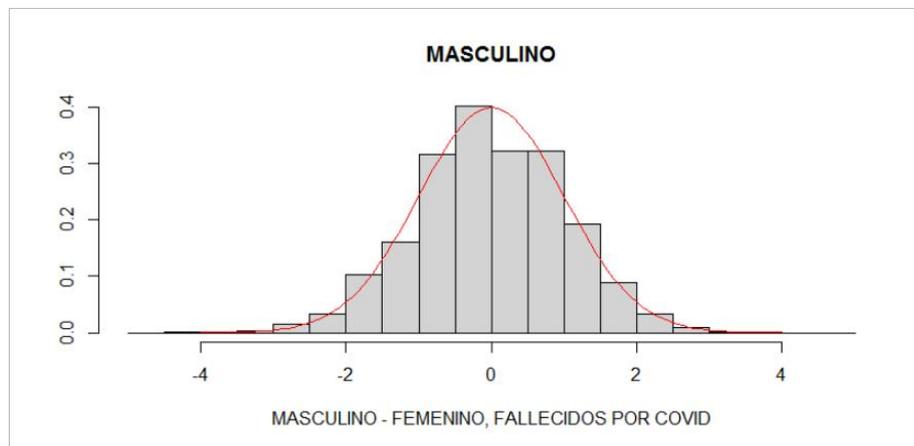


Figura 4. Proporciones Inhumaciones por COVID-19 - Masculino.
 Realizado por: Santillán et al 2024.

Tabla 3. Intervalos de confianza COVID – Masculino - Femenino.

	FEMENINO	MASCULINO
2.5%	0.2732767	0.6308608
97.5%	0.3646524	0.7217056

Realizado por: Santillán et al 2024.

Intervalos de confianza de proporciones con nivel de significancia $\alpha=0.05$

IC ($\alpha=0.05$)= [27.33; 36.47] % inhumaciones sexo FEMENINO.

IC ($\alpha=0.05$)= [63.09; 72.17] % inhumaciones sexo MASCULINO.

El resultado dice que hay una confianza del 95% que el intervalo [27.33; 36.47] % contenga a la proporción de Inhumaciones sexo FEMENINO, y [63.09; 72.17] % contenga a la proporción de Inhumaciones sexo MASCULINO.

6.3. Proporciones ciclos de vida - inhumaciones por COVID

$n=418$

A MAYOR = 283

ADULTO = 131

JUVENTUD = 3

INFANCIA = 1

NIÑEZ = 0

ADOLESCENCIA = 0

$$p = \frac{A \text{ MAYOR}}{n}$$

$$q = \frac{ADULTO}{n}$$

$$q_1 = \frac{JUVENTUD}{n}$$

$$p = \frac{283}{418}$$

$$q = \frac{131}{418}$$

$$q_1 = \frac{3}{418}$$

$$p * n \geq 5$$

$$q * n \geq 5$$

$$q_1 * n \geq 5$$

$$\frac{283}{418} * 418 \geq 5$$

$$\frac{131}{418} * 418 \geq 5$$

$$\frac{3}{418} * 418 \geq 5$$

$$283 \geq 5$$

CUMPLE

$$131 \geq 5$$

CUMPLE

$$3 \geq 5$$

NO
CUMPLE

$$q_2 = \frac{INFANCIA}{n}$$

$$q_3 = \frac{NIÑEZ}{n}$$

$$q_4 = \frac{ADOLESCENCIA}{n}$$

$$q_2 = \frac{1}{418}$$

$$q_3 = \frac{0}{418}$$

$$q_4 = \frac{0}{418}$$

$q_2 * n \geq 5$		$q_3 * n \geq 5$		$q_4 * n \geq 5$	
$\frac{1}{418} * 418 \geq 5$		$\frac{0}{418} * 418 \geq 5$		$\frac{0}{418} * 418 \geq 5$	
$1 \geq 5$	NO CUMPLE	$283 \geq 5$	NO CUMPLE	$0 \geq 5$	NO CUMPLE

Los dos valores para A Mayor y Adulto cumplen la condición, con esto se puede aplicar el método BOOTSTRAP para proporciones, por lo cual se puede aplicar el algoritmo y la función dentro del programa estadístico R, con el cual se determinan los límites superior e inferior para determinar los Intervalos de confianza, para los que se aplicaran un nivel de significancia $\alpha=0.05$, teniendo las siguientes gráficas y datos de los intervalos generados para cada una de las variables cualitativas analizadas; para los demás valores debido a que son menores que la condición no cumplen, debido a eso no se los puede aplicar el método, se analizará de manera semejante para visualizar los efectos o desventajas del método.

Con esto se puede tomar los valores de las variables que no cumplen con la condición como datos atípicos.

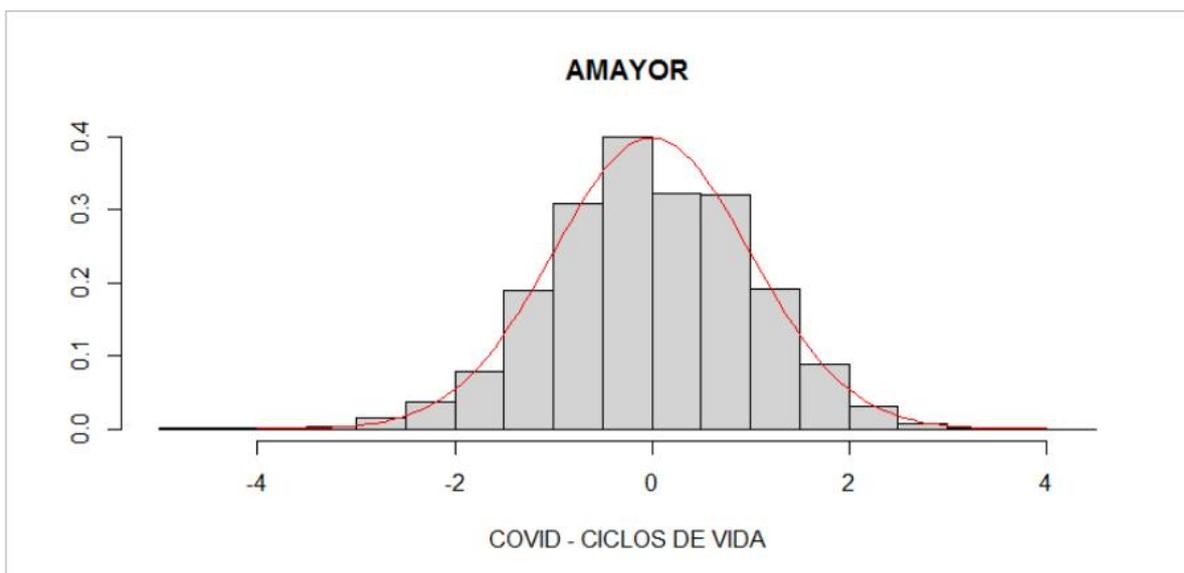


Figura 5. Proporciones Inhumaciones por COVID-19 – CICLOS DE VIDA – Adulto Mayor.

Realizado por: Santillán et al, 2024.

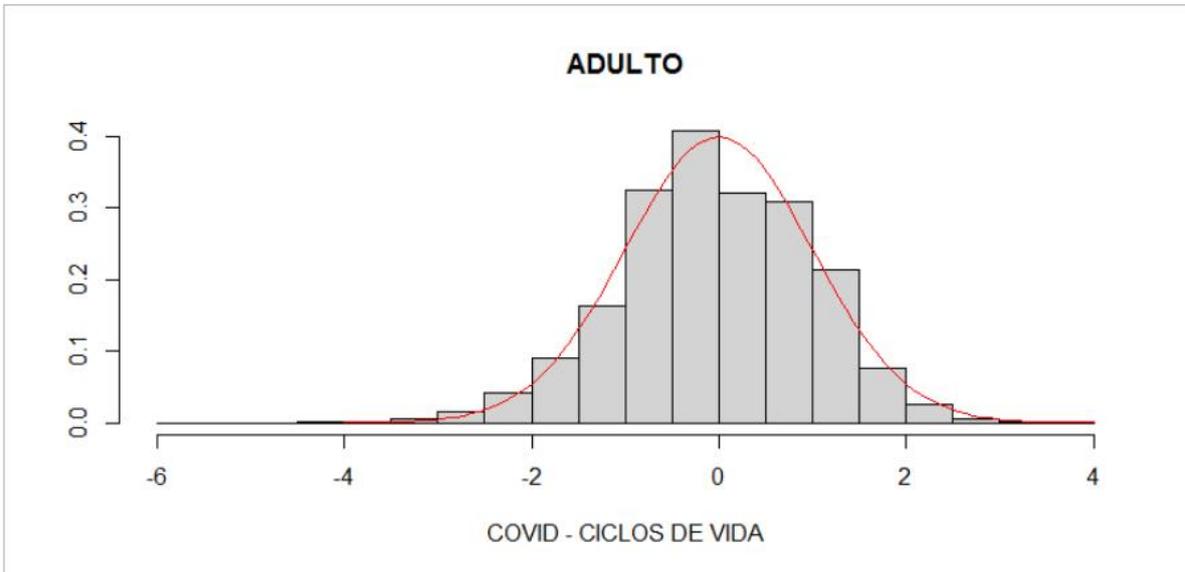


Figura 6. Proporciones Inhumaciones por COVID-19 – CICLOS DE VIDA – Adulto.
Realizado por: Santillán et al, 2024.

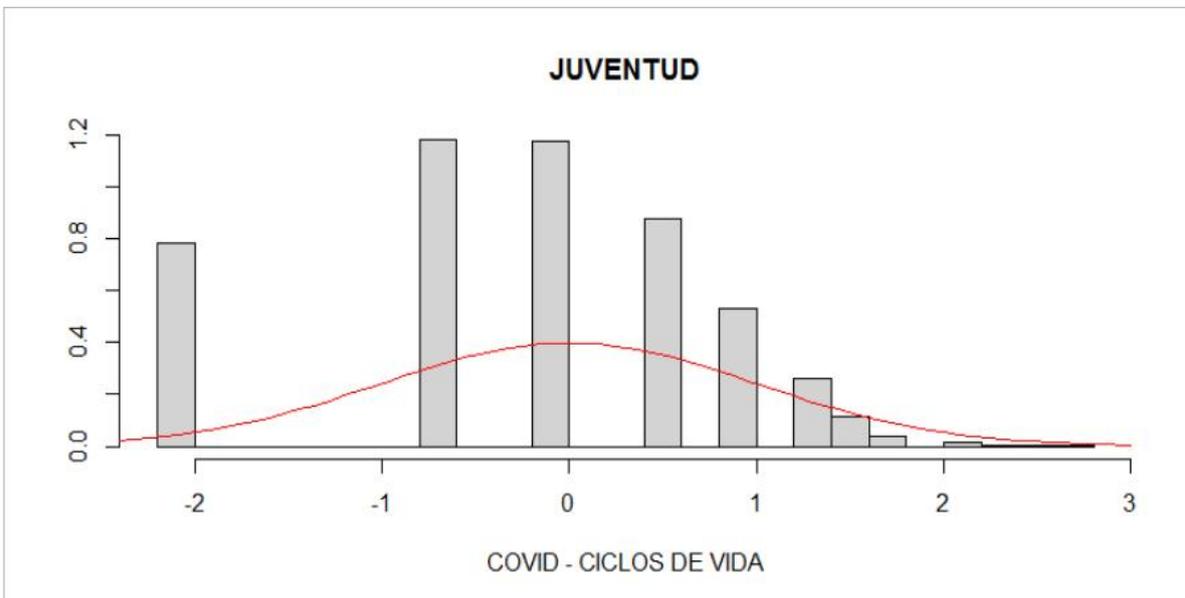


Figura 7. Proporciones Inhumaciones por COVID-19 – CICLOS DE VIDA – Juventud.
Realizado por: Santillán et al, 2024.

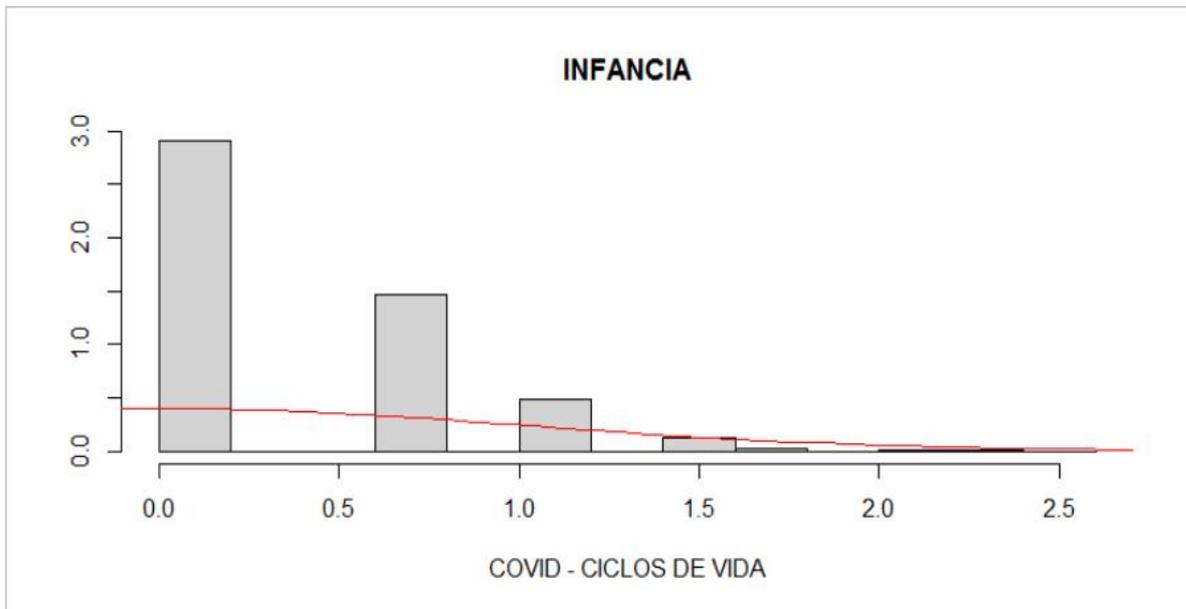


Figura 8. Proporciones Inhumaciones por COVID-19 – CICLOS DE VIDA – Infancia.
Realizado por: Santillán et al 2024.

Tabla 4. Intervalos de confianza COVID – Ciclos de vida

	AMAYOR	ADULTO	JUVENTUD	INFANCIA
2.5%	0.6284682	0.2688091	-Inf	-Inf
97.5%	0.7193141	0.3574331	0.01340958	0.005142483

Realizado por: Santillán et al 2024.

Intervalos de confianza de proporciones con nivel de significancia $\alpha=0.05$

IC ($\alpha=0.05$)= [62.85; 71.93] % inhumaciones ciclo de vida ADULTO MAYOR.

IC ($\alpha=0.05$)= [26.88; 35.74] % inhumaciones ciclo de vida ADULTO.

IC ($\alpha=0.05$)= [0; 1.34] % inhumaciones ciclo de vida JUVENTUD.

IC ($\alpha=0.05$)= [0; 0.51] % inhumaciones ciclo de vida INFANCIA.

El resultado dice que hay una confianza del 95% que el intervalo [62.85; 71.93] % contenga a la proporción de Inhumaciones ciclo de vida ADULTO MAYOR, [26.88; 35.74] % contenga a la proporción de Inhumaciones ciclo de vida ADULTO, [0; 1.34] %

contenga a la proporción de Inhumaciones ciclo de vida JUVENTUD, [0; 0.51] % contenga a la proporción de Inhumaciones ciclo de vida INFANCIA.

6.4. Proporciones ciclos de vida – masculino inhumaciones por COVID

$$n=283$$

$$A \text{ MAYOR} = 187$$

$$ADULTO = 95$$

$$INFANCIA = 1$$

$$JUVENTUD = 0$$

$$NIÑEZ = 0$$

$$ADOLESCENCIA = 0$$

$$p = \frac{A \text{ MAYOR}}{n}$$

$$p = \frac{187}{283}$$

$$p * n \geq 5$$

$$\frac{187}{283} * 283 \geq 5$$

$$187 \geq 5$$

CUMPLE

$$q = \frac{ADULTO}{n}$$

$$q = \frac{95}{283}$$

$$q * n \geq 5$$

$$\frac{95}{283} * 283 \geq 5$$

$$95 \geq 5$$

CUMPLE

$$q_1 = \frac{INFANCIA}{n}$$

$$q_1 = \frac{1}{283}$$

$$q_1 * n \geq 5$$

$$\frac{1}{283} * 283 \geq 5$$

$$1 \geq 5$$

NO
CUMPLE

$$q_2 = \frac{JUVENTUD}{n}$$

$$q_2 = \frac{0}{283}$$

$$q_2 * n \geq 5$$

$$q_3 = \frac{NIÑEZ}{n}$$

$$q_3 = \frac{0}{283}$$

$$q_3 * n \geq 5$$

$$q_4 = \frac{ADOLESCENCIA}{n}$$

$$q_4 = \frac{0}{283}$$

$$q_4 * n \geq 5$$

$$\frac{0}{283} * 283 \geq 5$$

$$\frac{0}{283} * 283 \geq 5$$

$$\frac{0}{283} * 283 \geq 5$$

$$0 \geq 5$$

NO
CUMPLE

$$0 \geq 5$$

NO
CUMPLE

$$0 \geq 5$$

NO
CUMPLE

Los dos valores para A Mayor y Adulto cumplen la condición, con esto se puede aplicar el método BOOTSTRAP para proporciones, por lo cual se puede aplicar el algoritmo y la función dentro del programa estadístico R, con el cual se determinan los límites superior e inferior para determinar los Intervalos de confianza, para los que se aplicaran un nivel de significancia $\alpha=0.05$, teniendo las siguientes gráficas y datos de los intervalos generados para cada una de las variables cualitativas analizadas, para los demás valores debido a que son menores que la condición no cumplen, debido a eso no se los puede aplicar el método, se analizará de manera semejante para visualizar los efectos o desventajas del método.

Con esto se puede tomar los valores de las variables que no cumplen con la condición como datos atípicos.

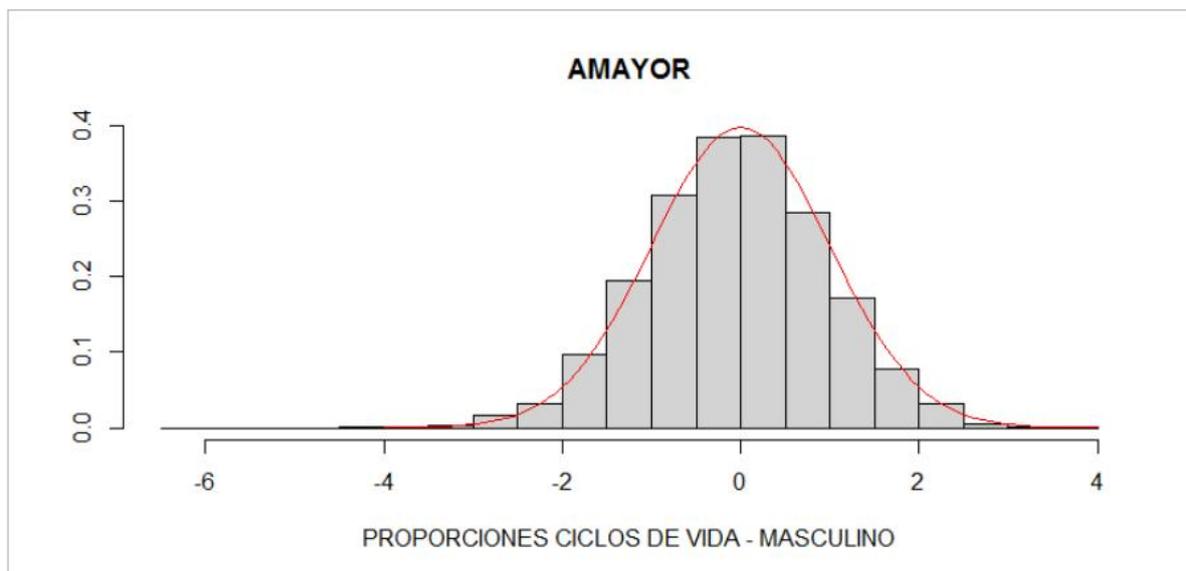


Figura 9. Proporciones Inhumaciones por CICLOS DE VIDA – MASCULINO – Adulto Mayor.

Realizado por: Santillán et al 2024.

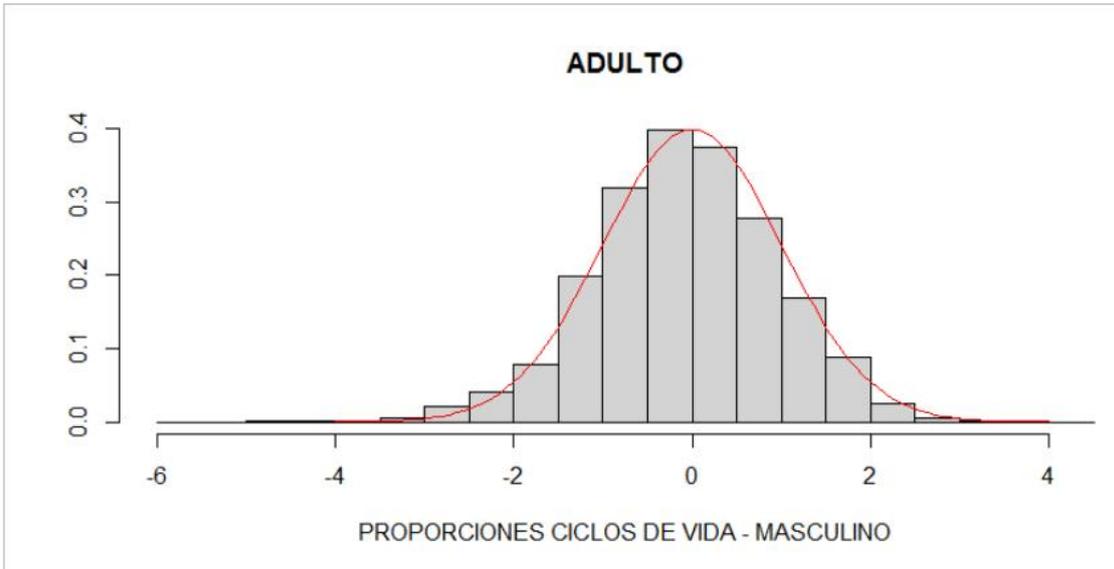


Figura 10. Proporciones Inhumaciones por CICLOS DE VIDA – MASCULINO – Adulto.

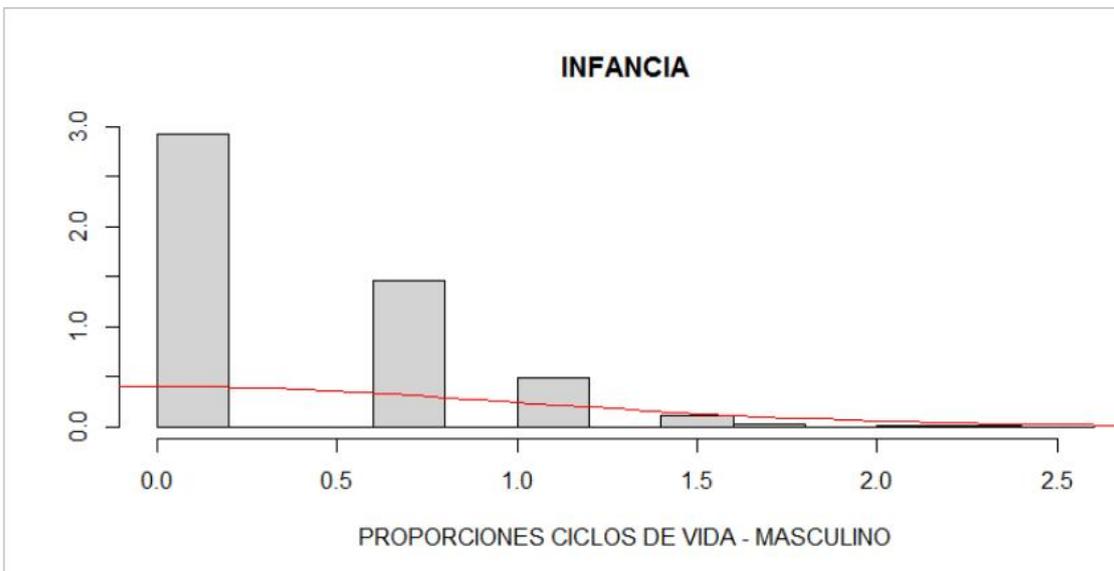


Figura 11. Proporciones Inhumaciones por CICLOS DE VIDA – MASCULINO – Infancia.

Tabla 5: Intervalos de confianza COVID – Ciclos de vida - Masculino.

	AMAYOR	ADULTO	INFANCIA
2.5%	0.5995359	0.276313	-Inf
97.5%	0.7125230	0.390187	0.007586974

Realizado por: Santillán et al 2024.

Intervalos de confianza de proporciones con nivel de significancia $\alpha=0.05$

IC ($\alpha=0.05$)= [59.95; 71.25] % inhumaciones ciclo de vida ADULTO MAYOR.

IC ($\alpha=0.05$)= [27.63; 39.02] % inhumaciones ciclo de vida ADULTO.

IC ($\alpha=0.05$)= [0; 0.76] % inhumaciones ciclo de vida INFANCIA.

El resultado indica que hay una confianza del 95% que el intervalo [59.95; 71.25] % contenga a la proporción de Inhumaciones ciclo de vida ADULTO MAYOR, [27.63; 39.02] % contenga a la proporción de Inhumaciones ciclo de vida ADULTO, [0; 0.76] % contenga a la proporción de Inhumaciones ciclo de vida INFANCIA.

6.5. Proporciones ciclos de vida – femenino inhumaciones por COVID

n=135

A MAYOR = 96

ADULTO = 36

JUVENTUD = 3

INFANCIA = 0

NIÑEZ = 0

ADOLESCENCIA = 0

$$p = \frac{A \text{ MAYOR}}{n}$$

$$p = \frac{96}{135}$$

$$p * n \geq 5$$

$$\frac{96}{135} * 135 \geq 5$$

$$187 \geq 5$$

CUMPLE

$$q = \frac{ADULTO}{n}$$

$$q = \frac{36}{135}$$

$$q * n \geq 5$$

$$\frac{36}{135} * 135 \geq 5$$

$$95 \geq 5$$

CUMPLE

$$q_1 = \frac{JUVENTUD}{n}$$

$$q_1 = \frac{3}{135}$$

$$q_1 * n \geq 5$$

$$\frac{3}{135} * 135 \geq 5$$

$$1 \geq 5$$

NO
CUMPLE

$$q_2 = \frac{INFANCIA}{n}$$

$$q_2 = \frac{0}{135}$$

$$q_3 = \frac{NIÑEZ}{n}$$

$$q_3 = \frac{0}{135}$$

$$q_4 = \frac{ADOLESCENCIA}{n}$$

$$q_4 = \frac{0}{135}$$

$$q_2 * n \geq 5$$

$$\frac{0}{135} * 135 \geq 5$$

$$q_3 * n \geq 5$$

$$\frac{0}{135} * 135 \geq 5$$

$$q_4 * n \geq 5$$

$$\frac{0}{135} * 135 \geq 5$$

$$0 \geq 5$$

NO
CUMPLE

$$0 \geq 5$$

NO
CUMPLE

$$0 \geq 5$$

NO
CUMPLE

Los dos valores para A Mayor y Adulto cumplen la condición, con esto se puede aplicar el método BOOTSTRAP para proporciones, por lo cual se puede aplicar el algoritmo y la función dentro del programa estadístico R, con el cual se determinan los límites superior e inferior para determinar los Intervalos de confianza, para los que se aplicaran un nivel de significancia $\alpha=0.05$, teniendo las siguientes gráficas y datos de los intervalos generados para cada una de las variables cualitativas analizadas, para los demás valores debido a que son menores que la condición no cumplen, debido a eso no se los puede aplicar el método, se analizará de manera semejante para visualizar los efectos o desventajas del método.

Con esto se puede tomar los valores de las variables que no cumplen con la condición como datos atípicos.

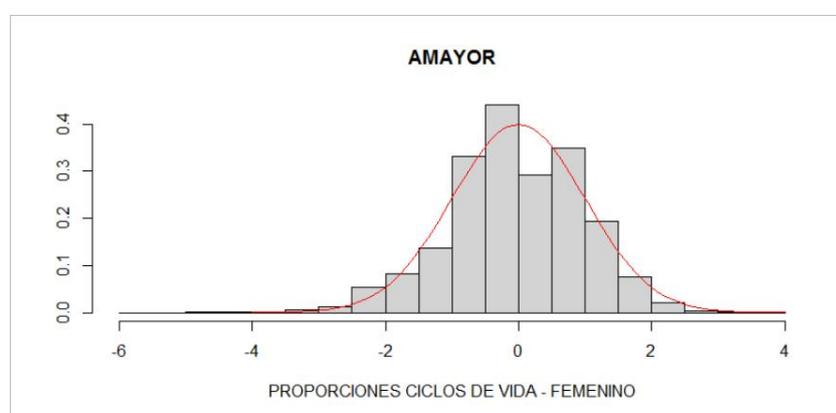


Figura 12. Proporciones Inhumaciones por CICLOS DE VIDA – FEMENINO – Adulto Mayor.

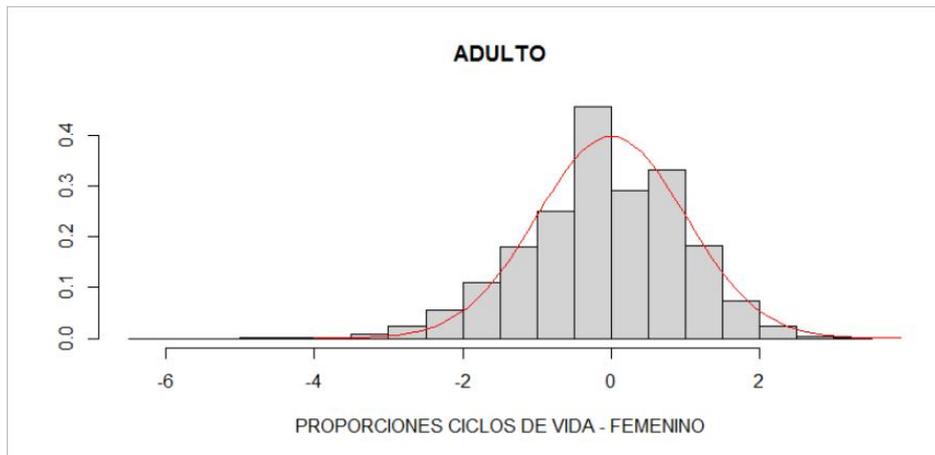


Figura 13. Proporciones Inhumaciones por CICLOS DE VIDA – FEMENINO – Adulto.

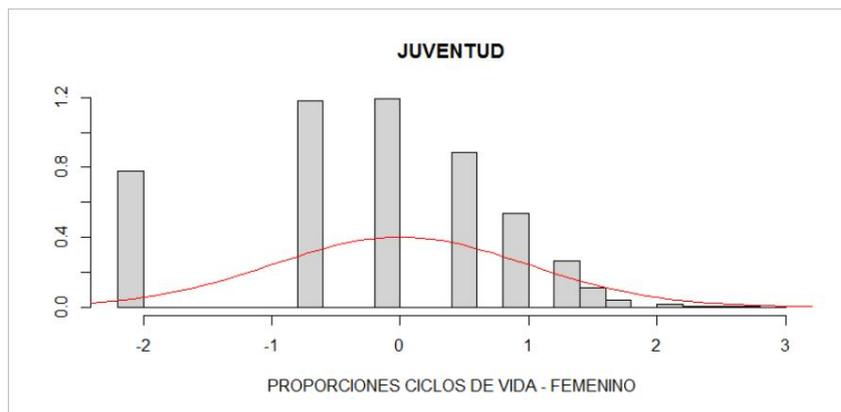


Figura 14. Proporciones Inhumaciones por CICLOS DE VIDA – FEMENINO – Juventud.

Tabla 6. Intervalos de confianza COVID – Ciclos de vida - Femenino.

	AMAYOR	ADULTO	JUVENTUD
2.5%	0.6328488	0.1795039	-Inf
97.5%	0.7816046	0.3321964	0.04161937

Realizado por: Santillán et al 2024

Intervalos de confianza de proporciones con nivel de significancia $\alpha=0.05$

IC ($\alpha=0.05$)= [63.28; 78.16] % inhumaciones ciclo de vida ADULTO MAYOR.

IC ($\alpha=0.05$)= [17.95; 33.22] % inhumaciones ciclo de vida ADULTO.

IC ($\alpha=0.05$)= [0; 4.16] % inhumaciones ciclo de vida INFANCIA.

El resultado dice que hay una confianza del 95% que el intervalo [63.28; 78.16] % contenga a la proporción de Inhumaciones ciclo de vida ADULTO MAYOR, [17.95; 33.22] % contenga a la proporción de Inhumaciones ciclo de vida ADULTO, [0; 4.16] % contenga a la proporción de Inhumaciones ciclo de vida INFANCIA.

Conclusiones

a) Se estimaron los intervalos de confianza BOOTSTRAP de proporciones de factores asociados a inhumaciones por causa de COVID-19, en el cementerio general de Riobamba, periodo marzo 2020 – abril 2021, de los cuales se distribuyeron o separaron en varias variables como sexo y edad, además para el análisis de inhumaciones por COVID-19 dentro de cada uno de los sexos se determinó los intervalos de confianza con los cuales se identifican los grupos vulnerables.

b) Se conoció que el método BOOTSTRAP para re muestreos se puede aplicar en cualquier disciplina o área de las ciencias, si excepciones de ningún tipo, especialmente en investigaciones donde se obtienen distintos datos o mediciones repetidas, proporcionándonos estimaciones y errores de los estadísticos utilizados con los cuales se generan los intervalos de confianza.

c) Se analizó el método BOOTSTRAP para intervalos de confianza de proporciones, encontrando que es uno de los métodos más confiables debido a que este método utiliza el re muestreo con reemplazo donde se pueden generar más muestras y así proyectar, simular o inferir resultados de la población, encontrando como en la investigación planteada grupos vulnerables, a los cuales se los puede tomar medidas de protección y cuidados para evitar pérdidas humanas.

d) De acuerdo con el análisis de vulnerabilidad de los grupos, en cuanto a sexo, el grupo más vulnerable es el sexo masculino. A su vez, dentro de intervalos de edades determinadas mediante los ciclos de vida, se encontró que el grupo más vulnerable es el

conocido como adultos mayores, que son las personas que tienen una edad superior a 60 años, tanto para los sexos masculino y femenino.

Referencias

- Cirera, L., Segura, A., & Hernández, I. (2021). Defunciones por COVID-19: no están todas las que son y no son todas las que están. *Gaceta Sanitaria*, 35(6), 590-593.
- Diaconis, P., y Efron, B. (1983). Computer intensive methods in statistics. *Scientific American*, 248 (5), 116-13.
- Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*. Vol. 7, No. 1.
- Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap, and other resampling methods. *Biometrika*, 68, 589-599.
- Efron, B. y Tibshirani, R. (1993) An Introduction to the Bootstrap. Chapman & Hall.
- Llaro-Sánchez, M. K., Gamarra-Villegas, B. E., & Campos-Correa, K. E. (2020). Características clínico-epidemiológicas y análisis de sobrevivencia en fallecidos por COVID-19 atendidos en establecimientos de la Red Sabogal-Callao 2020. *Horizonte Médico (Lima)*, 20(2).
- Manrique Abril, F., González-Chordá, V. M., Gutiérrez Lesmes, O. A., Tellez Piñerez, C. F., & Herrera-Amaya, G. M. (2020). Modelo SIR de la pandemia de Covid-19 en Colombia.
- Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures. *Psychological Bulletin*, 105(1), 156-166.
- Mosier, C. (1951). Problems and Designs of Cross-Validation. *Educational and Psychological Measurement*, 11. Noreen, E. (1989). Computer intensive methods for testing hypotheses. New York: John Wiley & Sons, Ltd.
- Quenouille, M. H. (1956). Notes on Bias in Estimation. *Biometrika*, 43, 353-36.
- Sangrador, C. O., Leiza, J. R. G., Boillos, M. J. P., Ara, F. P., Lobato, M. D. P. L., & de Llano, J. M. A. (2021). Impacto de la COVID-19 en la mortalidad de la comunidad autónoma de Castilla y León. *Gaceta Sanitaria*, 35(5), 459-464.
- Simon, J. L. (Ed.). (1969). *Basic Research Methods in Social Science*. (Vol. acceso en Marzo 2003). New York: Random House.
- Tukey (1958). Bias and Confidence in Not-Quite Large Samples (Abstract). *Annals of Mathematical Statistics*, 29, 614.