Polo del Conocimiento

Pol. Con. (Edición núm. 85) Vol. 8, No 8
Agosto 2023, pp. 2463-2472
ISSN: 2550 - 682X
DOI: 10.23857/pc.v8i8

Estudio estadístico inferencial del incremento de Homicidios Intencionales para determinar la afectación en el comportamiento delictual a la población nacional a partir de la base de datos de Homicidio Intencional utilizando programación en R.

Inferential statistical study of the increase in Intentional Homicide to determine the affectation of criminal behavior in the national population from the Intentional Homicide database using programming in R.

Estudo estatístico inferencial do aumento dos Homicídios Intencionais para determinar a afetação do comportamento criminoso na população nacional a partir da base de dados de Homicídios Intencionais utilizando programação em R.

Juan Carlos Yungán Cazar [I]
jyungan@espoch.edu.ec
https://orcid.org/0000-0001-5682-0399

Katherine Adriana Merino Villa [II]
kathetine.merino@espoch.edu.ec
https://orcid.org/0009-0001-0616-9611

Edgar Gualberto Salazar Álvarez [III]
edgar.salazar@espoch.edu.ec
https://orcid.org/0000-0003-0988-0641

Diego Alejandro Cáceres Veintimilla [IV]
diego.caceres@espoch.edu.ec
https://orcid.org/0000-0003-0498-1240

Correspondencia: jyungan@espoch.edu.ec

Ciencias Tecnologías de la Información y la Comunicación
Artículo de Investigación

*** Recibido:** 23 de junio de 2023 ***Aceptado:** 12 de julio de 2023 *** Publicado:** 28 de agosto de 2023

I. Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.
II. Ingeniera en Electrónica Telecomunicaciones y Redes Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.
III. Escuela Superior Politécnica de Chimborazo. Riobamba, Ecuador.
IV. Escuela Superior Politécnica de Chimborazo. Riobamba, Ecuador.

Resumen

La programación estadística en R es un enfoque clave en la ciencia de datos que involucra el uso del lenguaje de programación R para realizar análisis estadísticos, manipulación de datos, visualización y generación de informes. Es ampliamente utilizado por profesionales en diversas disciplinas para analizar y comprender conjuntos de datos complejos. El proceso de recolección, almacenamiento, análisis y visualización de datos contenidos en Bases de Datos gubernamentales en Ecuador mediante la programación estadística en R puede desglosarse de la siguiente manera: Se inicia accediendo a las Bases de Datos públicas. Los datos recolectados se importan y almacenan en el entorno de trabajo de R. Es común que los datos requieran preparación antes del análisis. Esto incluye la limpieza de valores atípicos, el tratamiento de valores faltantes y la transformación de datos en el formato adecuado para su análisis. Una vez que los datos están preparados, se pueden realizar análisis estadísticos utilizando diversas técnicas, como estadísticas descriptivas, pruebas de hipótesis, análisis de regresión, clustering, entre otros. La visualización de datos es fundamental para comunicar los resultados de manera efectiva. R proporciona paquetes como ggplot2, que permiten crear gráficos de alta calidad. La generación de Informes se lo realiza utilizando la herramienta R Markdown. Una vez que se han realizado los análisis y generado los informes, los resultados pueden ser compartidos.

Palabras Clave: Programación estadística; Bases de datos estructuradas; prueba de hipótesis; índices de homicidio.

Abstract

Statistical programming in R is a key approach in data science that involves using the R programming language to perform statistical analysis, data manipulation, visualization, and report generation. It is widely used by professionals in various disciplines to analyze and understand complex data sets. The process of collection, storage, analysis and visualization of data contained in government databases in Ecuador through statistical programming in R can be broken down as follows: It begins by accessing public databases. The collected data is imported and stored in the R framework. It is common for the data to require preparation before analysis. This includes cleaning outliers, handling missing values, and transforming data into the appropriate format for analysis. Once the data is prepared, statistical analyzes can be performed using various techniques, such as descriptive statistics, hypothesis testing, regression analysis, clustering, among others. Data

visualization is critical to communicating results effectively. R provides packages such as ggplot2, which allow you to create high-quality plots. Report generation is done using the R Markdown tool. Once the analyzes have been carried out and the reports generated, the results can be shared.
Keywords: Statistical programming; Structured databases; hypothesis testing; homicide rates.

Resumo

A programação estatística em R é uma abordagem chave na ciência de dados que envolve o uso da linguagem de programação R para realizar análise estatística, manipulação de dados, visualização e geração de relatórios. É amplamente utilizado por profissionais de diversas disciplinas para analisar e compreender conjuntos de dados complexos. O processo de coleta, armazenamento, análise e visualização dos dados contidos nas bases de dados governamentais do Equador através da programação estatística em R pode ser dividido da seguinte forma: Começa pelo acesso às bases de dados públicas. Os dados coletados são importados e armazenados na estrutura R. É comum que os dados exijam preparação antes da análise. Isso inclui a limpeza de valores discrepantes, o tratamento de valores ausentes e a transformação de dados no formato apropriado para análise. Uma vez preparados os dados, as análises estatísticas podem ser realizadas utilizando diversas técnicas, como estatística descritiva, teste de hipóteses, análise de regressão, agrupamento, entre outras. A visualização de dados é fundamental para comunicar resultados de forma eficaz. R fornece pacotes como ggplot2, que permitem criar gráficos de alta qualidade. A geração do relatório é feita através da ferramenta R Markdown. Uma vez realizadas as análises e gerados os relatórios, os resultados podem ser compartilhados.
Palavras-chave: Programação estatística; bases de dados estruturadas; testes de hipótesis; taxas de homicídios.

Introducción

La programación estadística inferencial en R se erige como una herramienta poderosa para extraer conocimientos profundos de los datos. R es un lenguaje de programación y un entorno especializado en estadísticas y análisis de datos, que permite realizar análisis inferenciales con eficiencia y precisión.

En este contexto, la utilización de bases de datos estructuradas cobra vital importancia. Estas bases organizan la información en tablas con filas y columnas, lo que facilita la gestión y manipulación

de datos. R se integra de manera perfecta con bases de datos estructuradas, permitiendo la conexión directa y la extracción de información relevante para el análisis estadístico.

Mediante la programación estadística inferencial en R, es posible realizar estimaciones de parámetros, pruebas de hipótesis y modelado predictivo a partir de los datos estructurados. Las librerías y funciones especializadas en estadísticas de R proporcionan las herramientas necesarias para llevar a cabo análisis sofisticados y visualizar los resultados de manera efectiva.

Estadística Inferencial - Descripción "Contrastes de Hipótesis".

Muchos problemas requieren decidir si se acepta o rechaza un enunciado acerca de algún parámetro (estadístico). Para esto, se considera hipótesis, y el procedimiento para la toma de decisiones en torno a probar o no la hipótesis, recibe el nombre de prueba de hipótesis. (Triola, 2004)

Contrastar una hipótesis consiste en probar un valor observado con un valor definido por el investigador, los mismos que se han desarrollado en muchos campos tales como:

- Agricultura, para decidir que fertilizante entrega los mejores resultados
- Medicina, para probar si un medicamento es mejor que otro
- Industria, para validar procesos orientados a mejorar y mantener estándares de calidad
- Entre muchos otros.

Metodología

Algoritmo para realizar una prueba de hipótesis:

1. Definir el valor a contrastar, por lo general es un valor determinado a priori por experiencia, regulaciones o experiencias realizadas anteriormente.
2. Definir la hipótesis nula y la alternativa
3. Determinar el nivel de error de la prueba (esto refiere al Error tipo I)
4. Determinar la distribución de contraste, por lo general depende del número de observaciones con que estemos trabajando
5. Definir una región de aceptación o rechazo
6. Estimar el valor a contrastar de las observaciones disponibles
7. Decidir si aceptar o no la hipótesis nula.

Modelo para aplicar contraste de hipótesis

En el Ecuador a partir del año 2018, se tiene un incremento en los Homicidios Intencionales (HI), afectando este comportamiento delictual a la población nacional, estamos interesados en poder comprobar si este fenómeno delictual afecta a personas jóvenes o no, y si afecta a hombres y mujeres jóvenes. Para esto se planteará dos tipos de Contrastes de hipótesis:

- La primera comprobaremos si la media de las edades de fallecidos es igual a la edad de jóvenes que según la Organización Mundial de la Salud (OMS), se consideran personas jóvenes, las comprendidas entre 10 y los 24 años, para el propósito de nuestro análisis consideraremos el punto medio es decir 17 años en relación con el total de fallecidos.
- Otra prueba que realizaremos consistirá en probar si las edades medias que se tiene en HI son iguales en hombres y mujeres. (Santana, 2014)

Planteamiento del estudio:

El estudio consistirá en dos actividades, la primera probaremos si el promedio de edad de HI en Ecuador es igual a la edad de jóvenes definido anteriormente en 17 años, es decir nuestra prueba es la siguiente:

$$H_0: \mu = 17$$
$$H_1: \mu \neq 17$$

Esa será nuestra primera estimación, y la segunda consiste en:

El segundo contraste considera dos poblaciones de Hombres y Mujeres de HI, se quiere probar si la edad de los HI es iguale entre hombres y mujeres así:

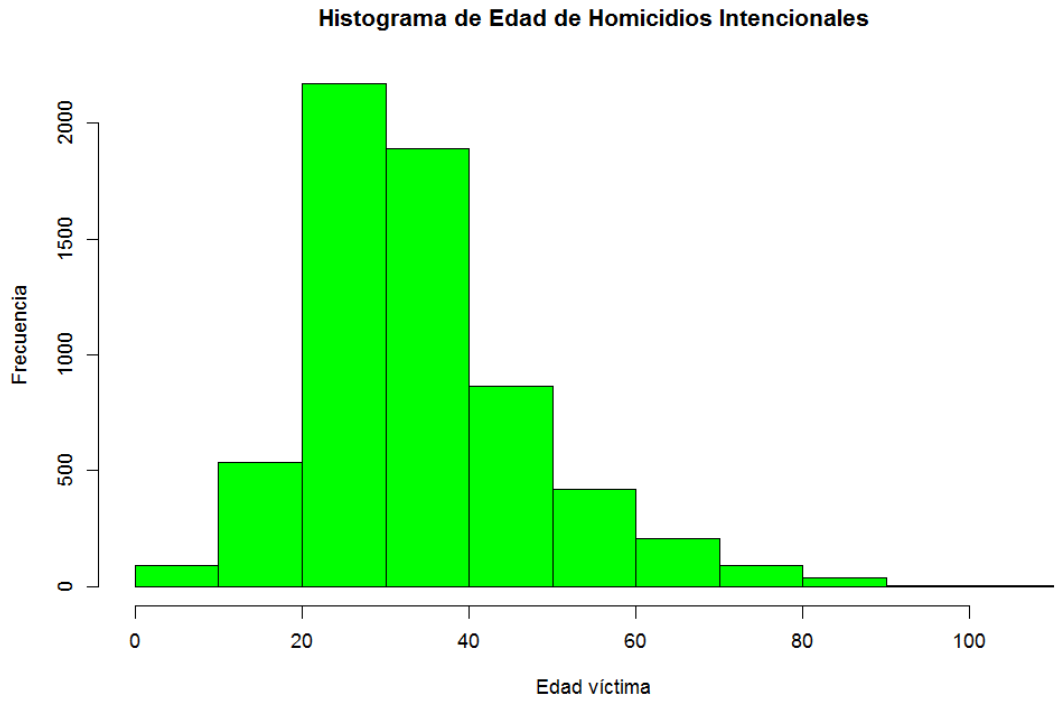
$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Tipo de base de datos a recabar para el estudio

Para ejecutar los dos contrastes planteados, se procedió a recabar la base de datos de Homicidios Intencionales, anonimizada para guardar la reserva de la información sensible las variables a considerar es la edad y sexo de la víctima, se tiene los datos en el periodo de 2018 hasta el 2021, desde Medicina Legal, institución que realiza el protocolo de autopsia sobre toda muerte violenta y particularmente sobre los homicidios intencionales. (Hernández, 2022)

Desarrollo del estudio

Para ejecutar los dos Contrastes de hipótesis, se utilizará la base de datos de Homicidio Intencional, por edad y sexo de la víctima desde el año 2018 hasta el año 2021 que consta de 6362 registros. Iniciamos con un análisis exploratorio, iniciamos con un histograma (Donoso 2023)



Gráfica 1. Edad de homicidios intencionales

La figura muestra el histograma de las edades en años de los Homicidios intencionales registrados entre el 2018 y 2021 en Ecuador.

Contraste de la prueba:

data: pruebah$EDAD

t = 103.36, df = 6300, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 17

95 percent confidence interval:

34.24478 34.91154

sample estimates:

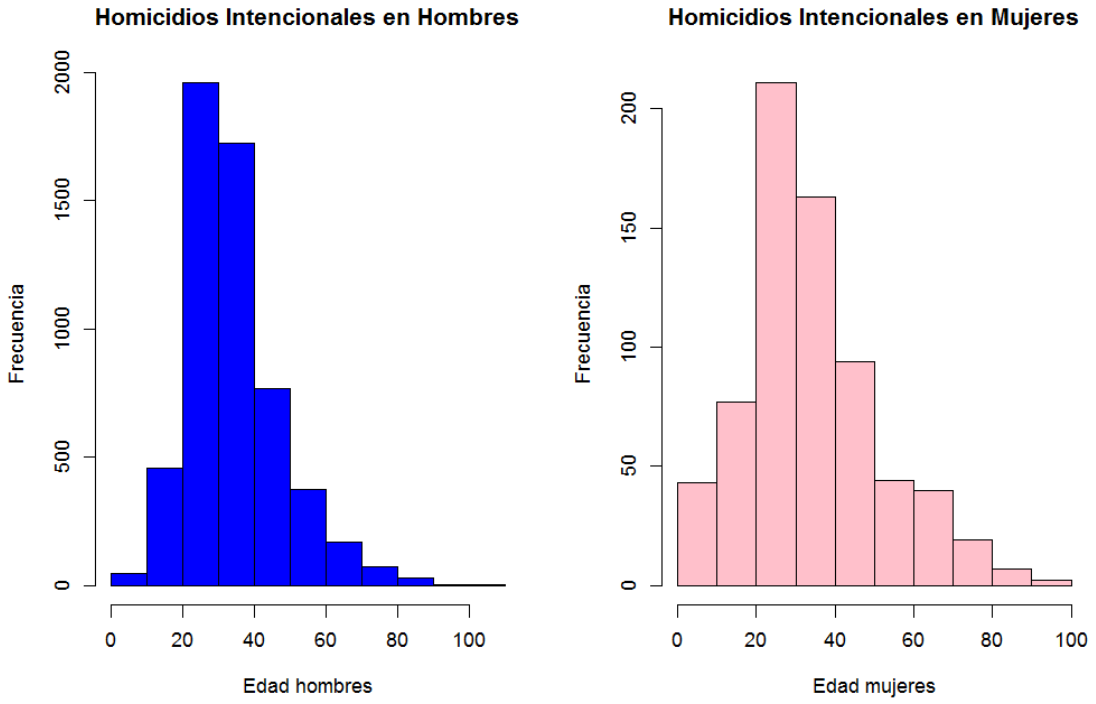
mean of x

34.57816

La prueba rechaza H0 con un nivel de error muy bajo, esto es los homicidios intencionales no se tiene en personas jóvenes de 17 años, la media estimada es de 35 años aproximadamente.

Ahora realicemos la prueba de hipótesis para determinar si las edades en hombres y mujeres son iguales.

Para esto consideramos dos muestras independientes Hombres y Mujeres (Bruce, 2022)



Gráfica 2. Edad de homicidios intencionales en hombres y mujeres

Se presenta un histograma de homicidios intencionales en hombres y mujeres según la edad.

data: HombresHI and MujeresHI
t = 0.52708, df = 6294, p-value = 0.5982
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.7756959 1.3462156
sample estimates:
mean of x mean of y
34.61383 34.32857

Pol. Con. (Edición núm. 85) Vol. 8, No 8, Agosto 2023, pp. 2463-2472, ISSN: 2550 - 682X

Al contrastar la prueba se tiene que no se rechaza la igualdad de edades medias, esto ya que, el p-valor es del 0.5982, esto indica que las edades en homicidios intencionales son similares entre hombres y mujeres. (Vásquez, 2022).

Discusión y resultados

Para la realización del presente taller, se trabajó con una base de datos reales de homicidios intencionales, adicional a esto para la prueba de la muestra total de edades como para la prueba de igualdad de edades medias entre hombres y mujeres, consideramos que provienen de una distribución normal, así como la varianza desconocida pero constante y finita.

Los principales resultados que se obtuvo fue que la edad de los homicidios difiere de la edad de joven que es 17 años, esto hace suponer que el comportamiento de este tipo de violencia puede estar asociado a diferentes factores que puede aumentar conforme aumenta la edad así en la prueba se tiene que el promedio de homicidios se tiene alrededor de los 34 años.

Adicional a esto, al comparar la edad media de hombres y mujeres, se tiene que son iguales, lo que indica que las edades en hombres y mujeres que sufren este tipo de violencia es la misma, lo interesante sería profundizar el análisis y poder determinar si existen otros factores asociados a este fenómeno delictual que pueda diferenciar el comportamiento en hombres y mujeres.

Conclusiones

1. La programación estadística inferencial en R emerge como una herramienta poderosa para extraer conocimientos profundos de los datos. R, como lenguaje y entorno especializado en estadísticas y análisis de datos, posibilita realizar análisis inferenciales con precisión y eficiencia.
2. La utilización de bases de datos estructuradas cobra vital importancia en este contexto. Estas bases organizan la información en tablas, facilitando la manipulación y gestión de datos. La integración fluida entre R y bases de datos estructuradas permite conexiones directas y extracción de información relevante para análisis estadísticos.
3. La programación estadística inferencial en R permite realizar estimaciones de parámetros, pruebas de hipótesis y modelado predictivo a partir de datos estructurados. Las librerías y

funciones especializadas en estadísticas de R proporcionan las herramientas necesarias para análisis sofisticados y visualización efectiva de resultados.

4. Los contrastes de hipótesis son esenciales para la toma de decisiones en diferentes campos como agricultura, medicina e industria. En este contexto, el texto presenta ejemplos específicos de contraste de hipótesis, como la comparación de edades de víctimas de homicidios intencionales con una definición de personas jóvenes y la comparación de edades entre hombres y mujeres.
5. El análisis de los homicidios intencionales en Ecuador revela que las edades de las víctimas difieren de la definición de jóvenes, lo que sugiere que el fenómeno delictual puede estar asociado a diversos factores a medida que aumenta la edad. Además, se concluye que las edades medias de hombres y mujeres víctimas son similares, lo que plantea la posibilidad de profundizar en factores que puedan influir en este comportamiento delictivo diferencial entre géneros.

Referencias

Triola, F. M. (2004). Probabilidad y estadística - Mario F. Triola - Google Libros. In Probabilidad y estadística (Vol. 1).

Santana Sepúlveda, S., & Mateos Farfán, E. (2014). El arte de programar en R: un lenguaje para la estadística.

Donoso, M. E. A., Maurisaca, N. E. C., & Reyes, J. E. A. (2022). Análisis de Correspondencias Múltiples para el Estudio de los Homicidios Intencionales en el Ecuador. Revista Politécnica, 50(3), 43-52.

Hernández Bringas, H. (2022). Homicidios en América Latina y el Caribe: magnitud y factores asociados. Notas de población.

Bruce, P., Bruce, A., & Gedeck, P. (2022). Estadística práctica para ciencia de datos con R y Python. Marcombo.

Vásquez Sánchez, E., & Ortiz Basauri, G. M. (2022). Estadística Inferencial en la lógica de la investigación científica.