

Universidad Nacional del Altiplano

Facultad de Ingeniería Estadística e Informática

Docente: Fred Torres Cruz

Autor: Nelson Catunta Huisa

Trabajo Encargado - N° 003

<https://github.com/CAISA25/TRABAJO003.git>

<https://caisa.shinyapps.io/publicacion003/>

Documentación – Aplicación Web de Análisis Estadístico en R Shiny

1. Introducción

Esta documentación proporciona una guía detallada para utilizar la aplicación web desarrollada en R Shiny, diseñada para realizar análisis estadísticos de datos. La aplicación es intuitiva y está orientada a usuarios sin conocimientos previos en estadística o programación. Su objetivo es automatizar tareas estadísticas complejas, ofrecer resultados precisos e interactivos, y facilitar la visualización y exportación de los resultados obtenidos.

2. Objetivo General

El principal objetivo de esta aplicación es proporcionar una herramienta accesible que permita a los usuarios realizar análisis estadísticos sin necesidad de habilidades avanzadas en programación o matemáticas. La aplicación ofrece funcionalidades de importación de datos, análisis descriptivo, pruebas estadísticas, y visualización gráfica de los resultados.

3. Definiciones Clave

A continuación se detallan los términos esenciales para comprender el funcionamiento de la aplicación. Estas definiciones están orientadas a un público sin experiencia previa en estadística o programación.

R Shiny

R Shiny es un marco de trabajo (framework) para construir aplicaciones web interactivas utilizando el lenguaje de programación R. Shiny permite a los usuarios crear interfaces gráficas sin necesidad de escribir código HTML, CSS o JavaScript. Gracias a Shiny, es posible desarrollar aplicaciones donde los usuarios interactúan con los datos a través de botones, menús y gráficos, todo en tiempo real.

Análisis Estadístico Descriptivo

El análisis estadístico descriptivo es el proceso mediante el cual se resumen y describen las características principales de un conjunto de datos. Este tipo de análisis no busca hacer

inferencias o predicciones sobre los datos, sino simplemente proporcionar una visión general de las características de los mismos.

- **Medidas de tendencia central:** Ayudan a identificar un valor representativo de los datos.
 - **Media:** Es el promedio de todos los valores.
 - **Mediana:** Es el valor central cuando los datos se ordenan de menor a mayor.
 - **Moda:** Es el valor que más se repite en un conjunto de datos.
- **Medidas de dispersión:** Indican qué tan dispersos están los datos respecto a un valor central.
 - **Desviación estándar:** Mide el grado de variabilidad de los datos con respecto a la media.
 - **Varianza:** Es el cuadrado de la desviación estándar y también mide la dispersión de los datos.
 - **Rango intercuartílico (IQR):** Representa la distancia entre el primer cuartil (Q1) y el tercer cuartil (Q3), indicando la dispersión de los valores centrales.

Visualización de Datos

La visualización de datos es el uso de representaciones gráficas para explorar, analizar y presentar datos. Utilizar gráficos adecuados puede ayudar a comprender patrones, identificar anomalías y comunicar hallazgos de manera más efectiva. Los tipos comunes de gráficos incluyen:

- **Histogramas:** Muestran la distribución de frecuencias de una variable.
- **Boxplots:** Ayudan a identificar valores atípicos y la variabilidad de los datos.
- **Gráficos de barras:** Permiten comparar categorías entre sí.
- **Gráficos de líneas:** Usados para mostrar tendencias a lo largo del tiempo.
- **Diagramas de dispersión:** Muestran la relación entre dos variables numéricas.

Pruebas Estadísticas

Las pruebas estadísticas son procedimientos matemáticos utilizados para tomar decisiones sobre un conjunto de datos. Estas pruebas permiten validar hipótesis, como la relación entre dos variables o si un conjunto de datos sigue una distribución específica. Algunas pruebas comunes son:

- **Prueba de normalidad de Shapiro-Wilk:** Determina si un conjunto de datos sigue una distribución normal. Esto es fundamental porque muchas pruebas estadísticas requieren que los datos sean normales.

- Prueba t de Student: Compara las medias de dos grupos para ver si son significativamente diferentes.
- Prueba ANOVA (Análisis de varianza): Compara las medias de tres o más grupos para determinar si al menos uno de ellos es diferente de los demás.
- Prueba de Levene: Evalúa si varias muestras tienen la misma varianza, lo que es necesario para algunas pruebas estadísticas.
- Correlación de Pearson y Spearman: Mide la fuerza y dirección de la relación entre dos variables numéricas.

Importación de Archivos

La importación de archivos es el proceso de cargar datos desde un archivo externo (como un archivo CSV o Excel) a la aplicación para su posterior análisis. La aplicación permite importar los siguientes formatos:

- **.csv** (Comma Separated Values): Un formato de texto donde los valores están separados por comas.
- **.xlsx** (Excel): Un formato de hoja de cálculo que puede contener múltiples hojas y diferentes tipos de datos.
- **.txt** (Texto plano): Archivos de texto donde los datos están organizados de manera simple.

4. Estructura General de la Aplicación

La aplicación se divide en cinco secciones principales, cada una diseñada para facilitar tareas específicas:

1. **Importar Datos:** Cargar el archivo con los datos que serán analizados.
2. **Resumen Estadístico:** Generar estadísticas descriptivas para comprender mejor los datos.
3. **Visualización Gráfica:** Crear representaciones gráficas interactivas de los datos.
4. **Pruebas Estadísticas:** Realizar pruebas estadísticas para verificar hipótesis.
5. **Exportación de Resultados:** Descargar los resultados y gráficos generados.

5. Modo de Uso

5.1. Importación de Datos

Archivos admitidos: .csv, .xlsx, .txt

Pasos:

1. Hacer clic en el botón “*Seleccionar archivo*”.
2. Seleccionar el archivo desde su equipo.
3. Verificar que los datos se visualicen correctamente en la tabla de vista previa.

Recomendación: Asegúrese de que los archivos tengan nombres de columnas claros y sin caracteres especiales.

5.2. Análisis Estadístico Descriptivo

En esta sección, la aplicación calcula las medidas estadísticas básicas del conjunto de datos, tales como:

- Medias (media, mediana, moda)
- Desviación estándar y varianza
- Rango intercuartílico (IQR)
- Mínimo y máximo de los datos

Estas métricas se utilizan para obtener una comprensión general de la distribución y dispersión de los datos.

5.3. Visualización Gráfica

La aplicación permite generar los siguientes gráficos:

- Histograma: Para examinar la distribución de una variable.
- Boxplot: Para detectar valores atípicos y entender la dispersión.
- Gráfico de barras: Útil para comparar categorías de datos.
- Gráfico de líneas: Ideal para observar tendencias a lo largo del tiempo.
- Gráfico de dispersión: Para ver la relación entre dos variables.

5.4. Pruebas Estadísticas

Esta sección permite realizar varias pruebas inferenciales, como la prueba t, ANOVA, y análisis de correlación. Los resultados incluirán valores p, lo que indica si los resultados son estadísticamente significativos o no.

5.5. Exportación de Resultados

Los resultados pueden ser exportados en varios formatos:

- Archivos de datos: `.csv`, `.xlsx`
- Gráficos: `.png`, `.pdf`
- Informes: `.html`, `.pdf`

6. Metodología de Desarrollo

La aplicación fue desarrollada utilizando R Shiny para la interfaz web interactiva. Se emplearon bibliotecas como:

- `shiny` y `shinydashboard` para la creación de la interfaz.
- `ggplot2` y `plotly` para la visualización gráfica.
- `dplyr` y `tidyr` para la manipulación de los datos.
- `car`, `nortest`, `psych` para realizar análisis estadísticos y pruebas de normalidad.

La arquitectura modular permite una fácil expansión, mantenimiento y actualización de la aplicación.

7. Resultados Esperados

Se espera que la aplicación logre:

- Reducir significativamente el tiempo necesario para realizar análisis estadísticos manualmente.
- Ofrecer a los usuarios una plataforma accesible para la generación de análisis descriptivos y pruebas estadísticas.
- Proporcionar informes claros y gráficos que faciliten la toma de decisiones basadas en datos.

8. Recomendaciones Finales

- **Verificación de Datos:** Asegúrese de que los datos no tengan valores faltantes o errores de formato antes de cargarlos.
- **Comprensión del Contexto:** Asegúrese de interpretar los resultados dentro del contexto de su investigación o análisis.
- **Frecuencia de Actualización:** Revise la aplicación regularmente para estar al tanto de posibles actualizaciones o nuevas funcionalidades.

9. Referencias Bibliográficas

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning*. Springer.
- Wickham, H., & Grolemund, G. (2016). *R for Data Science*. O'Reilly Media.
- Navarro, D. (2021). *Learning statistics with R: A tutorial for psychology students and other beginners*.
- RStudio. (2025). Shiny Documentation. <https://shiny.posit.co/>
- ggplot2 Documentation. <https://ggplot2.tidyverse.org/>