

Universidad Nacional del Altiplano

Facultad de Ingeniería Estadística e Informática

Docente: TORRES CRUZ FRED

Autor/autores: Harrison Capia Tintaya

Código matrícula: 221301

Link GitHub: <https://github.com/HacapoXd/Finesi-lp3/tree/main/homeworks>

Link app: <https://hacapoXd.shinyapps.io/statistics/>

Pruebas de Hipótesis desde Excel utilizando Shiny

1. Introducción

La estadística es una herramienta fundamental en el análisis de datos, permitiendo realizar inferencias y tomar decisiones basadas en pruebas rigurosas. En el contexto de las pruebas de hipótesis, es común que los investigadores necesiten realizar cálculos y análisis sobre conjuntos de datos que provienen de diversas fuentes. Sin embargo, realizar estas pruebas de forma manual puede ser tedioso y propenso a errores. En este artículo, se presenta una aplicación web desarrollada utilizando el entorno de programación R y el framework Shiny, que permite a los usuarios realizar pruebas de hipótesis directamente desde archivos Excel, facilitando el análisis y la toma de decisiones.

El objetivo de este trabajo es desarrollar una herramienta interactiva y accesible para realizar pruebas estadísticas, tales como t-tests, pruebas de proporciones, análisis de varianzas, entre otros, directamente desde una interfaz gráfica de usuario (GUI). Esta herramienta está diseñada para usuarios que requieren realizar análisis rápidos y eficientes sin necesidad de un profundo conocimiento en programación.

2. Metodología

La aplicación se desarrolló utilizando el paquete **Shiny** de R, que permite crear aplicaciones web interactivas de manera sencilla y eficiente. A través de esta aplicación, los usuarios pueden cargar archivos Excel (.xls o .xlsx) y seleccionar las columnas de los datos que desean analizar. Una vez seleccionadas las columnas y la prueba de hipótesis a realizar, el sistema ejecuta el análisis estadístico y presenta los resultados de forma inmediata.

2.1. Pruebas de Hipótesis y Requerimientos

Cada tipo de prueba de hipótesis tiene ciertos requerimientos y condiciones que deben cumplirse para que el análisis sea válido. A continuación se describen las pruebas disponibles en la aplicación y sus respectivas condiciones:

- **Pruebas para Medias:**

- **Z para una media:** Esta prueba se utiliza para comparar la media de una muestra con un valor conocido (como la media poblacional). Se requiere que la muestra sea de distribución normal y que se conozca la desviación estándar de la población.

- **t para una media:** Similar a la prueba Z, pero se utiliza cuando la desviación estándar de la población es desconocida. Requiere que los datos sigan una distribución normal o, si la muestra es grande (mayor a 30), que los datos tengan una distribución aproximadamente normal.
 - **t para dos medias independientes:** Compara las medias de dos muestras independientes. Para que la prueba sea válida, se asume que ambas muestras son independientes, los datos son numéricos y siguen una distribución normal. Además, las varianzas de las dos muestras deben ser aproximadamente iguales.
 - **t para dos medias pareadas:** Se utiliza para comparar dos medias de muestras relacionadas. Se requiere que los datos sean numéricos y que se cumpla la suposición de normalidad de las diferencias entre las parejas de datos.
- **Pruebas para Proporciones:**
 - **Z para una proporción:** Compara la proporción de éxitos en una muestra con una proporción conocida. Requiere que los datos sean binarios (0 o 1), y que el tamaño de la muestra sea suficientemente grande (generalmente, tanto np como $n(1 - p)$ deben ser mayores a 5).
 - **Z para dos proporciones:** Compara las proporciones de éxitos entre dos muestras independientes. Se requiere que cada grupo tenga un tamaño suficientemente grande y que las proporciones sean aproximadas a una distribución normal.
 - **Pruebas de Varianzas:**
 - **Chi-cuadrado para una varianza:** Se utiliza para probar si la varianza de una muestra es igual a un valor especificado. Se asume que los datos siguen una distribución normal. La prueba es sensible a la falta de normalidad.
 - **F para comparar dos varianzas:** Compara las varianzas de dos muestras independientes. Se requiere que los datos en ambas muestras sigan una distribución normal y que las muestras sean independientes.
 - **Pruebas de Normalidad:**
 - **Shapiro-Wilk:** Esta prueba evalúa si una muestra sigue una distribución normal. Es adecuada para muestras pequeñas (menos de 50 observaciones). Los datos deben ser numéricos.
 - **Kolmogorov-Smirnov:** Compara la distribución de una muestra con una distribución teórica (como la normal). También requiere que los datos sean numéricos y que se cumpla la suposición de normalidad.
 - **Análisis de Varianza (ANOVA):**
 - **ANOVA de un factor:** Compara las medias de más de dos grupos. Requiere que los datos sean numéricos, y que los grupos sean independientes y sigan una distribución normal. También se asume que las varianzas entre los grupos son homogéneas (homocedasticidad).

2.2. Interfaz de Usuario

La interfaz de usuario de la aplicación se organiza en dos paneles principales. En el panel lateral, el usuario puede cargar un archivo Excel, seleccionar las columnas a analizar y elegir la prueba de hipótesis que desea realizar. El panel principal muestra los resultados de la prueba seleccionada, que pueden incluir tanto valores numéricos como gráficos que visualizan la distribución de los datos.

3. Resultados

Una vez cargados los datos y seleccionada la prueba de hipótesis, el sistema realiza el análisis estadístico y devuelve los resultados. Dependiendo del tipo de prueba, los resultados incluyen valores como la media, el valor p , los grados de libertad, el estadístico de prueba, entre otros.

En el caso de una prueba t para una media, por ejemplo, la aplicación ejecuta la función `t.test()` de R, que devuelve el valor p y los intervalos de confianza para la media de la muestra. En pruebas de normalidad, como la prueba de Shapiro-Wilk, la aplicación devuelve un valor p que indica si los datos siguen una distribución normal. Además, se genera un gráfico de distribución para visualizar los datos y sus posibles desviaciones de la normalidad.

4. Conclusión

La aplicación web desarrollada utilizando Shiny proporciona una herramienta interactiva para realizar pruebas de hipótesis estadísticas de manera rápida y eficiente, sin la necesidad de conocimiento profundo en programación. Al permitir la carga de datos desde archivos Excel y realizar múltiples tipos de pruebas estadísticas, la aplicación es una herramienta útil tanto para estudiantes como para profesionales en el área de la estadística. Esta herramienta facilita la toma de decisiones basadas en análisis de datos rigurosos y puede aplicarse en una amplia gama de contextos, desde investigaciones académicas hasta análisis en el ámbito empresarial.

El futuro desarrollo de la aplicación podría incluir la expansión de las pruebas disponibles, la integración con bases de datos más grandes y la mejora de la visualización de los resultados para facilitar la interpretación de los mismos.

5. Referencias

- Chang, W. (2023). shiny: Web Application Framework for R. R package version 1.7.1. <https://cran.r-project.org/web/packages/shiny/index.html>
- Wickham, H., Grolemund, G. (2017). R for Data Science. O'Reilly Media.
- R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.r-project.org/>