

Dashboard en Shiny para modelación estadística *

Oscar Daniel Rivera Baena *Universidad de Antioquia*
Olga Cecilia Usuga Manco *Universidad de Antioquia*

Resumen: El objetivo de este trabajo es desarrollar una estrategia didáctica, a partir del lenguaje de programación R, que apoye el proceso de enseñanza-aprendizaje de la Estadística, en particular, de regresión lineal múltiple. La estrategia didáctica se desarrolló a partir de la creación de un Dashboard en Shiny que analiza la distribución de la variable respuesta, las relaciones entre variables, el problema de multicolinealidad, el análisis de residuales y por último evalúa el aprendizaje obtenido con la interacción del Dashboard. A través de esta aplicación cualquier persona que esté comenzando en el estudio del modelado estadístico puede comprender el paso a paso a seguir para obtener una predicción o evaluar el efecto de alguna variable predictora en la variable respuesta.

Keywords: Dashboard, Regresión lineal múltiple, Shiny

Introducción

En la enseñanza-aprendizaje de la Estadística en el sistema de educación superior el uso de la computación presenta grandes ventajas con respecto a otros medios, debido a la posibilidad de interacción y de individualización que ofrecen. Con la computación pueden desarrollarse ambientes de aprendizaje donde el estudiante avanza a su propio ritmo y recibe refuerzo diferencial según sus capacidades y destrezas (Enoa, 2016).

En ambientes de aprendizaje con computación el aprendizaje activo, autónomo y colaborativo es un reto importante, lo que ha obligado a la revisión de las metodologías docentes utilizadas y a llevar a cabo un cuidadoso diseño de las actividades a realizar dentro y fuera del aula, así como de los correspondientes materiales. Por otra parte, la necesidad de feedback del alumno en un proceso de esta naturaleza obliga al planteamiento de una evaluación continua dinámica que proporcione rápidamente al estudiante información acerca de su desempeño (Pérez et al., 2019).

En esta línea de investigación, se ha iniciado el desarrollo de una serie de aplicaciones interactivas en Shiny con objeto de orientar el aprendizaje autónomo de los estudiantes del curso de Diseño de Experimentos y Análisis de regresión, así como evaluar el grado de consecución de conocimientos.

El trabajo presentado a continuación se presenta en cuatro secciones. En la primera sección se muestra la metodología, seguido de ello se muestra una descripción del Dashboard desarrollado. Por último, en la tercera y cuarta Sección se presentan los Resultados y Conclusiones.

* Autor de contacto: odaniel.rivera@udea.edu.co

Metodología

Los datos utilizados en el Dashboard están disponibles públicamente en la página del Anuario Estadístico de Antioquia del Departamento Administrativo de Planeación. Los datos extraídos de esta entidad pública corresponden a una serie de variables sociales, educativas, deportivas y de seguridad de las zonas rurales de Antioquia, Colombia para el año 2016.

Para el análisis exploratorio se utilizaron los paquetes **ggplot2** (Wickham, 2016) y **RcmdrMisc** (Fox, 2018). En la validación del modelo, específicamente en el análisis de residuales, se usaron funciones de los paquetes **car** (Fox and Weisberg, 2019), **nortest** (Gross and Ligges, 2015), **tseries** (Trapletti and Hornik, 2019) y **lmtest** (Zeileis and Hothorn, 2002). Finalmente, para la construcción del Dashboard se usaron los paquetes **shiny** (Chang et al., 2019), **shinythemes** (Chang, 2018) y **DT** (Xie, Cheng and Tan, 2019).

Dashboard

El Dashboard se encuentra disponible en <https://danielrivera1.shinyapps.io/Regression2/> y presenta el análisis de regresión lineal múltiple a través de cinco pasos. En el primer paso se analiza la distribución de la variable respuesta a través de gráficos y pruebas de normalidad. Por su parte, en el segundo paso se presenta el análisis exploratorio entre la variable respuesta y las variables predictoras. En el tercer paso se analiza el problema de multicolinealidad a través del factor de inflación de varianza. A su vez, en el cuarto punto se presenta el análisis de residuales para validar el modelo. Por último, se presenta una sección de evaluación a través de tres niveles de dificultad. También, el dashboard dispone de una sección de glosario en caso de que no se haya tratado un concepto de manera teórica. Una vista general del Dashboard se observa en la Figura 1

Didactic modeling process: Linear regression for a safety issue in rural areas of Antioquia - Colombia

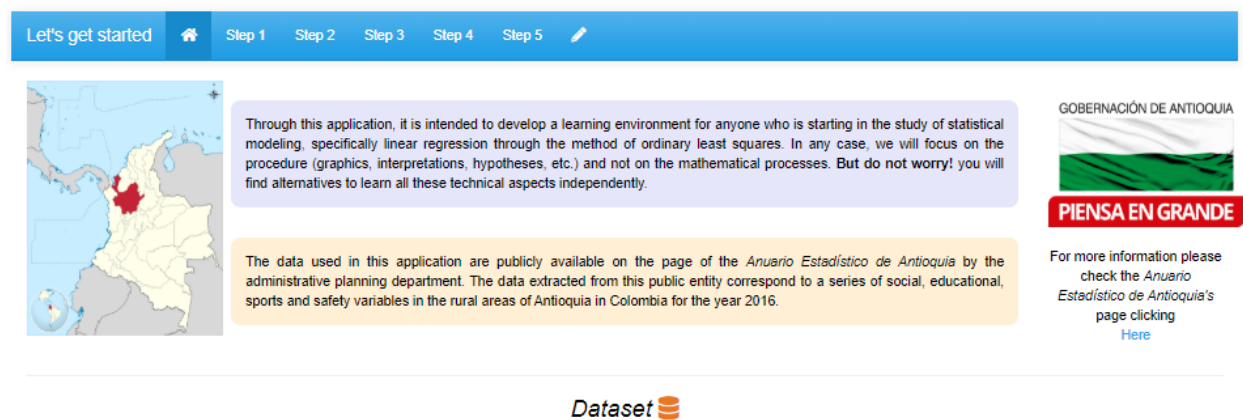


Figure 1: Dashboard para modelación estadística.

Resultados

La variable respuesta de interés es **Lesiones personales** ya que representa un problema de seguridad donde la integridad física de las personas se ve amenazada. Para el análisis de esta variable el

Dashboard muestra los gráficos de histograma, boxplot y qqplot, además, si es necesaria alguna transformación sobre la variable respuesta, el usuario puede seleccionar la potencia en un rango de -3 a 3. Todo el análisis gráfico va acompañado de los resultados de la prueba normalidad de Shapiro-Wilk. La figura 2 muestra la forma en la cual el usuario puede interactuar con la selección de la potencia.

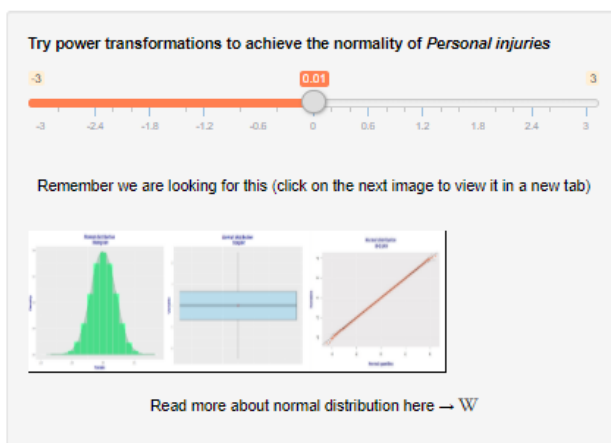


Figure 2: Selección de potencia.

En el análisis de relaciones entre la variable respuesta y las variables predictoras se presentan gráficos de dispersión y sus respectivos valores de correlación. Las variables seleccionadas para el análisis fueron: población proyectada, número de robos, número de accidentes de tráfico, cantidad de homicidios, número de estudiantes que desertan del colegio, sedes deportivas y cantidad de extorsiones.

En el análisis de relaciones lineales entre la variable respuesta y las variables predictoras se presenta el Factor de inflación de varianza (VIF) y un resumen estadístico del modelo de regresión de acuerdo a las variables predictoras seleccionadas. Esta fase de análisis es importante para verificar si existe algún efecto importante de la multicolinealidad en los valores de las estimaciones de los parámetros y su significancia dentro del modelo de regresión.

Para determinar las variables que son significativas dentro del modelo de regresión el Dashboard permite seleccionar las variables predictoras que se van a incluir dentro del modelo. Además, se puede seleccionar el método y criterio de selección de variables.

Conclusiones

El Dashboard desarrollado pretende apoyar el proceso de enseñanza-aprendizaje de estadística, en educación superior, y se basa en los contenidos desarrollados en un curso de Análisis de regresión. La aplicación es atractiva, intuitiva y motivadora para aquellas personas que deseen desarrollar habilidades en el análisis de modelos de regresión lineal múltiple.

References

- Chang, W. 2018. *shinythemes: Themes for Shiny*. R package version 1.1.2.
- Chang, W., J. Cheng, JJ. Allaire, Y. Xie and J. McPherson. 2019. *shiny: Web Application Framework for R*. R package version 1.3.2.
- Enoa, L. 2016. "Estrategia didáctica para el proceso de enseñanza-aprendizaje de la estadística mediante la herramienta estadística R." *Campus Virtuales* 5(2):112–121.
- Fox, J. 2018. *RcmdrMisc: R Commander Miscellaneous Functions*. <https://www.r-project.org>, <http://socserv.socsci.mcmaster.ca/jfox/>.
- Fox, J. and S. Weisberg. 2019. *An R Companion to Applied Regression*. Third ed. Thousand Oaks CA: Sage.
- Gross, J. and U. Ligges. 2015. *nortest: Tests for Normality*. R package version 1.0-4.
- Pérez, J.J., F. Arnaldos, M. T. Delfa, U.F. Martínez and L.M. Peris. 2019. "Docencia y evaluación en Estadística utilizando aplicaciones interactivas Shiny." *e-Publica* (24):37–55.
- Trapletti, A. and K. Hornik. 2019. *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-47.
URL: <https://CRAN.R-project.org/package=tseries>
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Xie, Y., J. Cheng and X. Tan. 2019. *DT: A Wrapper of the JavaScript Library 'DataTables'*. R package version 0.9.
URL: <https://CRAN.R-project.org/package=DT>
- Zeileis, A. and T. Hothorn. 2002. "Diagnostic Checking in Regression Relationships." *R News* 2(3):7–10.