

Actividad: Modelos estadísticos en R y creación de documentos dinámicos

Esta actividad forma parte de la evaluación del curso en la **convocatoria ordinaria**.

El envío consistirá en un fichero comprimido único (zip or rar) que contenga un documento PDF dinámico (creado usando *knitr* o *rmarkdown*, y referido más abajo como informe) y todos los ficheros fuente, incluyendo los ficheros de datos.

El informe no debe exceder el límite de 10 páginas. Esto implica realizar una selección de los gráficos y resultados de R que sean más relevantes para incluirlos.

La actividad se evaluará de acuerdo a los siguientes criterios: (i) originalidad; (ii) completitud y resolución correcta; (iii) descripción detallada del procedimiento de resolución y los resultados; (iv) presentación del documento.

Descripción de la tarea

Esta actividad consiste en analizar un conjunto de datos utilizando un modelo de regresión adecuado. El primer paso es encontrar un conjunto de datos para analizar, que preferiblemente sea original y, en cualquier caso, diferente de los utilizados en los ejemplos proporcionados en los materiales del curso.

Después de proporcionar una descripción del contexto de los datos, se debe formular una pregunta de investigación, que pueda responderse mediante la formulación y el ajuste de un modelo de regresión lineal múltiple o un modelo lineal generalizado, dependiendo del tipo de variable de respuesta.

El análisis debe abordar correctamente los siguientes aspectos:

- Motivación y objetivo del análisis: Incluyendo referencias completas al conjunto de datos utilizado (fuente y referencias de documentos/libros donde estos datos puedan haber sido utilizados antes).
- Formulación matemática del modelo elegido: Indicando claramente las suposiciones e hipótesis del modelo.
- Ajuste del modelo: Estimación e interpretación de los parámetros del modelo.
- Evaluación de la bondad de ajuste.

- Inferencia básica relacionada con el modelo (intervalos de confianza, contrastes de hipótesis).
- Diagnóstico del modelo: Incluyendo la identificación y tratamiento de posibles observaciones anómalas y/o influyentes.
- Posible simplificación del modelo: A través de la selección de variables explicativas utilizando procedimientos automáticos adecuados (por ejemplo, selección paso a paso, *stepwise*).

Todos los puntos anteriores deben analizarse a fondo. Dependiendo de la complejidad del modelo o del problema en cuestión, esto puede requerir la consulta de referencias bibliográficas, como las proporcionadas en los materiales del curso o en las fuentes indicadas a continuación.

Será necesario incluir tanto la ejecución en R (código y resultados) como los comentarios explicativos. Para ello, se elaborará un documento dinámico donde todos los elementos (código, salida y explicaciones) estén adecuadamente integrados y descritos. Dado que hay una limitación de páginas (no más de 10), se deben seleccionar los resultados y gráficos más relevantes para el informe. Por ejemplo, algunos gráficos o resultados muy extensos pueden no mostrarse y en su lugar añadir un breve comentario de lo más relevante.

Estructura del informe

1. **Introducción:** Describir la motivación y los objetivos del análisis, relacionándolos con el contenido del curso. Explicar por qué el conjunto de datos elegido es de interés y cómo permite aplicar los conceptos desarrollados los materiales del curso en clase, o por qué es necesario ampliarlos utilizando algunas referencias bibliográficas.
2. **Los datos:** Describir los datos utilizados, incluyendo referencias completas: fuente de los datos, dónde obtenerlos y notas sobre trabajos previos en los que estos datos puedan haber sido utilizados.
3. **El modelo:** Proporcionar la formulación matemática del modelo elegido, incluyendo las suposiciones (hipótesis) del mismo.
4. **Análisis estadístico:** Describir el ajuste del modelo, escribiendo la ecuación del modelo ajustado y analizando la bondad de ajuste, inferencia básica relacionada con el modelo, diagnóstico del modelo y posible simplificación del mismo.

5. **Conclusiones:** Resumir el análisis realizado y destacar los resultados más relevantes. También mencionar las limitaciones del análisis realizado y sugerir posibles mejoras o extensiones.
6. **Referencias:** Incluir las fuentes y enlaces que se hayan consultado y que sean relevantes para el trabajo. Estas pueden ser libros, artículos, tutoriales, blogs o apuntes del curso que hayan contribuido a la comprensión o implementación.

Referencias recomendadas y links

Manuales para complementar y ampliar los contenidos del curso:

- Faraway, J.J. (2004). *Linear Models with R*. Chapman & Hall/CRC. Texts in Statistical Science Series.

Se puede obtener una versión gratuita desde <https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>.

- Faraway, J.J. (2006) *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC. Texts in Statistical Science Series.

Se puede obtener una versión gratuita desde <https://englianhu.wordpress.com/wp-content/uploads/2016/01/faraway-extending-the-linear-model-with-r-e28093-2006.pdf>

- García-Portugués, E. (2025). *Notes for Predictive Modeling*. <https://bookdown.org/egarpor/PM-UC3M/PM-UC3M.pdf>

Algunos links de donde de pueden obtener datos:

- OpenIntro: <https://www.openintro.org/data/>
- TidyTuesday: <https://github.com/rfordatascience/tidytuesday>
- FiveThirtyEight. GitHub repo: <https://data.fivethirtyeight.com/>. R package: <https://fivethirtyeight-r.netlify.app/>
- Kaggle: <https://www.kaggle.com/datasets>
- Data is Plural: <https://www.data-is-plural.com/>