

Análisis de Regresión (2021-3)



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Especialización en Estadística Aplicada

Prof. [Sébastien Lozano Forero](mailto:slozanof@libertadores.edu.co) (slozanof@libertadores.edu.co)

Estimada(o) estudiante, sea bienvenida(o) al espacio académico **Análisis de regresión**, versión 2021-3, de la Especialización en Estadística Aplicada en su metodología virtual de la Facultad de Ingeniería y Ciencias Básicas de la Fundación Universitaria Los Libertadores. En el presente *Jupyter Notebook* encontrará información relevante para el correcto desarrollo de nuestro espacio académico. A continuación, una pequeña tabla de contenidos.

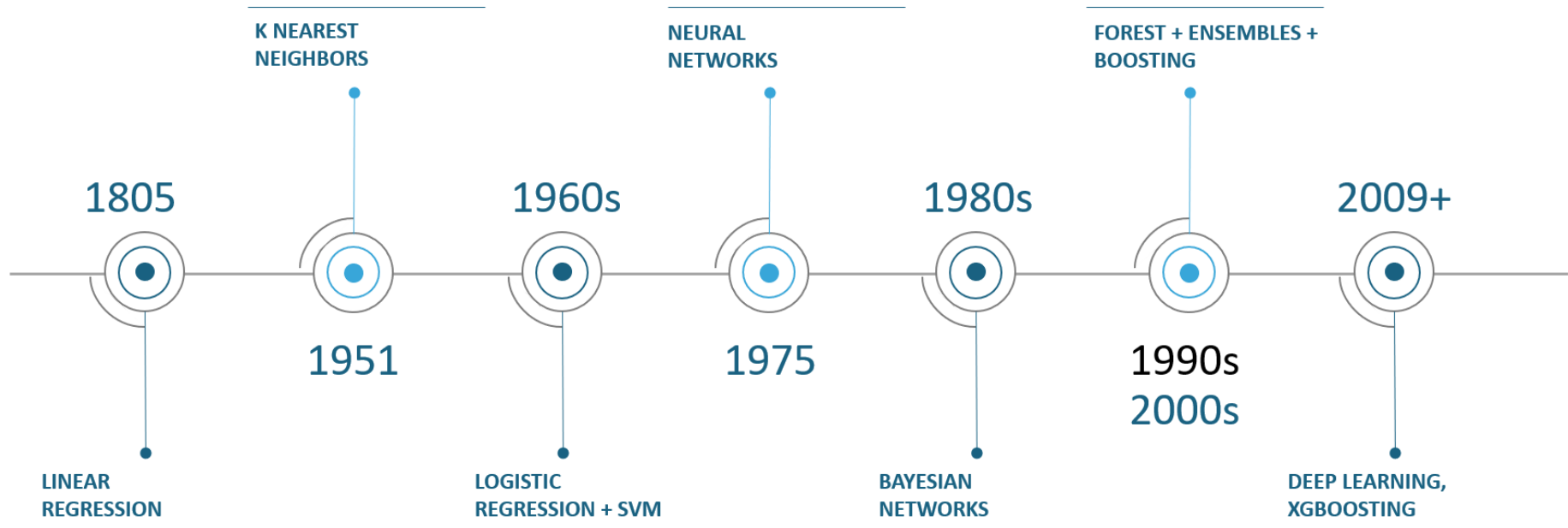
Tabla de contenidos

- [Presentación](#)
 - [Contexto](#)
 - [Contenido](#)
 - [Preliminares](#)
 - [Modelo de Regresión Lineal Simple](#)
 - [Modelo de Regresión Lineal Múltiple](#)
 - [Modelo de Regresión Logística](#)
 - [Evaluación](#)
 - [Bibliografía recomendada](#)
- [Dinámica](#)
 - [Encuentros Sincrónicos](#)
 - [Blackboard](#)
 - [Slack](#)
- [Requerimientos](#)
- [Código de Ética](#)

- [Repositorio](#)

Presentación

El modelo de regresión lineal fue inicialmente introducido por el matemático y físico alemán Johann Carl Friedrich Gauss a inicios del siglo XIX (~1800) para poder predecir la trayectoria de cuerpos celestes. Ha sido objeto de estudio de grandes pensadores de la matemática y la estadística a través los siglos XIX, XX y XXI; y es, hoy en día uno de los modelos estadísticos más (muchos dirán que el más) importantes en términos de usos e ideas. Es decir, muchos de los algoritmos más populares usados en Machine Learning actualmente no son regresiones, está claro, pero se deben a las ideas de la misma.



(tomado de [jnexttech](#))

En este contexto, puede ser natural la pregunta **¿Es el análisis de regresión relevante a más de 200 años de su creación?**. Para la respuesta (que es un contundente **SÍ**), considere las siguientes fuentes de información:

- Revista [Harvard Business Review](#)
- Revista [Analytics India Magazine](#)
- Revista [Smallbusiness](#)

- Revista [Lancet](https://www.lancet.com/article/S0140-6736(16)30294-7/fulltext)30294-7/fulltext)

Evaluación

- Por tratarse de un curso de carácter más aplicado, se evitará el esquema de exámenes escritos.
- Se realizarán dos (2) talleres (entregas: Domingo **29 de Agosto** y Domingo **5 de Septiembre**). Entregable esperado: Reporte en pdf (o Jupyter Notebook) con la solución a un problema planteado.
- Se realizará un proyecto de aplicación en formato de artículo enviado. Podrán ser grupos de máximo tres personas con estricta política anti parásitos. (Entrega: Domingo **12 de Septiembre**). Entregable esperado: archivo pdf (Únicamente un archivo pdf) con la presentación de un problema, desarrollo y solución de alguna base de datos de interés del grupo de estudiantes.
- Podremos considerar bonos en clase por participación y demás.



Así, la nota final del espacio académico será calculada como:

$$\text{Calificación Final} = 0.3 \times \text{Calificación Taller 1} + 0.3 \times \text{Calificación Taller 2} + 0.4 \times \text{Calificación del Proyecto de Aplicación.}$$

Fuentes de datos

En caso de no contar con datos propios requerir alguna base de datos para análisis, es posible considerar las siguientes fuentes de bases de datos.

- <https://www.kaggle.com/> (repositorio online colaborativo de bases de datos. También cuenta con cursos, retos y comunidad. Altamente recomendado)
- <https://datos.bancomundial.org/>
- <https://www.datos.gov.co/>
- <https://datasetsearch.research.google.com/> (ver <https://www.datacentric.es/blog/google-open-data-datasets-gratuitos/>)

kaggle



Otras opciones están citadas en este [artículo](#). Comentario personal: recomiendo mucho que tengan presente a *medium* (portal que publica el anterior artículo). Allí se publican cosas muy interesantes de asuntos muy variados en varios idiomas. Incluso, si fuera su deseo, es posible escribir y publicar artículos allí con relativamente pocos filtros.

Bibliografía Recomendada

Como siempre debe pasar en un espacio académico, deberán haber referencias bibliográficas para poder consultar el formalismo de la metodología que se está empleando. En el caso de nuestro espacio académico, serán las siguientes (El trabajo sucio ya está hecho [acá](#)).



Dinámica

A continuación, algunos elementos para tener presente sobre la dinámica de nuestro espacio académico. Tenga presente que el espacio académico va desde el **Lunes 16 de Agosto de 2021** hasta **Domingo el 12 de noviembre de 2021**.

Encuentros Sincrónicos

Los encuentros sincrónicos serán los días **martes y Jueves en horario de 18h00 a 20h00**. Recuerde que al tratarse de un programa concebido en metodología virtual, éstos encuentros suponen un contacto *complementario* a su estudio independiente. De la misma forma, se propone que la dinámica del espacio parta de suministrar algunos *jupyter notebooks* o archivos pdf para proveer a las y los estudiantes de los materiales del caso para que puedan acompañar tal encuentro.

La asistencia a los encuentros sincrónicos no es obligatoria, pero sí altamente recomendada. Tenga presente que si no es posible asistir en estos días a estos horarios, todas las sesiones quedarán grabadas y estarán disponibles en la plataforma, junto con los materiales del caso. Será responsabilidad de cada estudiante que no asista al encuentro sincrónico, consultar el video de la sesión y revisar el material disponible en plataforma.

Blackboard

Al igual que la totalidad de los demás espacios académicos que componen el programa, la principal plataforma mediadora para todo el proceso académico, será la plataforma Blackboard. Por ésta, tendremos los encuentros sincrónicos del caso, estarán disponible para consulta los videos asincrónicos, se disponibilizarán los materiales del curso, se hará el depósito de entregables, se retroalimentará, etc. Sin embargo, y a manera de complementar su formación profesional, considero que puede ser interesante presentar, y eventualmente usar, algunas herramientas tecnológicas que actualmente son populares en los entornos de trabajo de desarrollo de software, análisis de datos (Data Science), entre otros.



Slack

Slack puede entenderse como un híbrido entre correo electrónico, servicio de mensajería instantánea y red social profesional que, considero, aumenta el involucramiento de los estudiantes en cursos virtuales. No en vano, hay comunidades enteras alrededor del mundo de temas muy variados que se consolidan usando Slack. Permite la creación de un entorno de trabajo para un determinado grupo de personas y a través de canales (temas) pre-establecidos.



Habrá un espacio de trabajo para las y los estudiantes de este espacio académico por Slack. Por tanto, se recomienda inmensamente la descarga de esta herramienta, idealmente tanto en pc personal como en smartphone, de acuerdo al sistema operativo del caso:

1. Mac: <https://slack.com/downloads/mac>
2. Windows: <https://slack.com/downloads/windows>
3. Android: <https://slack.com/downloads/android>

La invitación para usar nuestro espacio de trabajo será enviada al **correo institucional**. Por tanto, Slack se configura como herramienta no institucional pero sí de uso oficial. En dicho espacio de trabajo, encontrará los canales:

- **#clase**: Se recomienda el uso de canal para adelantar el diálogo académico respectivo a las sesiones sincrónicas.
- **#general**: discusiones sobre el espacio, anuncios, etc,
- **#musica**: recomendaciones musicales
- **#networking**: presentación de cada estudiante (Se recomienda compartir perfil de LinkedIn).
- **#recursos**: ayudas para el aprendizaje de los contenidos de este espacio
- **#social**: (cualquier cosa es buen tema para este canal)

Para el uso de Slack se recomienda:

- Siempre ser absolutamente respetuosa o respetuoso del espacio académico y de sus participantes
- Seguir las reglas que encontrarán en Slack al momento de entrar con la invitación
- Actualizar el perfil de cada uno con nombre, fotografía y datos de contacto
- Tomar los recursos que se consideren importantes durante el espacio. Una vez concluido el espacio académico, se cerrará el espacio de trabajo.

Requerimientos tecnológicos

Para un correcto desarrollo del curso, es necesario que los estudiantes tengan disponibles las siguientes herramientas para poder acompañar los encuentros sincrónicos.

Slack

Ver esta [sección](#)

R y Rstudio

Descargar e instalar, en estricto orden y de acuerdo al sistema operativo correspondiente, las siguientes distribuciones



1. R (consola original): <https://www.icesi.edu.co/CRAN/>
2. Rstudio (Interfaz mucho más agradable): <https://rstudio.com/products/rstudio/download/>

Cómo opción alternativa, se tiene <https://rstudio.cloud/>, que es una versión en la nube de R (no es la más recomendada pues, al ser gratuita, la velocidad depende de la demanda de uso en el servidor principal).

Algunos recursos adicionales recomendados para aprender R (además de los [atajos](#) de Rstudio):

- <https://learn.datacamp.com/>
- <https://www.edx.org/search?q=R>
- <https://www.coursera.org/search?query=R&>
- <https://www.udemy.com/courses/search/?q=R>
- <https://online-learning.harvard.edu/course/statistics-and-r?delta=1>
- <https://www.codecademy.com/learn/learn-r>
- <https://cognitiveclass.ai/courses/r-101>
- <https://www.lynda.com/R-training-tutorials/1570-0.html>

Anaconda (Opcional)

Anaconda es una de las principales herramientas actuales en los círculos profesionales de **Data Science**. Genera entornos virtuales (entornos que permiten construir de forma independiente a las especificaciones propias de cada máquina) para poder utilizar, entre otros, herramientas

como *Jupyter Notebooks*. También es muy usado para programar en código Python (recomendaría ver este [tutorial](#) para aprender a correr los jupyter notebooks en código R). Descargar Anaconda: <https://www.anaconda.com/>

Jupyter Notebooks (Opcional)

Jupyter Notebook es una aplicación que permite crear y compartir documentos web en formato JSON que siguen un esquema versionado y una lista ordenada de celdas de entrada y de salida. Estas celdas albergan, entre otras cosas, código, texto (en formato Markdown), fórmulas matemáticas y ecuaciones, o también contenido multimedia. El programa se ejecuta desde la aplicación web cliente que funciona en cualquier navegador estándar. Los documentos creados en Jupyter pueden exportarse, entre otros formatos, a HTML, PDF, Markdown o Python y también pueden compartirse con otros usuarios por correo electrónico, utilizando Dropbox o GitHub. Para saber más sobre qué es un jupyter notebook, entre [aquí](#).



Código de Ética

La integridad académica es fundamental para la institución, todas sus dependencias y autoridades académicas en todos los niveles. El compromiso de integridad académica parte de mostrar respeto por el trabajo propio y el trabajo ajeno.

Respuestas a tareas, cuestionarios y exámenes deberán ser **trabajo propio**. No está permitido compartir soluciones de tareas, códigos o exámenes con entre estudiantes. Esto incluye cualquier cosa escrita por cada estudiante, como también cualquier solución. Habrá *Tolerancia cero* con los casos de plagio (Entenderemos plagio como la copia de palabras, ideas o cualquier otro material de otra fuente sin el debido otorgamiento de créditos).

Repositorio Github del material del espacio académico

Git es un sistema de control de versiones distribuido con licencia open source diseñado para manejar todo tipo de proyectos en todas sus etapas de desarrollo. En la práctica, irá registrando las novedades en las versiones del proyecto de interés del autor, permitiendo el trabajo colaborativo remoto. Para saber más de Git, vea [aquí](#)



GitHub es un repositorio que permite alojar proyectos utilizando el sistema de control de versiones Git. Se utiliza principalmente para la creación de código fuente de programas de ordenador. Para saber más de Github, vea [aquí](#)

Todo el material del espacio académico estará disponible para descarga desde el repositorio de [Github](#). Por favor, consulte la descripción de los archivos.

In []: