

Análisis de regresión

10 de agosto
(semana 1)

Plan de trabajo

1. Características del curso
2. Datos del docente, horario de atención
3. Presentación de estudiantes
4. Evaluación y contenido
5. Garantías académicas
6. Introducción a la temática del curso

Características del curso

- ❖ Código: 2016360, grupo 1
- ❖ Agrupación: Disciplinar (obligatoria)
- ❖ Créditos: 4
- ❖ Tipología: Teórica
- ❖ Prerrequisitos: Inferencia estadística. Recomendable: cálculo multivariable, álgebra lineal/matricial y análisis multivariado.
- ❖ Horario y modalidad: Miércoles y viernes de 7am a 9am, **presencial** y/o virtual remota. Presentaciones magistrales, desarrollo de ejercicios y actividades por parte de estudiantes.
- ❖ Uso de software R (Rstudio) o Python.

Objetivos del curso

Objetivo general

Familiarizar al estudiante con los conceptos y fundamentos de modelación predictiva y explicativa de la relación entre diferentes variables.

Objetivos específicos

- ☐ Identificar la estructura general de los modelos de regresión.
- ☐ Teorizar la construcción, evaluación y selección de un modelo de regresión adecuado.
- ☐ Brindar los elementos de análisis estadístico adecuados para una correcta implementación, presentación y discusión de los hallazgos.

Datos docente y horario de atención

Mario Enrique Arrieta Prieto
Ing. Industrial y estadístico, UNAL
MS. En Estadística, UNAL
MS en matemática aplicada, RPI, NY, EEUU
Ph.D. en DSES, RPI, NY, EEUU
Docente Depto. Estadística
mearrietap@unal.edu.co

- ❖ Horario de atención: Miércoles y viernes (11am a 12:30pm, Of. 325 FEM) y martes (10am a 12m, Zoom), o en otro horario previamente acordado.
- ❖ La respuesta a inquietudes por correo se dará en **1-2 días hábiles** en horario de oficina.

Evaluación

Item	Porcentaje
Primer parcial (sem. 6)	20%
Segundo parcial (sem. 11)	20%
Tercer parcial (sem. 16)	20%
Proyectos (sem. 4, 9 y 16)	25%
<i>*Ver opciones</i>	15%

❖ Cada examen tendrá un componente en clase/MOODLE (70%) y la entrega vía correo de unos ejercicios por CLASSROOM (30%). El profesor podrá convocar a sustentaciones orales que valdrán el 50% del parcial junto con la parte de MOODLE que valdrá otro 50%.

❖ ***Máximo entre:**

- a) Opción 1: 15% quices.
- b) Opción 2: 10% quices; 5% participación

Evaluación (II)

- ❖ Se dejarán varios tipos de talleres (no calificables) de preparación para los exámenes, y de allí saldrán algunos ejercicios a entregar (uso de un editor de texto y ecuaciones: Latex, Word, Google docs).
- ❖ La participación será tomada en cuenta cuando el estudiante participe voluntariamente o cuando el profesor realice preguntas. Sólo se tendrá en cuenta el número de participaciones positivas.
- ❖ La calificación será asignada de acuerdo al número relativo de participaciones respecto a los demás estudiantes y tendrá una valoración final de entre 2.5 y 5.0.

Evaluación (III)

- ❖ El proyecto buscará complementar la formación teórica con el uso de software, con el fin de afianzar conceptos relacionados con la visualización, aplicación y explicación de resultados en contextos reales.
- ❖ En clase, se buscará familiarizar a los estudiantes con las diferentes alternativas disponibles en R para la modelación.
- ❖ La elección del software para el proyecto es de libertad del grupo de trabajo (entre R y Python).

Horas o Créditos	Tema o Capítulo
Módulos 1-2: 4 semanas	<p>Introducción al análisis de varias variables en estadística.</p> <ul style="list-style-type: none"> • Diferencia entre correlación y causación. • Diferentes coeficientes de asociación. • Pruebas de independencia para variables cualitativas. • Prueba ANOVA de un factor. • Análisis de regresión lineal simple. <p>Modelo lineal general desde el punto de vista matricial.</p>
Módulos 3-4: 4 semanas	<p>Regresión lineal múltiple</p> <ul style="list-style-type: none"> • Supuestos y principales resultados. • Construcción de un modelo de regresión lineal múltiple. • Estimación e inferencia en el modelo de regresión lineal múltiple. • Validación de supuestos. <p>Mecanismos remediales en caso de violación de supuestos.</p> <ul style="list-style-type: none"> • Datos atípicos e influyentes en el modelo. • Transformaciones estabilizadoras de varianza. • Mínimos cuadrados ponderados y generalizados.
Módulos 5-6: 4 semanas	<p>Modelos lineales desde el punto de vista de la analítica.</p> <ul style="list-style-type: none"> • Indicadores de capacidad predictiva de un modelo. • Mecanismos de selección de variables. • Criterios de información. • Regresión ridge y LASSO. • Uso de componentes principales. <p>El modelo lineal general.</p> <ul style="list-style-type: none"> • Problemas de inferencia en el modelo lineal general. • Funciones estimables. <p>Alternativas al modelo lineal general.</p> <ul style="list-style-type: none"> • Modelos de efectos aleatorios y multinivel. • Modelos no paramétricos y semiparamétricos.
Módulo 7-8: 4 semanas	<p>Alternativas al modelo lineal general.</p> <ul style="list-style-type: none"> • Modelos lineales generalizados (binomial, Poisson). • Modelos de aprendizaje automático (árboles de decisión, máquinas de soporte de regresión, redes neuronales)

Bibliografía

- (1) Hogg, Robert V., and Allen T. Craig. *Introduction to mathematical statistics. (5th edition)*. Englewood Hills, New Jersey, 1995.
- (2) Casella, George, and Roger L. Berger. *Statistical inference*. Vol. 2. Pacific Grove, CA: Duxbury, 2002.
- (3) **Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley and Sons, 2021.**
- (4) Johnson, Richard Arnold, and Dean W. Wichern. *Applied multivariate statistical analysis*. Vol. 6. London, UK:: Pearson, 2014.
- (5) Wooldridge, Jeffrey M. *Introductory econometrics: A modern approach*. Cengage learning, 2015.
- (6) **James, Gareth, et al. *An introduction to statistical learning*. Vol. 112. New York: springer, 2013.**
- (7) Hinkelmann, Klaus, and Oscar Kempthorne. *Design and analysis of experiments, volume 1: Introduction to experimental design*. Vol. 1. John Wiley and Sons, 2007.
- (8) Demidenko, Eugene. *Mixed models: theory and applications with R*. John Wiley and Sons, 2013.
- (9) Dobson, Annette J., and Adrian G. Barnett. *An introduction to generalized linear models*. Chapman and Hall/CRC, 2018.

Fechas importantes y garantías académicas

- ❖ Inicio semestre: 8 agosto
Fin semestre: 3 diciembre
Reporte calificaciones en SIA: 10 diciembre, 5pm
Adiciones y cancelaciones sin pérdida de créditos: primeras dos semanas.
- ❖ Solo se mantuvo como garantía el poder ver carga inferior a la mínima sin solicitud (Acuerdo CSU 008, 2008).
- ❖ Acuerdos de clase y programa sujetos a contingencias y dificultades.
 - No se tiene en cuenta asistencia, pero algunos quices en clase.
 - Participación y asistencia a clase deseada (notas de clase disponibles en drive).
 - Exámenes con ventanas de tiempo mayores. **¡Ojo con el plagio!**
 - Disponibilidad de comunicación frente a las dificultades que ustedes puedan tener.
 - Sesiones extra para complementar temas/reponer festivos.