

Analítica de datos aplicada a estudios sobre desarrollo
CIDE- 4 créditos. Período académico: 2025-20
Profesoras(es): Giovany Babativa-Márquez



Analítica de datos aplicada a estudios sobre desarrollo

Equipo docente en orden de participación:

Giovany Babativa-Márquez, j.babativamarquez@uniandes.edu.co

Felipe Tautiva, monitor, davidtautiva2806@gmail.com

Modalidad: Semipresencial

Horario: (cada 15 días)

- Modalidad presencial: jueves de 6:30 am a 9:20am
- Modalidad Virtual: jueves de 7:00 am a 8:50am

Salón: Pendiente por confirmar

Sesiones virtuales en este [link de zoom](#):

I. DESCRIPCIÓN

El proceso de investigación puede ser entendido como el ensamble de cuatro fases: formulación de la(s) pregunta(s) de investigación (y todo lo asociado al diseño de la investigación), recopilación de datos, análisis de datos y redacción. Este curso se centra en la fase de análisis de datos.

Exploraremos y comprenderemos los métodos y herramientas cuantitativas ampliamente utilizadas en los estudios sobre desarrollo en las últimas dos décadas. Al final del curso los estudiantes estarán en capacidad de aplicarlos en sus proyectos de investigación o consultoría.

Los estudios sobre desarrollo requieren de la ejecución de la investigación interdisciplinaria. El valor de este tipo de investigación radica en ver lo que una disciplina por sí sola no logra percibir, y en incorporar métodos complementarios para brindar soluciones a problemas complejos.

El curso está compuesto por nueve sesiones que buscan introducir los aspectos básicos y aplicados del análisis exploratorio de datos – AED, del análisis de regresión y del análisis multivariado como herramientas para el análisis de datos que permitan responder las preguntas de investigación. Comenzaremos con las ideas y métodos más simple, luego trabajaremos otras técnicas un poco más robustas. A medida que el curso va avanzado, el trabajo será más desafiante y relevante para los y las estudiantes. Es importante resaltar que el curso no presupone una amplia y sólida base matemática, se requerirán conceptos básicos como derivadas, logaritmos (calculo diferencial), probabilidad, tipos de variables. Sin embargo, los conceptos centrales no serán los números sino las variables y sus

relaciones. Por lo tanto, el objetivo de este módulo es comprender la relación entre las diferentes variables que afectan la realidad mundial, nacional y local.

Al final del curso, los estudiantes dominarán los aspectos básicos del análisis de datos y su aplicabilidad en los estudios de desarrollo en diferentes contextos. Este módulo dotará a los estudiantes con las capacidades básicas para aplicar estos métodos de investigación a problemas del mundo real y para evaluar su uso en investigaciones publicadas. También, permitirá a los y las estudiantes iniciar un proyecto cuantitativo, le dará bases para sus futuros estudios de posgrado y mejorará sus perspectivas laborales.

El curso es semipresencial con una duración de 16 semanas; comprende actividades sincrónicas, asincrónicas y presenciales.

II. OBJETIVOS DE ENSEÑANZA

1. Familiarizar a los estudiantes en el uso de herramientas para el análisis cuantitativo y espacial de datos aplicada a los estudios de desarrollo.
2. Presentar las cuestiones prácticas y críticas en el uso de métodos de análisis cuantitativo de datos.
3. Introducir a los participantes en el uso de herramientas para el análisis cuantitativo.
4. Mejoramiento del aprendizaje independiente: aprenderán cómo podría pensar y abordar sus propios temas de investigación de forma cuantitativa;
5. Escritura: los estudiantes aprenden a informar y discutir los resultados del análisis cuantitativo.

III. RESULTADO DE APRENDIZAJE

Al finalizar el curso, los y las estudiantes estarán en capacidad de:

- Aplicar los principales métodos de análisis de regresión.
- Aplicar los principales métodos de análisis multivariado.
- Combinar distintas herramientas para el análisis cuantitativo.
- Uso de programas estadísticos para analizar datos cuantitativos.

IV. METODOLOGÍA

Es un curso de nivel intermedio, para el que los conceptos e intuiciones básicas de estadística son requeridos (por ejemplo, estadística descriptiva, teoría de la probabilidad, estadística inferencial, análisis de correlación, análisis de regresión bivariada y múltiple).

El enfoque del curso es práctico, a través de ejercicios hechos con los estudiantes. Usaremos el software estadístico R Studio. No nos concentraremos en demostraciones y pruebas formales. El curso tiene un énfasis aplicado.

En cada módulo los estudiantes realizarán talleres en los que se aplicará la temática abordada. Este es un curso de cuatro créditos por lo que se recomienda dedicarle aproximadamente doce horas semanales, incluyendo el trabajo independiente, individual o grupal, y las actividades que se describen a continuación para cada uno de los módulos.

Trabajo sincrónico:

Los encuentros sincrónicos están orientados a la presentación de los conceptos que aborda cada sesión, el debate de algunas preguntas sugeridas con base en las lecturas, y la presentación de algunas funcionalidades del software que logran apoyar esta aproximación en un caso específico.

Para el desarrollo de cada clase habrán asignadas lecturas, obligatorias y complementarias. Esto se contempla como parte del trabajo asincrónico previo a cada encuentro sincrónico.

Trabajo asincrónico:

El trabajo asincrónico se desarrollará de forma individual. Comprende las lecturas asignadas, la revisión de funcionalidades del software, la consulta de información pertinente, y el desarrollo y entrega del taller correspondiente para cada sesión.

Este trabajo asincrónico está acompañado de recursos tales como:

- Funcionalidades (R Studio)
- Referencias a casos específicos para el ejercicio aplicado
- Recursos electrónicos

Se tiene previsto una dedicación de tres horas para el desarrollo de cada guía aplicada. Estas se podrán desarrollar en encuentros asincrónicos con otros estudiantes. Se tendrá un espacio en Zoom acordado con los participantes para poder tener el acompañamiento del tutor académico en estas sesiones.

V. CONTENIDO TEMÁTICO POR SESIÓN

Pasos previos necesarios para la primera sesión

El curso utilizará el software R (software de acceso abierto). Es importante que cada estudiante instale en su computador el software estadístico R y R-Studio. Disponible aquí: <https://posit.co/download/rstudio-desktop/>

Las sesiones presenciales se realizarán en una sala de computadores.

Sesión 1: *Giovany Babativa-Márquez*

Semana 1 – **agosto 07** (*Sesión de autoaprendizaje*): Comprenderemos los aspectos clave de la visualización de datos e haremos una introducción al software R.

Ver video:

Bibliografía

- R for data science. Section 3. Data visualization. Disponible aquí: <https://r4ds.had.co.nz/>
- Dougherty, J. and Ilyankou, I. (2022) Hands-On Data Visualization. Interactive Storytelling from Spreadsheets to Code. Section 6. Chart Design Principles. Disponible aquí: <https://handsondataviz.org/>
- Williams, G. (2022) Data Science Desktop Survival Guide. Togaware. Disponible aquí: <https://onepager.togaware.com/introducing-rstudio.html> Chapter 2 *introducing R*.
- Herramienta de Excel XY Charterl label. Disponible aquí: <http://www.appspro.com/Utilities/ChartLabeler.htm>

Semana 2. *Encuentro virtual, asistencia voluntaria, la sesión la imparte el docente, el video queda grabado. agosto 14 a las 7 am. Link:*

Sesión 2: *Giovany Babativa-Márquez*

Semana 3. – **agosto 21** (*Híbrida: presencial o virtual 6:30 am*): *Visualización de datos y análisis exploratorio de datos*

Bibliografía

- Peng, R. and Matsui, E. (2017) *The Art of Data Science*. Chapter 4 Exploratory Data Analysis and Chapter 5 Using Models to Explore Your Data. Disponible aquí: <https://bookdown.org/rdpeng/artofdatascience/>

Semana 4. Taller 1. Taller visualización de datos y análisis exploratorio. Se publicará oportunamente en BloqueNeon junto con los términos de referencia.

Fecha de entrega: agosto 28, a la media noche

Sesión 3: Giovany Babativa-Márquez

Semana 5 – septiembre 04 (presencial, 6:30am): Modelo de regresión lineal

Bibliografía

- Çetinkaya-Rundel, M. and Hardin, J. (2021) Introduction to modern statistics. Sections of Regression modeling: 7, 8, 9 y 10. Disponible aquí: <https://openintro-ims.netlify.app/>

Semana 6. Taller 2 sobre Modelo de regresión lineal múltiple. Se publicará oportunamente en BloqueNeon junto con los términos de referencia.

Fecha de entrega: septiembre 24 a la media noche.

Sesión 4: Giovany Babativa-Márquez

Semana 7 – septiembre 18 (Virtual, 7am). Modelo de regresión lineal múltiple – parte 2

Bibliografía

- Çetinkaya-Rundel, M. and Hardin, J. (2021) Introduction to modern statistics. Sections of Regression modeling: 7, 8, 9 y 10. Disponible aquí: <https://openintro-ims.netlify.app/>

Examen 1. octubre 03. Oportunamente se publicarán las instrucciones en BloqueNeon.

Sesión 5: Giovany Babativa

Semana 8 – septiembre 25 (Híbrida: Presencial y Virtual, 6:30 am). Análisis de regresión no lineal

Bibliografía

- Çetinkaya-Rundel, M. and Hardin, J. (2021) Introduction to modern statistics. Sections of Regression modeling: 7, 8, 9 y 10. Disponible aquí: <https://openintro-ims.netlify.app/>

Semana 9. Taller 3 análisis de regresión no lineal. Se publicará oportunamente en BloqueNeon junto con los términos de referencia.

Fecha de entrega: octubre 08 a la media noche.

Sesión 6: Giovany Babativa

Semana 11 – octubre 09 (virtual 7am): Introducción a las técnicas multivariadas -análisis de componentes principales.

Bibliografía

- Husson, F., Lê, S., & Pagès, J. (2017). Exploratory multivariate analysis by example using R. CRC press.

- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis 6th Edition.
- Aldás Manzano, J., & Uriel Jiménez, E. (2017). Análisis multivariante aplicado con R. Ediciones Paraninfo, SA.
- https://bookdown.org/brian_nguyen0305/Multivariate_Statistical_Analysis_with_R/
- https://bookdown.org/jsalinas/tecnicas_multivariadas/presentacion.html
- <https://bookdown.org/mpfoley1973/data-sci/>

Semana 12. Taller 4 análisis de componentes principales. Se publicará oportunamente en BloqueNeon junto con los términos de referencia.

Fecha de entrega: octubre 22 a la media noche.

Sesión 7: Giovany Babativa

Semana 13- octubre 23 (virtual 7am): Introducción a las técnicas multivariadas: análisis de correspondencia simple y múltiple.

Bibliografía

- Husson, F., Lê, S., & Pagès, J. (2017). Exploratory multivariate analysis by example using R. CRC press.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis 6th Edition.
- Aldás Manzano, J., & Uriel Jiménez, E. (2017). Análisis multivariante aplicado con R. Ediciones Paraninfo, SA.
- https://bookdown.org/brian_nguyen0305/Multivariate_Statistical_Analysis_with_R/
- https://bookdown.org/jsalinas/tecnicas_multivariadas/presentacion.html
- <https://bookdown.org/mpfoley1973/data-sci/>
- https://bookdown.org/brittany_davidson1993/bookdown-demo/

Sesión 8: Giovany Babativa

Semana 14- noviembre 06 (presencial 6:30 am): Análisis de correspondencia simple y múltiple.

Bibliografía

- Husson, F., Lê, S., & Pagès, J. (2017). Exploratory multivariate analysis by example using R. CRC press.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis 6th Edition.
- Aldás Manzano, J., & Uriel Jiménez, E. (2017). Análisis multivariante aplicado con R. Ediciones Paraninfo, SA.
- https://bookdown.org/brian_nguyen0305/Multivariate_Statistical_Analysis_with_R/

- https://bookdown.org/jsalinas/tecnicas_multivariadas/presentacion.html
- <https://bookdown.org/mpfoley1973/data-sci/>
- https://bookdown.org/brittany_davidson1993/bookdown-demo/

Examen 2. noviembre 14. Oportunamente se publicarán las instrucciones en BloqueNeon.

Sesión 9: Giovany Babativa

Semana 15- noviembre 20 (virtual 7am): Análisis de conglomerados (clúster).

Bibliografía

- Husson, F., Lê, S., & Pagès, J. (2017). Exploratory multivariate analysis by example using R. CRC press.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis 6th Edition.
- Aldás Manzano, J., & Uriel Jiménez, E. (2017). Análisis multivariante aplicado con R. Ediciones Paraninfo, SA.
- https://bookdown.org/brian_nguyen0305/Multivariate_Statistical_Analysis_with_R/
- https://bookdown.org/jsalinas/tecnicas_multivariadas/presentacion.html

Sesión 10: Giovany Babativa

Semana 16 – noviembre 27 (Presencial 6:30am): Examen/exposición

VI. EVALUACIÓN

La nota final del curso corresponde al promedio ponderado de las notas obtenidas en las siguientes evaluaciones:

Actividad	Ponderación	Fecha de entrega
Talleres (4)	30%	Definida en Bloque Neón
Examen 1	20%	03 de octubre
Examen 2	20%	14 de noviembre
Presentación final	30%	27 de noviembre

VII. REGLAS GENERALES DE COMPORTAMIENTO ACADÉMICO

a. Asistencia

De acuerdo con los Artículos 40 y 41 (especializaciones) y 43 y 44 (maestrías) del Reglamento General de Estudiantes de Posgrado de la Universidad de los Andes, la Dirección de Programas Académicos del CIDER, DPA, recuerda que la asistencia a clase es un factor determinante para la adecuada formación de los profesionales inscritos en los programas del CIDER. Por tanto, el porcentaje mínimo de asistencia en todas las materias es del 85% (especializaciones) y 80% (maestrías). Los estudiantes que no cumplan con dicho

requerimiento obtendrán automáticamente como nota final 1,5 en la materia respectiva. Conforme a esta directiva, el profesor tomará regularmente registro de asistencia.

Excusas:

De acuerdo con el Artículo 42 del Reglamento General de Estudiantes de Posgrado de la Universidad de los Andes, la DPA del CIDER ha establecido las siguientes reglas:

- En caso de falla justificada a clase o a la realización de una evaluación, el estudiante deberá presentar excusa válida al profesor y a la DPA (correo: cidercoordinacion@uniandes.edu.co).
- Se consideran excusas válidas: a) Incapacidad médica; b) incapacidades expedidas por la Decanatura de Estudiantes; c) muerte del cónyuge o de parientes hasta el segundo grado de consanguinidad o de afinidad; d) autorización para participar en eventos deportivos, expedida por la Decanatura de Estudiantes; e) autorización para asistir a actividades académicas y culturales, expedida por la respectiva dependencia académica; f) Citación a diligencias judiciales, debidamente respaldada por el documento respectivo.
- En caso de ser aceptada la justificación por parte del profesor, se le anulará la inasistencia a clase al estudiante, o en caso de tratarse de un examen, el profesor fijará nueva fecha, hora y forma de evaluación.

Material de soporte para recordar y reforzar conceptos previos

- Pérez-Tejada, H (2008) Estadística para las ciencias sociales, del comportamiento y de la salud. 3a.edición. Disponible aquí: <https://www.uv.mx/rmipe/files/2015/09/Estadistica-para-las-ciencias-sociales-del-comportamiento-y-de-la-salud.pdf>
- Leer capítulos: 2. Descripción de datos, 5. Muestreo, 6 cálculo probabilístico
- El sitio web Academia Khan ofrece lecciones para aprender estadística y otros temas de matemáticas. Es un sitio académico reconocido internacionalmente por el éxito de su método. Disponible en <https://es.khanacademy.org/>
- Hancock, G. R., Mueller, R. O., & Stapleton, L. M. (2010). The reviewer's guide to quantitative methods in the social sciences. Routledge.
- Everitt, Skrondal. (2010). The Cambridge Dictionary of Statistics 4th Edition.
- Good, P. I., & Hardin, J. W. (2012). Common errors in statistics (and how to avoid them). John Wiley & Sons.
- Temas de estadística básica: <https://bookdown.org/aquintela/EBE/>