

PROGRAMA DE CURSO

ANÁLISIS DE DATOS E INFERENCIA CAUSAL

A. Antecedentes generales del curso:

Departamento	Ingeniería Industrial (DII)					
Nombre del curso	Análisis de datos e inferencia causal	Código	IN4143	Créditos	6	
Nombre del curso en inglés	<i>Data Analysis and Causal Inference</i>					
Horas semanales	Docencia	3,0	Auxiliares	1,5	Trabajo personal	5,5
Carácter del curso	Obligatorio	X		Electivo		
Requisitos	IN3242: Estadística					

B. Propósito del curso:

El curso tiene como propósito que los y las estudiantes apliquen distintas técnicas avanzadas, estadísticas, econométricas y de *machine learning*, a fin de generar una mejor calidad de información y mejorar la toma de decisiones. Para esto, utiliza modelos de análisis de datos para interpretar inferencias causales en casos aplicados a gestión, economía y políticas públicas, entre otros.

Finalmente, los y las estudiantes utilicen y apliquen métodos y software de análisis de datos, pudiendo interpretar resultados, los que explica de manera clara y argumentada.

El curso tributa a las siguientes competencias específicas (CE) y genéricas (CG):

CE2: Concebir y diseñar soluciones que crean valor para resolver problemas de las organizaciones, utilizando los conocimientos provenientes de la gestión de operaciones, tecnologías de información y comunicaciones, finanzas, economía y marketing.

CE3: Modelar, simular y evaluar problemas de gestión, para encontrar soluciones óptimas, a necesidades de la ingeniería industrial.

CE4: Emplear y aplicar los conocimientos de las distintas disciplinas constitutivas de la ingeniería industrial: gestión de operaciones, tecnologías de información y comunicaciones, finanzas, economía y marketing, en las respectivas áreas funcionales de las organizaciones.

CG1: Comunicación académica y profesional

Comunicar en español de forma estratégica, clara y eficaz, tanto en modalidad oral como escrita, puntos de vista, propuestas de proyectos y resultados de investigación fundamentados, en situaciones de comunicación compleja, en ambientes sociales, académicos y profesionales.

CG2: Comunicación en inglés

Leer y escuchar de manera comprensiva en inglés una variedad de textos e informaciones sobre temas concretos o abstractos, comunicando experiencias y opiniones, adecuándose a diferentes contextos y a las características de la audiencia.

CG3: Compromiso ético

Actuar de manera responsable y honesta, dando cuenta en forma crítica de sus propias acciones y sus consecuencias, en el marco del respeto hacia la dignidad de las personas y el cuidado del medio social, cultural y natural.

C. Resultados de aprendizaje:

Competencias específicas	Resultados de aprendizaje
CE3	RA1: Utiliza y propone modelos de análisis de datos para aplicaciones en gestión, economía y políticas públicas, entre otros, considerando las limitaciones de cada tipo de modelo, sesgo y el uso de medidas correctivas, entre otros aspectos.
CE2	RA2: Interpreta inferencias causales sobre problemáticas asociadas a temas de gestión, economía, políticas públicas, entre otros, a fin de generar información que favorezca la toma de decisiones.
CE2, CE4	RA3: Resuelve problemas de gestión, economía y políticas públicas mediante el uso de modelos y herramientas de inferencia causal, considerando el origen y validez de los datos, así como las ventajas y limitaciones de cada modelo, herramienta o instrumento.
Competencias genéricas	Resultados de aprendizaje
CG1	RA4: Produce, con claridad y precisión, textos breves donde reporta el análisis de datos e interpretación de resultados sobre ejemplos de aplicaciones de inferencia causal en problemas de gestión, economía, políticas públicas, entre otros.
CG1, CG2	RA5: Lee en inglés y español múltiples fuentes como artículos, videos, datos de páginas web, entre otros, extrapolando información sobre modelos y conceptos estadísticos con la cual respaldar argumentos sobre aplicaciones del análisis de datos e inferencia causal y su importancia.
CG3	RA6: Analiza ejemplos de marketing, gestión, políticas públicas, entre otros, en los que se emplea el análisis de datos para interpretar información y que podrían conllevar dilemas éticos respecto del uso de los datos, considerando el tipo de modelo usado y las consecuencias derivadas de esa elección.

D. Unidades temáticas:

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
1	RA1, RA4, RA5, RA6	Introducción a los métodos de investigación empírica y sus aplicaciones	2 semanas
Contenidos		Indicador de logro	
1.1. Introducción: Rol de evidencia empírica en conocimiento científico y toma de decisiones. 1.2. Tipos de datos y métodos de recolección. 1.3. Visualización y ética en el análisis de datos.		El/la estudiante: <ol style="list-style-type: none"> Determina la importancia de la evidencia empírica para el manejo e interpretación de datos. Identifica tipos de datos y métodos de recolección, distinguiéndolos según la naturaleza de estos. Utiliza tipos de datos y métodos de recolección para la resolución de problemas, seleccionando métodos correspondientes, según la naturaleza del dato a trabajar. Identifica y analiza casos que evidencian problemas éticos asociados al análisis de datos, enfatizando la importancia de trabajar de manera ética y responsable con los datos. Lee en inglés y español diversos textos donde se ejemplifican varios conceptos cubiertos en la unidad, extrayendo las ideas centrales sobre temas de análisis de datos e inferencia causal. Produce textos breves donde reporta los resultados del análisis de datos sobre el fenómeno trabajado, exponiendo, con claridad y precisión, conclusiones consistentes con el análisis realizado. 	
Bibliografía de la unidad		-Wooldridge (2007), cap. 1 y 19. -Material específico de la unidad.	

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
2	RA1, RA3, RA4	Modelos lineales y aplicaciones	3 semanas
Contenidos		Indicador de logro	
2.1. Estimación e inferencia en el análisis regresión múltiple. 2.2. Interpretación de los coeficientes de variables continuas y categóricas. 2.3. Problemas en los datos y aplicación de medidas correctivas.		El/la estudiante: <ol style="list-style-type: none"> Explica las propiedades básicas del modelo lineal clásico, considerando sus implicancias en ciertos contextos teóricos. Selecciona y determina el tipo de modelo para trabajar con datos, según el problema a resolver. Realiza estimaciones de modelos lineales, test de 	

<p>2.4. Selección de modelos: variables omitidas, irrelevantes, formas funcionales.</p> <p>2.5. Aplicaciones a casos en economía y gestión, entre otros ejemplos.</p>	<p>hipótesis, interpretando los coeficientes en ejemplos.</p> <p>4. Determina el alcance de los problemas que se generan con el manejo de datos y el relajamiento de supuestos clásicos, aplicando medidas correctivas.</p> <p>5. Aplica el modelo de regresión lineal en diferentes aplicaciones de economía y gestión, considerando la naturaleza de los datos.</p> <p>6. Produce textos breves donde reporta los resultados del análisis de datos sobre el fenómeno trabajado, exponiendo, con claridad y precisión, conclusiones consistentes con el análisis realizado.</p>
<p>Bibliografía de la unidad</p>	<p>-Wooldridge (2007), cap. 2-9. -Material específico de la unidad.</p>

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
3	RA1, RA2, RA3, RA4	Causalidad e inferencia	3,5 semanas
Contenidos		Indicador de logro	
<p>3.1. El problema de sesgo de selección.</p> <p>3.2. Conceptos de experimentos naturales, cuasi- experimentos y experimentos con asignación aleatoria.</p> <p>3.3. Causalidad, variables de control y heterogeneidad.</p> <p>3.4. Tipos de efectos de tratamiento y amenazas a efectos causales.</p>		<p>El/la estudiante:</p> <p>1. Explica el problema de selección y su solución a través de asignación aleatoria.</p> <p>2. Reconoce la importancia de realizar experimentos aleatorios, y los supuestos utilizados para interpretar la evidencia en términos causales.</p> <p>3. Aplica modelos lineales para realizar inferencias causales en aplicaciones para economía y la gestión, entre otros ejemplos.</p> <p>4. Elabora textos breves donde reporta los resultados del análisis de datos sobre el fenómeno trabajado, exponiendo, con claridad y precisión, conclusiones consistentes con el análisis realizado.</p>	
<p>Bibliografía de la unidad</p>		<p>-Angrist, J. & Pischke, J-S. (2009), cap. 1 & 2. -Bernal, R. & Peña, X. (2012), cap. 2-4. -Material específico de la unidad.</p>	

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
4	RA1, RA2, RA3, RA4	Endogeneidad e identificación	4 semanas
Contenidos		Indicador de logro	
4.1. El problema de identificación, concepto de endogeneidad y efecto causal. 4.2. Estimador de variables instrumentales, tests de exogeneidad y debilidad de instrumentos. 4.3. Datos de panel y estimadores de diferencias en diferencias. 4.4. Regresión discontinua. 4.5. Aplicaciones.		El/la estudiante: 1. Aplica los conceptos de identificación y endogeneidad, reconociendo su presencia e importancia en un conjunto de aplicaciones en economía y gestión. 2. Utiliza diversas técnicas para la estimación de efectos causales, considerando datos con diversas estructuras. 3. Determina las limitaciones de los métodos existentes, tomando en cuenta la evaluación de resultados obtenidos en la aplicación de modelos y técnicas de análisis de datos. 4. Produce textos breves donde reporta los resultados del análisis de datos sobre el fenómeno trabajado, exponiendo, con claridad y precisión, conclusiones consistentes con el análisis realizado.	
Bibliografía de la unidad		-Wooldridge (2007), cap. 13, 14, 15 & 16. -Angrist, J. & Pischke, J-S. (2009), cap. 4, 5 & 6. -Bernal, R. & Peña, X. (2012), cap.5,7, 8. -Material específico de la unidad.	

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
5	RA3	Big Data e inferencia causal	2,5 semanas
Contenidos		Indicador de logro	
5.1. Aprendizaje automático (machine learning) y big data en el contexto de la inferencia causal. 5.2. Árboles de clasificación y causal random forest: Inferencia causal en efectos de tratamiento heterogéneos. 5.3. Aplicaciones de big data e inferencia causal.		El/la estudiante: 1. Determina la relevancia de los métodos causales en el ámbito del "big data", y sus beneficios. 2. Aplica modelos de bosques aleatorios causales a diversos problemas en economía, políticas públicas y gestión, relacionándolos con efectos heterogéneos.	
Bibliografía de la unidad		Lecturas complementarias y material provisto por el cuerpo docente.	

E. Estrategias de enseñanza - aprendizaje:

El curso considera las siguientes estrategias de enseñanza-aprendizaje:

- Clases expositivas y participativas
- Resolución de problemas.
- Análisis de caso y aplicaciones.

F. Estrategias de evaluación:

El curso tiene distintas instancias de evaluación que consideran:

- Controles.
- Tareas.
- Examen.

Al inicio de cada semestre, el cuerpo académico informará sobre la cantidad y tipo de evaluaciones, así como las ponderaciones correspondientes.

G. Recursos bibliográficos:

Bibliografía obligatoria:

- (1) Wooldridge, J.M. (2007) Introducción a la econometría: un enfoque moderno. Thomson-Paraninfo.
- (2) Angrist, J. y Pischke, J-S. (2009) "Mostly Harmless Econometrics: An empiricist companion". Princeton University Press.
- (3) Bernal, R. y Peña, X. (2012) "Guía práctica para la evaluación de impacto". Ediciones Universidad Católica de Chile y Universidad de Los Andes, Colombia.

-Material de clases provisto por el profesor.

Bibliografía complementaria

- (4) Angrist, J. D., & Pischke, J. S. (2014). Mastering' Metrics: The Path from Cause to Effect. Princeton University Press.
- (5) Gerber, A. S., & Green, D. P. (2012). Field experiments: Design, analysis, and interpretation. WW Norton.
- (6) Greene, W.H (1998) Análisis Econométrico, tercera edición, Prentice-Hall.
- (7) Gujarati, D. N. (2011). Econometrics by example. Hampshire, UK: Palgrave Macmillan.
- (8) Hoyle, R. H., Harris, M. J., & Judd, C. M. (2002). *Research methods in social relations*. Thomson Learning.
- (9) Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523),1228-1242.

H. Datos generales sobre elaboración y vigencia del programa de curso:

Vigencia desde:	Otoño, 2022
Elaborado por:	Daniel Schwartz
Validado por:	Validación académico par: Raimundo Undurraga COMDOC, CTD de Industrial
Revisado por:	Área de Gestión Curricular