

## Diplomado en Big Data para Políticas Públicas

### Bases Estadísticas y Herramientas Para el Análisis de Datos Programa de Estudios

**Código:** BDPP002  
**Período Académico:** Junio/Julio 2019  
**Duración:** Intensivo  
**Requisito de asistencia:** 75%  
**Profesor:** Eduardo Fajnzylber Reyes // Pablo Aguirre Hörmann  
**Email profesor:** eduardo.fajnzylber@uai.cl // [pablojavier.aguirre@edu.uai.cl](mailto:pablojavier.aguirre@edu.uai.cl)

#### I. INTRODUCCION

El análisis de datos para toma de decisiones utiliza la investigación aplicada, como sólido apoyo en la búsqueda de soluciones a los problemas (u oportunidades) administrativos, financieros, de ingeniería u otros. Por esta razón, las herramientas que entrega la ciencia estadística dentro de la metodología de investigación que se aplique para contribuir a resolver el problema gerencial de interés deben realizarse bajo estrictos análisis de su pertinencia y aplicabilidad al contexto del problema que se estudia. Son muchos los aspectos, desde el punto de vista estadístico, que deben tenerse en cuenta para obtener información de valor que esclarezca el problema o la oportunidad. En particular, la metodología estadística que se utilice tendrá definitivas implicaciones en los resultados que se obtengan. Por otro lado, y considerando que vivimos en una época donde la cantidad y disponibilidad de datos va en ascenso en prácticamente todos los ámbitos de la sociedad, el poder realizar análisis estadísticos requiere del uso de herramientas especializadas con el fin de poder procesar las grandes cantidades de información que eventualmente puedan encontrarse.

En este curso de estadística para el análisis de datos masivos, se revisarán los conceptos básicos de las metodologías estadísticas, y se pondrán en práctica a través del uso del lenguaje de programación R, preparando al alumno para una adecuada aplicación e interpretación que conduzca a resultados que contribuyan de forma eficiente.

#### II. OBJETIVOS

Entender de manera teórica y práctica los fundamentos y conceptos de estadística que permita realizar análisis de un conjunto de datos y entrar al estudio sistemático o tecnológico estadísticamente de bases de datos, siendo capaces de diferenciar conceptos de estadística, tales como asociatividad, correlación, modelos estadísticos, análisis estadístico, entre otros, con la finalidad de distinguir la modelación de la información.

Se busca también profundizar en el entendimiento y uso del lenguaje de programación R con el fin de adquirir herramientas avanzadas para el manejo y limpieza de datos además de aprender a comunicar y compartir hallazgos obtenidos a partir del análisis exploratorio de información.

### III. CONTENIDOS

#### **Clases 1, 2 y 3 (Herramientas estadísticas para el análisis de datos)**

1. Conceptos básicos en R y RStudio
2. Tipos de datos y funciones
3. Importación y exportación de datos
4. Manipulación de datos
5. Visualización de datos
6. Comunicar/Compartir trabajo: R Markdown

#### **Clases 4, 5 y 6 (Bases estadísticas para el análisis de datos)**

7. Experimentos, resultados y variables aleatorias
8. La distribución de probabilidad de una variable aleatoria
9. Estadísticos: Esperanza, Mediana, Varianza, Covarianza, Correlación
10. Distribuciones importantes: Normal, chi-cuadrado, T-student, Fisher
11. Ley de Grandes Números y Teorema Central del Límite
12. Parámetros, muestras y estimadores
13. Inferencia: Test de hipótesis, valores P, intervalos de confianza
14. El modelo de regresión lineal múltiple (MLRM): definición, estimación y estadísticos
15. Aplicaciones del MRLM a problemas de predicción

### IV. METODOLOGIA

Clases expositivas y prácticas. A lo largo de las sesiones de clases se expondrán los contenidos del curso y mostrarán casos de ejemplos de bibliografía o casos reales que deberán ser discutidos. También hay evaluaciones de contenido a través de tareas que podrían ser individuales o grupales. Durante el curso se realizarán una serie de ejercicios en clase y demostraciones prácticas en R que requerirán que cada estudiante interactúe con un computador personal.

Las tareas grupales o individuales que resulten deben ser entregadas según el plazo establecido previamente, la entrega puede ser físicamente o en línea en el sistema de webcursos. No hay entregas fuera de plazo. Cada uno de los documentos subidos a webcursos debe tener como título el nombre del individuo o grupo según sea el caso.

### V. EVALUACION

La nota final del curso será calculada como el promedio de todas las tareas individuales o grupales.

Fecha de entrega tarea 1: 1 de Julio

Fecha de entrega tarea 2: 29 de Julio

## VI. BIBLIOGRAFIA

Wickham y Golemund. 2016. R for Data Science. Disponible en: <https://r4ds.had.co.nz/>

Wickham, Hadley. Advanced R. 2014. Web del libro: <http://adv-r.had.co.nz/>

Fajnzylber, Eduardo. 2017. Apuntes de econometría.

J. M. Wooldridge, "Introducción a la Econometría: Un enfoque moderno", 4ª edición, Thomson, 2010. Apéndices A, B y C.

## VII. PLANIFICACION

Las clases se llevarán a cabo los días viernes y sábado, de acuerdo a la planificación que se presenta a continuación. Los horarios serán los siguientes:

Viernes:	15.30 - 16.40	Clases
	16.40 - 16.50	Pausa
	16.50 - 18.00	Clases
	18.00 - 18.20	Coffee break
	18.20 - 19.30	Clases
Sábados:	09.30 - 10.40	Clases
	10.40 - 11.00	Coffee break
	11.00 - 12.10	Clases
	12.10 - 12.20	Pausa
	12.20 - 13.30	Clases

<b>Parte 1 - Herramientas estadísticas para el análisis de datos</b>	
Viernes, 14 de junio	Clase: Temas 1 y 2
Sábado, 15 de junio	Clase: Temas 3 y 4
Viernes, 21 de junio	Clase: Temas 5 y 6
Por Definir	Ayudantía R (apoyo a tarea 1)
Lunes, 1 de Julio	Entrega Tarea 1
<b>Parte 2 - Bases estadísticas para el análisis de datos</b>	
Sábado, 22 de junio	Clase de nivelación estadística (obligatoria para quienes no hayan aprobado la prueba de evaluación)
Viernes, 28 de junio	Clase: Temas 7, 8, 9 y 10
Viernes, 5 de julio	Clase: Temas 11, 12 y 13
Sábado, 6 de julio	Clase: Temas 14, 15 y 16
Sábado 20 de julio	Ayudantía estadística (apoyo a tarea 3)
Lunes, 29 de julio	Entrega Tarea 2