
Programa del Curso de Exploración, Visualización y Mantenimiento de Datos, I-2020

1. Identificación

Nombre: Exploración, Visualización y Mantenimiento de Datos				
Código: MMD3201	SCT: 6	Modalidad: Presencial	Duración: Semestral	Docente: Raúl Valenzuela
Horas de Cátedra	Horas de Ayudantía		Horas de Laboratorio	Horas de Trabajo Personal
48	0		22,5	109,5

2. Descripción

Asignatura de tercer año de la carrera de Ingeniería Civil en Modelamiento Matemático de Datos de la Universidad de O'Higgins.

3. Resultados de Aprendizaje Esperados

Al final del curso se espera que el estudiante:

1. Conozca distintas herramientas de la estadística para describir datos.
2. Determine y describa las características principales y limitaciones de una base de datos real utilizando métodos estadísticos descriptivos.
3. Comprenda cómo abordar la **lectura de datos masivos** contenidos en distintos formatos.
4. Aplique técnicas de exploración visual de datos utilizando software y paquetes estadísticos.
5. Entienda y aplique los principios de visualización de datos para comunicar un análisis exploratorio y explicativo a una audiencia amplia y una audiencia técnica.
6. Esté familiarizado con las tareas de mantenimiento de datos e investigación reproducible.

4. Unidades Temáticas (UT) y Contenidos

A. Revisión de probabilidades y estadísticas

- A.1. Regresión lineal simple
- A.2. Residuos y medidas de bondad de ajuste
- A.3. Predicción de intervalos
- A.4. Regresión lineal múltiple

B. Medidas de tendencia central y dispersión

- B.1. Media, mediana, moda, trimedia
- B.2. Estimadores robustos y resistentes
- B.3. Rango intercuartil
- B.4. Desviación estándar
- B.5. Desviación media absoluta

Escuela de Ingeniería

- C. Datos, formatos y su manipulación**
 - C.1. Tipo y estructura de datos
 - C.2. Formatos de almacenamiento
 - C.3. Acondicionamiento de datos (Data Wrangling)
 - C.4. Imputación de datos faltantes

- D. Análisis exploratorio de datos (AED)**
 - D.1. Origen del AED
 - D.2. Técnicas de AED
 - D.3. Métodos de Visualización en AED

- E. Visualización de datos**
 - E.1. Los 6 principios de Edward Tufte
 - E.2. Visualización exploratoria vs explicativa
 - E.3. Escogiendo una visualización efectiva
 - E.4. Elementos de diseño gráfico aplicado a datos

- F. Mantenimiento de datos**
 - F.1. Ciclo de vida de los datos
 - F.2. Planificación para la administración de datos
 - F.3. Almacenamiento y respaldo

- G. Visualización espacial de datos**
 - G.1. Herramientas
 - G.2. Técnicas de visualización

5. Metodología Docente

La metodología de trabajo será activo-participativa, en donde se desarrollarán cátedras expositivas y sesiones demostrativas mediante videoconferencia. Además, los estudiantes trabajarán directamente con código Python para reforzar los conceptos. Al final del curso los estudiantes presentaran un proyecto donde demostraran los conocimientos adquiridos en el curso.

6. Evaluación

La evaluación permitirá que los estudiantes demuestren los resultados de aprendizaje alcanzados en los distintos momentos del proceso de enseñanza. La evaluación se realizará mediante 2 Control de Cátedra (CC), 3 Ejercicios de Python (EP), 1 Ensayo (EN) y un Examen (E) en las siguientes fechas:

Ítem	Fecha
Control Cátedra	18 mayo
Ejercicio Python	27 mayo
Ejercicio Python	10 junio
Ensayo sobre lectura	15 junio
Ejercicio Python	8 julio
Control Cátedra	22 julio
Examen	Por definir

1. La nota final (NF) está compuesta por la Nota de Cátedra (NC) y la Nota de Actividades Complementarias (NAC) con las siguientes ponderaciones:
$$NF = 0,7*NC + 0,3*NAC$$
2. La asignatura se aprueba con $NC \geq 4.0$ y $NAC \geq 4.0$.
3. La NC está compuesta por el promedio de los Controles de Cátedra (CC) y el Examen (E) con las siguientes ponderaciones:
$$NC = 0,5*CC + 0,5*E$$
4. La NAC está compuesta por las notas de los Ejercicios con las siguientes ponderaciones:
$$NAC = (EP1+EP2+EP3+EN)/4$$
5. Durante las evaluaciones no se permitirá lo siguiente: intercambio de materiales, mantener sobre la mesa elementos distintos de lápices, goma, corrector. No se permite el uso de calculadoras, el uso de celulares o elementos tecnológicos con capacidad de almacenar texto, video, audio o conexión a internet. Durante las evaluaciones se podría exigir la presentación de un documento de identidad en buen estado.
6. Los EP son de carácter práctico y se desarrollaran durante todo el bloque correspondiente. El contenido a evaluar es todo lo visto hasta la semana anterior al ejercicio.

7. Exención al Examen

1. El promedio simple entre los Controles de Cátedra corresponde a la nota de presentación (NP) del estudiante.
2. Cualquier estudiante cuya Nota de Presentación cumpla $NP \geq 5.5$, está exento de rendir el Examen.

8. Planificación

Tabla 1: Programa de Actividades por Semana

Semana	Lunes correspondiente	Actividad (Unidad Temática)
1	13 abril	A
2	20 abril	A
3	27 abril	B
4	4 mayo	C
5	11 mayo	C
6	18 mayo	D / CC
7	25 mayo	D / EP
8	1 junio	D
9	8 junio	E / EP
10	15 junio	E / EN
11	22 junio	F
12	29 junio	G
13	6 julio	Trabajo proyecto / EP
14	13 julio	Trabajo proyecto
15	20 julio	Trabajo proyecto / CC

9. Bibliografía y Material de Apoyo

1. Briney, K. (2015). *Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success*. Pelagic Publishing.
2. Devore, J. L. (2011). *Probability and Statistics for Engineering and the Sciences*. Cengage Learning.
3. Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.
4. Manski, Charles F. *Identification for prediction and decision*. Harvard University Press, 2009.
5. McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media.
6. Nussbaumer, C. (2015). *Storytelling with Data*. John Wiley & Sons.
7. Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press.

10. Otros

1. Para cualquier consulta relacionada con la asignatura se recomienda el uso de la plataforma U-Campus. El horario y lugar para consultas se informará oportunamente.
2. Cualquier asunto relacionado con el curso será comunicado mediante la plataforma U-Campus o en clases.

