

CARRERA: Licenciatura en Ciencias de Datos	CURSO LECTIVO: 2024
CÁTEDRA: Estadística Avanzada y Modelos de Regresión	CURSO: 2º año - 2º semestre
DURACIÓN: Semestral	Hs. TOTALES: 48 Hs. Reloj Totales
SEMANAS: 16	Hs. TEÓRICAS: 32 Hs. Reloj Totales Hs. PRÁCTICAS: 16 Hs. Reloj Totales

PROFESOR ADJUNTO: Benitez, Carlos Marcelo

PROFESOR ADJUNTO: Fernández Lozada, Marco Antonio

1. OBJETIVOS DE LA ASIGNATURA

Que los alumnos logren:

1. **Comprender los principios fundamentales de la estadística y su aplicación en la inferencia de datos**, con un enfoque particular en la distinción entre correlación y causalidad, y en cómo estos conceptos influyen en la interpretación y toma de decisiones basada en datos.
2. **Aplicar pruebas estadísticas paramétricas y no paramétricas para el análisis de datos complejos**, facilitando la toma de decisiones informadas mediante el uso adecuado de pruebas de hipótesis, análisis de varianza, y otras herramientas estadísticas para abordar diferentes tipos de datos y situaciones.
3. **Desarrollar habilidades para establecer inferencias causales**, utilizando una combinación de herramientas estadísticas tradicionales y técnicas avanzadas de Machine Learning, con el fin de identificar relaciones causales en conjuntos de datos.
4. **Dominar el uso de modelos de regresión**, incluyendo regresión lineal, no lineal y árboles de regresión, y aplicar estos modelos a problemas reales de predicción, y análisis de datos, asegurando una adecuada interpretación y ajuste a las características específicas del problema.
5. **Adquirir competencias en técnicas de transformación de datos y reducción de dimensionalidad**, mejorando el rendimiento de los modelos de Machine Learning mediante la aplicación de transformaciones adecuadas y la reducción de la complejidad computacional.

2. UNIDADES TEMÁTICAS

1. Fundamentos de la Estadística y

Pruebas de Hipótesis Estadística

Descriptiva:

Resumen y organización de datos: Tablas de frecuencias, distribuciones de frecuencias, medidas de tendencia central (media, mediana, moda).

Medidas de dispersión: Rango, varianza, desviación estándar, coeficiente de variación. Visualización de datos: Histogramas, diagramas de caja (boxplots), diagramas de dispersión, gráficos de barras.

Pruebas paramétricas:

Diferencia de medias entre dos muestras Normales independientes con varianza conocida: Prueba de Z para dos muestras independientes, condiciones y aplicaciones.

Diferencia de medias entre dos muestras Normales independientes con varianza no conocida: Prueba de T de Student para dos muestras independientes, estimación de la varianza y su importancia.

Diferencia de medias entre dos muestras cualesquiera: Prueba de Z para muestras independientes.

Prueba de media unilateral y bilateral: Pruebas de hipótesis para medias, interpretación y condiciones para su uso.

Prueba de la mediana: Prueba para muestras independientes, comparación de medianas. Prueba de igualdad de varianzas - Test de Levene: Prueba de homocedasticidad, comparación entre muestras para comprobar igualdad de dispersión.

Análisis de varianza: ANOVA, interpretación de los resultados.

Pruebas no paramétricas:

Chi-cuadrado: Test de independencia y bondad de ajuste.

Kolmogorov-Smirnov: Prueba de ajuste de distribuciones y comparación entre dos muestras. Shapiro-Wilk y Anderson-Darling: Pruebas de normalidad para una sola muestra.

2. Regresión y Modelos predictivos
Correlación y Causalidad

Definición, tipos de correlación (Pearson, Spearman) Diferenciación entre correlación y causalidad

Modelo de regresión lineal simple

Estimación de la recta de regresión, interpretación de coeficientes, análisis de residuos.

Modelo de Regresión lineal múltiple

Variables independientes múltiples, supuestos del modelo, multicolinealidad, evaluación de la bondad de ajuste (R^2 , R^2 ajustado).

Evaluación del ajuste y diagnóstico de modelos: Herramientas para analizar la calidad del modelo, como análisis de residuos, pruebas de heterocedasticidad.

Multicolinealidad: Identificación, diagnóstico y soluciones.

Regresión No Lineal

Modelos de regresión no lineales: Ejemplos como el modelo polinómico.

Regresión logística

Introducción a la regresión logística binaria y multinomial, interpretación de los coeficientes, función sigmoide, evaluación del modelo.

Árboles de Regresión

Fundamentos de árboles de regresión: Creación de árboles, crecimiento del árbol, criterio de división.

Regresión con Random Forest y XGBoost: Comparación entre Random Forest y XGBoost, importancia de las variables, afinamiento de parámetros (grid search, random search, validación cruzada, regularización, poda).

Overfitting y Underfitting: Identificación de modelos que se ajustan demasiado o no lo suficiente, técnicas para evitar estos problemas.

Entrenamiento y Métricas de

Evaluación Métricas comunes: RMSE, MAE, R^2 , R^2 ajustado. Interpretación de Resultados

3. Inferencia Causal y Machine Learning Causal

Métodos para establecer causalidad en datos observacionales.

Contrafactuales y causalidad: Introducción al concepto de causalidad contrafactual, uso de modelos causales.

Estimación del efecto causal. Técnicas.

Introducción a los bosques causales: Uso de árboles de decisión para estimar efectos causales. Evaluación de intervenciones: Técnicas para la estimación de efectos de tratamiento en datos.

3. BIBLIOGRAFÍA

3.1 BIBLIOGRAFÍA GENERAL OBLIGATORIA

- León-García, A. (2019). Probability, statistics, and random processes for electrical engineering (4th ed.). Pearson.
- Ness, O. (2023). Causal AI. MIT Press.
- Molak, A. (2023). Causal inference and discovery in Python: Unlock the secrets of modern causal machine learning with DoWhy, EconML, PyTorch and more. Packt Publishing.
- Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
- Chan, Y. L., Badano, A., & Rey, M. (2018). Análisis inteligente de datos con lenguaje R. Editorial Universitaria.
- Szeretter Noste, M. E. (2017). Apunte de regresión lineal.
- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: A guide for data scientists. O'Reilly Media

4. METODOLOGÍA

El curso está organizado en 3 unidades temáticas divididas en encuentros de 3hs reloj semanales, a realizarse en formato presencial. La modalidad adoptada para el dictado será **teórico-práctica**. En las clases se presentarán los temas de cada unidad, proponiendo espacios de intercambio con el docente y entre los estudiantes a partir de consignas específicas. Se facilitará material de lectura obligatoria y complementaria para complementar la comprensión de las unidades.

5. EVALUACIONES Y CRITERIOS PARA LA APROBACIÓN

La aprobación de la materia estará supeditada al cumplimiento de la condición de asistencia exigida por la Universidad, la aprobación de todas las actividades prácticas y la aprobación del examen parcial y final.

Los trabajos prácticos serán grupales, debiéndose cargar a través de la plataforma de Entornos Virtuales de Aprendizaje en tiempo y forma, otorgándole una única instancia de revisión y recuperación. Las actividades prácticas deberán contar con su aprobación para acceder a la instancia de evaluación final.

Para los trabajos prácticos y las evaluaciones se realizarán sesiones de consultas individuales y grupales, haciendo además puesta en común general si el caso lo requiera. A los estudiantes que presenten dificultades se les observará y se los guiará para resolver el conflicto.

La instancia de recuperación está prevista para estudiantes que no hayan aprobado el examen parcial o que hayan estado ausentes.

Criterios de Evaluación:

- Respeto de las consignas presentadas.
- Resolución correcta de los problemas planteados.
- Adecuada respuesta a los contenidos teóricos.

6. CRITERIOS y MODALIDAD PARA LA EVALUACIÓN DEL EXAMEN FINAL

El examen final consiste en una evaluación escrita, presencial e individual, donde el alumno deberá demostrar conocimientos teóricos y prácticos. El examen final se diferencia en que abarca todos los temas del programa. Los ejercicios prácticos tendrán un carácter integrador, articulando los distintos contenidos vistos en la materia. Finalmente, en las preguntas teóricas se pretende que el alumno demuestre un conocimiento profundo de los temas, relacionando conceptos entre sí.

Criterios de Evaluación:

- Respeto de las consignas presentadas.
- Adecuada respuesta a los contenidos teóricos.
- Relación de conceptos pertinente.
- Resolución correcta de los problemas planteados.
- Fundamentación bibliográfica de los temas.