

PDA: PLAN DE DESARROLLO DE LA ASIGNATURA 2023-I

Denominación del espacio formativo: **FUNDAMENTOS DE ESTADÍSTICA PARA ANALÍTICA DE DATOS**

Código de la Asignatura:	64491989		Número de Créditos:	CINCO (5)	
Naturaleza Académica:	Teórica		Teórico-Práctica	X	Práctica

Prerrequisitos Exigidos: **Ninguno**

Co-requisitos Exigidos: **Ninguno**

Dispositivo(s) Pedagógico(s) Requerido(s):

- La clase expositiva o dialógica
- Estudio de caso
- Laboratorio en centro de cómputo

FINALIDAD DEL ESPACIO FORMATIVO

- Adquirir claridad en las definiciones y conceptos básicos de estadística en la solución de problemas que involucran bases de datos.
- Aplicar conceptos y métodos matemáticos, estadísticos y computacionales, para extraer conocimiento de grandes volúmenes de datos.
- Interpretar resultados como insumo para la toma de decisiones.
- Apropiación por parte de los estudiantes de las diferentes técnicas estadísticas abordadas en el curso y correcta aplicación en una determinada situación del mundo real.
- Manejo de Software estadístico Python o R para el procesamiento y análisis de información. Visualización de datos.

DESCRIPCIÓN SEMANAL DEL DESARROLLO DE LA ASIGNATURA				
SEMA NA	SESI ÓN	TEMAS	FINALIDADES FORMATIVAS DEL TRAYECTO	REFERENCIAS
SEMA NA 1	1-2	Presentación del programa. Reglas de juego. Conceptos básicos de estadística Introducción a Python: carga de datos.	Apropiación de las definiciones y conceptos básicos de la Estadística. Reconocer Python y Colab.	Capítulos 1 al 3 Wes McKinney (2012)
SEMA NA 2	1 - 2	Tablas de frecuencias Análisis gráfico univariado: Histogramas, Barras, pie, líneas, etc., con Python. Medidas de tendencia Central	Conocer las formas de resumir la información en tablas y gráficos.	Capítulo 8 Wes McKinney (2012) Capítulos 1 al 2 Mendenhall (2009)
SEMA NA 3	1 - 2	Medidas de dispersión. Medidas de posición Medidas de forma	Identificar la medida posición central que resuma mejor una serie de datos. Cálculo e interpretación.	Capítulo 5 al 7 Wes McKinney (2012) Capítulos 1 al 2 Mendenhall (2009)
SEMA NA 4	1 - 2	Análisis gráfico de datos univariados: Boxplot. Trabajo en Python	Identificar la medida posición central que resuma mejor una serie de datos. Cálculo e interpretación.	Capítulo 5 al 7 Wes McKinney (2012) Capítulos 1 al 2 Mendenhall (2009)
SEMA NA 5	1 - 2	Teorema de Chebyschev Datos Bivariados.	Comparar la prueba empírica con el teorema de	Capítulo 5 al 7 Wes McKinney (2012)

		Correlación.	Chebyschev Reconocer el uso de tablas de contingencia para datos bivariados, gráficos de barras y el concepto de correlación.	Capítulo 3 Mendenhall (2009)
SEMA NA 6	Sesi ón 1	Probabilidad. Conceptos y propiedades.	Conocer y aprender los conceptos y definiciones básicas de probabilidad y utilizarla para presentar y resolver problemas no deterministas.	Capítulo 2 Montgomery (2003) Capítulo 4 Mendenhall (2009) Taller propuesto
	Sesi ón 2	Avance del Caso de Estudio 1 (30%)	Aplicar todos los elementos de estadística descriptiva a una base de datos particular	Presentación en grupo de los resultados obtenidos.
SEMA NA 7	1 - 2	Probabilidad condicional. Teorema de Bayes y probabilidad total.	Estudiar las diferentes distribuciones de probabilidad, su uso y ejecución en R	Capítulo 2 Montgomery (2003) Capítulo 4 Mendenhall (2009)
SEMA NA 8	1 - 2	Distribuciones: Binomial, Normal, Chi-Cuadrado, t-Student.	Describir la curva normal y utilizar sus propiedades para aplicarlas en conjuntos de datos que asuman normales.	Capítulo 7-8 Montgomery (2003) Capítulo 7 Mendenhall (2009)
SEMA NA 9	1 - 2	Introducción a la inferencia estadística. Conceptos de muestreo	Reconocer los elementos fundamentales de inferencia estadística y de muestreo.	Capítulo 7-8 Montgomery (2003) Capítulo 7 -8 Mendenhall (2009)
SEMA NA 10	1 - 2	Introducción a la inferencia estadística. Definiciones.	Apropiación de los conceptos de inferencia clásica.	Capítulo 7-8 Montgomery (2003)

		Estimadores puntuales y de intervalo.	Uso adecuado de intervalos de confianza y pruebas de hipótesis.	Capítulo 7 -9 Mendenhall (2009)
SEMA NA 11	1 - 2	Pruebas de hipótesis clásicas. Valor crítico. El p-valor	Apropiación de los conceptos de inferencia clásica. Uso adecuado de intervalos de confianza y pruebas de hipótesis.	Capítulo 9 Montgomery (2003) Capítulo 9 Mendenhall (2009)
SEMA NA 12	1 - 2	Regresión lineal simple. Criterios de ajuste	Reconocer los elementos, conceptos y usos de la regresión lineal	Capítulo 4 Géron, A. (2017). Capítulo 11 Montgomery (2003) Capítulo 12 Mendenhall (2009)
SEMA NA 13	1 - 2	Regresión lineal simple. Análisis de Residuos	Reconocer los elementos, conceptos y usos de la regresión lineal	Capítulo 11 Montgomery (2003) Capítulo 12 Mendenhall (2009)
SEMA NA 14	1 - 2	Regresión lineal simple. Transformación Box-Cox y otras	Reconocer los elementos, conceptos y usos de la regresión lineal múltiple	Capítulo 11 Montgomery (2003) Capítulo 12 Mendenhall (2009)
SEMA NA 15	1 - 2	Regresión lineal múltiple.	Reconocer los elementos, conceptos y usos de la regresión múltiple	Capítulo 4 Géron, A. (2017). Capítulo 11 Montgomery (2003) Capítulo 12 Mendenhall (2009)

SEMA NA 16	1 - 2	Regresión logística simple. Interpretación de los parámetros.	Reconocer los elementos, conceptos y usos de la regresión logística	Capítulo 4 Géron, A. (2017). Capítulo 11 Montgomery (2003) Capítulo 12 Mendenhall (2009)
SEMA NA 17	Sesi ón	Entrega y socialización del Caso de Estudio	<i>Presentación en grupo de los resultados obtenidos.</i>	

EVALUACIÓN DEL CURSO

Para el curso de Fundamentos De Estadística Para Analítica De Datos, se conformarán grupos de tres estudiantes y la evaluación estará dada de la siguiente forma:

- Entrega y sustentación estudio de caso (análisis descriptiva y visualización): 30 puntos. Distribuidos así:

Avance del documento en PDF y notebooks (PDF, Ipynb) 15 puntos.

Sustentación en grupo (PPT) 10 puntos.

Revisión de avances 5 puntos

- Entrega y sustentación estudio de caso (regresión): 45 puntos. Distribuidos así:

La entrega del documento en PDF y notebooks 15 puntos,

Tablero o reporte general: 10 puntos

Sustentación: 15 puntos

Revisión de avances 5 puntos

- Taller de probabilidad e inferencia (10 puntos)
- Ejercicios cortos de clase o casa (15 puntos)

La nota se calcula sobre 100. La nota de aprobación es 70.

Recuerde que:

- Las notas se calculan de 0.0 hasta 5.0, en múltiplos de 0.1.
- El material de las sesiones de clase quedarán en carpeta del Curso.

BIBLIOGRAFÍA

Bibliografía Obligatoria

- Wes McKinney (2012). Python for Data Analysis. Second Edition. O'Reilly.
- Milovanović, Foures and Vettigli (2015). Python Data Visualization Cookbook. Second Edition.
- Montgomery-Runger-Applied-Statistics-and-Probability-for-Engineers-3rd.2003
- Mendenhall (2009), W., Beaver, R. J., & Beaver, B. M. (2003). Probability and statistics 13rd. Thomson, Brooks/Cole.
- Géron, A. (2017). Hands-on machine learning with scikit-learn and tensorflow: Concepts. Tools, and Techniques to build intelligent systems.

Bibliografía Complementaria

- Roger D. Peng (2015). Exploratory Data Analysis with R.
- Narayananchar P. (2013). R statistical application. Development by example. Beginner's guide.
- Wes McKinney (2017). Python for Data Analysis. Second Edition. O'Reilly.
- Milovanović, Foures and Vettigli (2015). Python Data Visualization Cookbook. Second Edition.
- Larry Wasserman (2004). All of Statistics a Concise Course in Statistical Inference-Springer.

Docente	Director del Programa:	Fecha
Luis Andrés Campos		Agosto 2023