



Universidad de Puerto Rico
Recinto de Mayagüez
Facultad de Artes y Ciencias
DEPARTAMENTO DE MATEMATICAS

Curso: Statistical Data Analysis

Codificación: ESMA 3016

Número de horas/crédito: 3 créditos. 2 horas de conferencia y un laboratorio de dos horas por semana.

Prerrequisito: MATE 3031 y COMP 3010

Profesor: Edgar Acuña

Oficina: OP307

Extension:

Horas de Oficina: Martes y Jueves de 9.00am a 10.30 am y Martes de 12.30 a 3.30pm

Descripción del Curso: Estadística Descriptiva, Análisis exploratorio de datos univariados y multivariados, cálculo de probabilidades, variables aleatorias y distribuciones de probabilidad. Simulación. Inferencia estadística basada en una muestra, Comparación de dos grupos, Pruebas noparamétricas, Correlación y regresión Lineal Simple, Análisis de datos de categoría.

El propósito de este curso es que el estudiante aprenda a aplicar diversos métodos estadísticos haciendo uso en gran medida de un programa estadístico de computadoras. El concepto de probabilidad y variables aleatorias debe ser entendido como una medida del grado de incertidumbre que existe al extraer una conclusión de la población basada en una muestra tomada de ella.

Objetivos del Curso:

Al final del curso se espera que el estudiante pueda

- ◆ Resumir un conjunto de datos y presentarlos en tablas y gráficas
- ◆ Aplicar técnicas de Análisis Exploratorio de Datos para analizar un conjunto de datos.
- ◆ Calcular medidas estadísticas de centralidad y variabilidad basadas en la muestra tomada
- ◆ Establecer la relación entre dos variables cualitativas
- ◆ Entender el significado del concepto de correlación para relacionar dos variables cuantitativas
- ◆ Establecer una línea de regresión para representar la tendencia de la relación lineal de dos variables cualitativas.
- ◆ Determinar las probabilidades de eventos de experimentos aleatorios
- ◆ Aplicar herramientas de cálculo diferencial e integral a través del concepto de variables aleatorias para calcular probabilidades de eventos.
- ◆ Modelar experimentos aleatorios de acuerdo a los modelos de distribuciones conocidas como la Binomial, la Poisson y la Normal.

- ◆ Simular datos que siguen una distribución conocida, haciendo uso de un programa estadístico de computadoras
- ◆ Entender el significado del Teorema del Límite Central y la distribución de la media muestral.
- ◆ Aplicar los métodos de inferencia estadística tales como prueba de hipótesis e intervalos de confianza que le permitan para sacar conclusiones de la población usando la muestra extraída de ella

Bosquejo de contenido y distribución del tiempo:

Textos (recomendados):

1. Python for Data Analysis (2013) Wes McKinney. O'Reilly Media Inc , CA. Disponible gratis en <https://github.com/ICOUNTER/python-data-mining-and-text-mining/blob/master/Python4DataAnalysis.pdf>
- 2-Think Stats. Exploratory data Analysis in Python (2014). Allen B. Downey. Green Tea Press. Disponible en <http://greenteapress.com/thinkstats2/thinkstats2.pdf>
- 3- An introduction to statistics with Python (2016) Haslwanter, T. Springer Verlag. Material del libro esta disponible en <https://github.com/thomas-haslwanter/statsintro>
- 4-A Complete Tutorial to Learn Data Science with Python from Scratch. (2016) Analytics Vidhya. Disponible en <https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/>

Lección	Tema
1-2	Introducción to Python and Statistics
3	Tipos de Datos
4-5	Organización de datos
6	Manipulación de datos
7	Presentación de datos: Tablas,
8-9	Graficas Estadísticas Univariadas
10	Scatterplots y
11-12	Regresión y Correlación Lineal
13-14	Medidas Estadísticas
15	Box plots
16	EXAMEN I
17-18	Experimentos Aleatorios y Eventos
19	Definiciones y propiedades de Probabilidades

20-21	Probabilidad Condicional e Independencia
22-23	Variables aleatorias y sus propiedades
24	Distribución Binomial
25	Distribución Poisson
26-27	Distribuciones continuas y sus densidades
28-29	Simulación de distribuciones
30-31	La distribución Normal
32	Aplicaciones de la distribución Normal
33	Distribución de la Media Muestral
34	Distribución de la proporción Muestral
35	Examen II
36-37	Introducción a Inferencia Estadística
39	Estimación por Intervalos de Confianza
40-41	Prueba de hipótesis con una sola muestra
42-43	Prueba de hipótesis con dos muestras
44	Tablas de Contingencia
45	Bondad de Ajuste

Estrategias instruccionales: Conferencias en donde se presentan: los conceptos y métodos fundamentales de estadística, ejemplos, ejercicios y la solución de problemas. El uso de otras estrategias (tales como uso de tecnología avanzada, aprendizaje cooperativo, trabajo en clase, discusión abierta, laboratorios, etc.) se deja a discreción del profesor.

Recursos de aprendizaje o instalaciones mínimos disponibles o requeridos:

Todo el material del curso, incluyendo prontuario, textos recomendados, presentaciones y laboratorios esta en la cuenta de github del curso en <https://github.com/eacunafer/PyLabs-Analisis-de-Datos>.

Los trabajos del curso deben ser enviados al profesor a través de Piazza. Guardar evidencia de que el trabajo se sometió, porque el profesor no se hace responsable si el trabajo no es recibido.

Estrategias de evaluación:

La evaluación del curso puede incluir exámenes, asignaciones, pruebas cortas, y otros a discreción del profesor del curso.

Número de exámenes	2 exámenes parciales, y un examen final(70%)	- 3 -
--------------------	--	-------

Número de pruebas cortas	
Número de tareas	4 (30%)

Sistema de Calificación:

Se utiliza el siguiente sistema de calificación

90% - 100% A 80% - 89% B 65% - 79% C 60% - 64% D
0% - 59% F

EAF/Febrero 13, 2018