

Datos Generales

1. Nombre de la Asignatura: Métodos Estadísticos Básicos	2. Nivel de formación: Maestría	3. Clave de la Asignatura: IH590	
4. Prerrequisitos: Ninguno	5. Área de Formación: Básica común obligatoria	6. Departamento: Métodos Cuantitativos	
7. Modalidad: Híbrida	8. Tipo de Asignatura: Curso-Taller		
9. Carga Horaria: 60 horas			
Teoría: 40 horas	Práctica: 20 horas	Total: 60	Créditos: 7
10. Trayectoria de la asignatura:			

Contenido del Programa

11. Presentación

La ciencia de datos es una disciplina reciente conformada por un set de herramientas y técnicas utilizadas para extraer información útil de datos, y que ya es fundamental en el ámbito económico, financiero, gubernamental, entre otros. Los datos pueden ser de distintos tipos y provenir de diferentes fuentes tanto personales como dentro de la web, y conforme avanza el mundo tecnológico se ha cambiado el enfoque hacia conceptos como Ciencia de Datos, Big Data y Machine Learning. Las empresas empiezan a capitalizar con estos conceptos aplicados en gran cantidad de datos, capaces de predecir, por ejemplo, cómo es que las personas reaccionan y poder tomar decisiones en base a esto, y por tanto, tener un impacto en gran cantidad de sectores como salud, educación, energético, seguridad, deportes, etc, dando lugar a la mejora e innovación.

La ciencia de datos es un tema interdisciplinario, que contiene porciones considerables del área de matemáticas, estadística, programación, algoritmos de machine learning y análisis de grandes cantidades de datos. Es de interés de este curso proporcionar al estudiante la parte introductoria en el tema de estadística, presentando una variedad de conceptos y principios estadísticos básicos en el contexto de ciencia de datos, y utilizar diferentes técnicas y herramientas para analizar información y así tomar decisiones de acuerdo a una base estadística sólida. En un principio se aborda el tema de estadística descriptiva, en la que se trabaja la síntesis y presentación de datos relevantes, que incluye tópicos como medidas de tendencia central y de dispersión, representaciones gráficas de datos y correlación entre variables. Mucho del razonamiento estadístico depende de la teoría de probabilidad, donde se incluyen temas como métodos de conteo, valor esperado, error estándar y su inclusión en modelos probabilísticos. Finalmente en estadística inferencial se realizan generalizaciones válidas de muestras de datos como la estimación de intervalos de confianza y pruebas de hipótesis, entre otras.

12.- Objetivos del programa

Proveer al estudiante de herramientas para evaluar afirmaciones estadísticas y la capacidad de realizar un análisis propio en grupos de datos de interés.

Desarrollar un razonamiento e intuición de diferentes acercamientos para distintos tipos de datos, y tomar decisiones acordes en base a métodos propuestos.

Al finalizar el curso el estudiante es capaz de:

- * Demostrar su comprensión de la estadística descriptiva mediante el resumen, organización, presentación y visualización de datos de una manera clara, concisa y capaz de proveer información relevante.
- * Verificar su conocimiento en probabilidad y la capacidad de identificar la distribución de variables aleatorias en escenarios simples, y calcular cantidades representativas tales como valores esperados, desviación estándar y probabilidades asociadas.
- * Comprobar su habilidad en la aplicación de inferencia estadística básica, realizando el cálculo de intervalos de confianza y aplicando pruebas de hipótesis apropiadas para realizar generalizaciones válidas de diversos conjuntos de datos.

13.-Contenido temático

Unidad 1. Estadística descriptiva

Unidad 2. Probabilidad

Unidad 3. Distribuciones de probabilidad

Unidad 4. Estadística Inferencial

Contenido desarrollado

Unidad 1: Estadística descriptiva

Objetivo particular: Se introducen métodos numéricos y gráficos para describir y mostrar datos de una muestra o población. El alumno aplica métodos gráficos para mostrar e interpretar datos, además de calcular medidas de tendencia central y de dispersión, e identifica valores atípicos en los datos con la finalidad de resumir y evaluar la información.

Desarrollo:

1. Tipos de variable
 - 1.1. Categórica
 - 1.2. Numérica
2. Niveles de medida
3. Población y muestra
4. Tipos de gráfica
 - 4.1. Gráfico circular
 - 4.2. Gráfico de barras
 - 4.3. Histograma
 - 4.4. Polígono de frecuencia
 - 4.5. Diagrama de tallo y hoja
 - 4.6. Diagrama de caja
5. Medidas de tendencia central
 - 5.1. Media
 - 5.2. Mediana
 - 5.3. Moda
6. Medidas de dispersión
 - 6.1. Rango
 - 6.2. Desviación media
 - 6.3. Varianza
 - 6.4. Desviación estándar
7. Coeficiente de variación
8. Datos no agrupados
9. Datos agrupados
 - 9.1. Rangos
 - 9.2. Amplitud
 - 9.3. Frecuencia absoluta
 - 9.4. Frecuencia relativa
 - 9.5. Frecuencia acumulada
 - 9.6. Frecuencia relativa acumulada
 - 9.7. Cuartiles y percentiles
 - 9.8. Rango intercuartil
 - 9.9. Valores atípicos
10. Teorema de Chebyshev
11. Asimetría
12. Curtosis
13. Coeficiente de correlación y covarianza
14. Regresión lineal
 - 14.1. Simple

- 14.2. Múltiple
- 14.3. Multicolinealidad

Unidad 2: Probabilidad

Objetivo particular: Se extiende la noción de variabilidad descrita anteriormente en el contexto de datos para utilizarla en diversas situaciones. Se exponen diferentes conceptos acerca de teoría de probabilidad, métodos de conteo y su aplicación en la toma de decisiones, necesarios para el desarrollo e interpretación de una inferencia estadística.

1. Teoría de probabilidad
2. Métodos de conteo
 - 2.1. Regla fundamental de conteo
 - 2.2. Permutaciones
 - 2.3. Combinaciones
3. Operaciones con conjuntos
 - 3.1. Intersecciones
 - 3.2. Uniones
 - 3.3. Complementos
4. Dependencia e independencia de eventos
5. Reglas de adición y multiplicación
6. Probabilidad condicional
7. Teorema de Bayes

Unidad 3: Distribuciones de probabilidad

Objetivo particular: Se definen y describen diferentes modelos probabilísticos. El alumno identifica la distribución de una variable aleatoria en escenarios simples y calcula valores numéricos representativos tales como valor esperado, varianza y probabilidades asociadas.

1. Variables aleatorias
2. Distribuciones de probabilidad
3. Esperanza o valor esperado
4. Varianza y desviación estándar
5. Distribución uniforme
6. Distribución geométrica
7. Distribución binomial
8. Distribución normal
 - 8.1. Valores z
 - 8.2. Regla empírica
9. Aproximación normal de una distribución binomial
10. Distribución de Poisson
11. Distribución hipergeométrica
12. Ley de Pareto

Unidad 4: Estadística inferencial

Objetivo particular: Se exponen conceptos básicos y herramientas de estadística inferencial. El alumno construye y determina intervalos y niveles de confianza para inferir parámetros de una población o muestra de interés, con el objetivo de formular y realizar pruebas de hipótesis apropiadas utilizando diversas técnicas estadísticas.

1. Muestreo
2. Distribución muestral de la media
3. Teorema del límite central
4. Estimación de intervalos de confianza
 - 4.1. Para media
 - 4.2. Para proporciones
5. Estimación de tamaño de muestra
6. Distribución T Student
7. Pruebas de hipótesis
 - 7.1. Hipótesis nula
 - 7.2. Hipótesis alternativa
 - 7.3. Pruebas de una y dos colas
8. Errores tipo I y tipo II
9. Pruebas de Chi cuadrada
 - 9.1. Prueba de bondad de ajuste
 - 9.2. Prueba de asociación e independencia
10. Análisis de varianza (ANOVA)

14. Actividades Prácticas

Discusión y resolución de ejercicios en línea y en el aula, basados en conceptos descritos durante las sesiones. Tanto en modalidad presencial como virtual la herramienta principal utilizada es el lenguaje de programación Python y su implementación en los cuadernos de código (notebooks) de Jupyter, dentro de la distribución Anaconda. El estudiante emplea algoritmos construidos en clase y construye los propios para ejemplificar los conceptos abordados. Con esto, el estudiante aprovecha y emplea una herramienta práctica y capaz en el análisis de datos, en concordancia también con el plan de estudios de la maestría en el tema de programación.

15.- Metodología

Para lograr los objetivos propuestos, en primer lugar se exponen los fundamentos y conocimientos teóricos necesarios en base al programa descrito. En conjunto con un método de enseñanza basada en problemas, posteriormente se presenta al estudiante con ejercicios particulares dentro del tema abordado durante en la sesión, donde el alumno determina el conocimiento y herramientas necesarias para resolverlo. En la discusión (oral, en pizarrón y mediante algoritmos), se guía al estudiante con estrategias para mejorar el aprendizaje y desarrollar la intuición para alcanzar los objetivos. Además, un panorama de aula invertida es utilizada para complementar el aprendizaje fuera del aula, mediante la revisión de material complementario (notas, algoritmos y bibliografía adicional) y la resolución individual de ejercicios, que luego serán revisados y discutidos en clase con el objetivo de evaluar y proveer retro-alimentación.

16.- Evaluación

Los estudiantes serán evaluados en el curso según el siguiente criterio:

Sets de ejercicios (2 sets por unidad) ... 40%

Necesario para desarrollar las habilidades en el análisis estadístico utilizando los conceptos vistos en clase, y recibir retroalimentación en el proceso de aprendizaje del estudiante. Por esta razón, esto toma una parte importante en la evaluación. Posteriormente se toma una parte del tiempo en clase para la discusión entre estudiantes y resolución de la tarea asignada.

Dos exámenes parciales ... 40%

Los exámenes son escritos y se aplican a la mitad y final del curso. El primer examen incluye las dos primeras unidades (estadística descriptiva y probabilidad) y el segundo las dos últimas unidades (distribuciones y estadística inferencial). Se aplican con el objeto de precisar la comprensión del estudiante sobre los temas abordados.

Proyecto final ... 20%

En lugar de un examen final, se realiza un proyecto final, de manera escrita y oral, con el propósito de evaluar de manera general la mayor parte de los temas expuestos y aplicados durante el curso. La actividad consiste en la elección de un set de datos concreto, identificar un objetivo o pregunta de interés y realizar un análisis estadístico completo para resolverlo, utilizando los conceptos y herramientas revisadas durante el curso.

17.- Bibliografía

1. Lind D., Marchal W., & Wathen S. (2017) “Statistical Techniques in Business and Economics”
2. McClave, Benson & Sincich (2018), Edición 13, “Statistics for Business and Economics”

Otros materiales:

1. <https://www.lynda.com/Business-Skills-tutorials/Statistics-Fundamentals-Part-1-Beginning/427473-2.html>
2. <https://www.lynda.com/Business-Intelligence-tutorials/Statistics-Fundamentals-Part-2-Intermediate/495322-2.html>
3. [Curso Udemy: https://www.udemy.com/course/probability-and-statistics-for-business-and-data-science/](https://www.udemy.com/course/probability-and-statistics-for-business-and-data-science/)
4. [Curso Udemy: https://www.udemy.com/course/machinelearningpython/](https://www.udemy.com/course/machinelearningpython/)
5. <http://math.oxford.emory.edu/site/math117/>

18.- Perfil del profesor

Profesor con nivel mínimo de maestría, de reconocido prestigio, con amplia experiencia en la docencia y análisis de datos.



19.- Nombre de los profesores que imparten la materia.

Dr. Eduardo de la Fuente Acosta

20.- Lugar y fecha de su aprobación (incluyendo la última actualización)

Ultima Actualización: Zapopan, Jalisco. 17 de Marzo de 2020

Ultima Modificación: Zapopan, Jalisco. 01 de Febrero de 2021

21.- Instancias que aprobaron el programa (Junta Académica y/o Coordinación del programa)

Junta Académica de la MCD