**Programa de Asignatura**

**Inferencia Estadística**

**A. Antecedentes Generales**

| 1. **Unidad Académica** | Facultad de Ingeniería | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1. **Carrera** | Ingeniería Civil Industrial | | | | | | |
| 1. **Código** | IIP225A | | | | | | |
| 1. **Ubicación en la malla** | 2 año, II semestre | | | | | | |
| 1. **Créditos** | 10 | | | | | | |
| 1. **Tipo de asignatura** | Obligatorio | X | Electivo |  | Optativo | |  |
| 1. **Duración** | Bimestral |  | Semestral | X | Anual | |  |
| 1. **Módulos semanales** | Clases Teóricas | 1 | Clases Prácticas | 1 | Ayudantía | | 1 |
| 1. **Horas académicas** | Clases | 68 | Ayudantía | | | 34 | |
| 1. **Pre-requisito** | Probabilidades y Estadística | | | | | | |

**B. Aporte al Perfil de Egreso**

En el curso de **Inferencia estadística**, perteneciente al ciclo de Bachillerato, los estudiantes comenzaran a introducirse en el mundo del modelado estadístico, pieza fundamental en el análisis de datos. Comprenderán la importancia de los test de hipótesis, como la clave para validar la relación entre variables tanto al caracterizar como al predecir. El estudiante será capaz de construir, seleccionar, validar, mejorar y reportar, modelos estadísticos fundamentales en la ingeniería, como los son los modelos lineales y algunos adicionales provenientes de la familia exponencial. Se espera que el estudiante conozca las métricas básicas de validación de un modelo, los alcances que tiene y también las limitaciones. El estudiante estudiara el modelo Logit, como una introducción a los modelos de clasificación y score hoy en día usados en las consultoras y la banca.

Se usará el software R en su interfaz Rstudio, en la manipulación, limpieza y presentación de datos, y en el ajuste, selección y validación de los modelos aprendidos. El estudiante al final del curso podrá llevar a cabo un proyecto de análisis de datos, con una variedad de herramientas tanto del tipo teóricas como prácticas las cuales marcaran la diferencia en el mundo laboral.

Consta de cinco unidades: Estimación puntual e intervalar, Test de hipótesis, Test de datos categóricos (χ 2)**.**, Regresión lineal múltiple y Modelos lineales generalizados.

El curso comenzara con el complemento a lo aprendido por los estudiantes en el curso anterior de probabilidades y estadística, completando su formación básica en estimación intervala y test de hipótesis sobre dos parámetros para finalizar con el test de Pearson. En este punto el estudiante reconocerá situaciones en donde el test de Pearson aplicará toda su potencia y simplicidad en el estudio de datos categóricos y bondad de ajuste. En el capítulo de modelos lineales y modelos lineales generalizados el estudiante comprenderá el concepto de regresión del punto de vista del modelo estadístico. Evidenciara como la estimación por máxima verosimilitud como método de estimación, entrega mayor información a la obtenida mediante el punto de vista geométrico puro.

El estudiante, además, estudiara en detalle la manera de validar un modelo y decidir de manera correcta cuando un modelo es mejor que otro. El estudiante también comprenderá la importancia de la ética en la manipulación de los datos, evitando, por ejemplo, el exceso de datos en un modelo.

Este curso pertenece al área formativa de ciencias básicas y tributa a las competencias genéricas de Eficiencia y Visión Analítica, así como aporta a la competencia específica de Resolución de problemas bajo un enfoque sistémico y Dominio de TIC´s para el desempeño de la profesión, declarada en el perfil de egreso de la carrera.

**C. Objetivos de Aprendizajes Generales de la asignatura**

* Reforzar y complementar los contenidos básicos de estimación y test de hipótesis del curso previo.

* Entregar al estudiante métodos estadísticos clásicos de manera de introducirlo al análisis de datos sin perder la rigurosidad y seriedad que la estadística necesita, pero con una mirada moderna con énfasis en un ajuste e interpretación correctos. Todo a través de un software apropiado y pertinente a las necesidades en la ingeniería de hoy y en virtud de la era digital y el proyecto UDD futuro.

**D. Unidades de Contenido y Objetivos de Aprendizaje**

| **Unidades de Contenidos** | **Objetivos de Aprendizaje** |
| --- | --- |
| **UNIDAD I: Estimación puntual e intervalar.**   * Intervalo de confianza para diferencia de medias. * Intervalo de confianza para el cociente de varianzas. * Intervalos de confianza para diferencia de proporciones. | * Reconocer un problema que involucra 2 poblaciones. * Reconocer los parámetros involucrados en un problema de 2 poblaciones. * Planear un modelo estadístico para estudiar la diferencia de medias o el cociente de varianzas. * Realizar una estimación con el nivel de confianza pedido. * Interpretar correctamente el intervalo de confianza construido. * Realizar estimación intervalar en un problema real. * Aprender las hipótesis bajo las cuales construimos nuestro intervalo de confianza y evidenciar la importancia de validar los supuestos. |
| **UNIDAD II: Test de hipótesis**   * Test de hipótesis para diferencia de medias. * Test de hipótesis sobre diferencia de proporciones. * Test de hipótesis sobre cociente de varianzas. | * Reconocer un problema que involucra comparar 2 poblaciones. * Reconocer los parámetros involucrados en la comparación de 2 poblaciones. * Realizar un contraste de hipótesis de diferencia de medias o cociente de varianzas. * Aplicar la teoría aprendida a problemas reales. * Aprender las hipótesis bajo las cuales construimos nuestro test de hipótesis y evidenciar la importancia de validarlas. |
| **UNIDAD III: Test sobre datos categóricos (χ 2).**   * Bondad de ajuste (caso continuo y discreto). * Independencia en Tablas de contingencia. * Homogeneidad. | * Reconocer un problema de datos que involucra categóricos. * Conocer el estadístico de Pearson y reconocer su simplicidad y efectividad como test. * Estudiar la distribución del estadístico de Pearson y calcular correctamente sus grados de libertad. * Estudiar la bondad de ajuste de un conjunto de datos a un modelo pedido. * Estudiar la relación entre 2 variables en una tabla de contingencia mediante el test de independencia. * Resolver problemas reales usando el test de Pearson. |
| **UNIDAD IV: Regresión Lineal Múltiple.**   * Definición y supuestos del modelo de regresión lineal múltiple. * El método de Mínimos Cuadrados Ordinarios. * Evaluación del ajuste de un modelo de Regresión Lineal. * Residuos. * Intervalos de Predicción. * Selección Backward y Forward. | * Estudiar el concepto de regresión. * Relacionar el concepto de combinación lineal con la media, como definición de una regresión lineal. * Estudiar el modelo normal asociado y relacionar los test de hipótesis aprendidos con el ajuste del modelo. * Realizar un estudio de la calidad del ajuste del modelo a través de los test de hipótesis asociados. * Seleccionar el “mejor modelo” y realizar interpretación de sus coeficientes en un contexto real. * Realizar predicciones basados en los intervalos de confianza aprendidos en las secciones pasadas. * Conocer las limitantes de los modelos lineales. |
| **Unidad V: Modelos Lineales Generalizados.**   * Ajuste de los Modelos de Regresión. Logística y Poisson. * Estimación y significancia. * Selección del Modelo. * Interpretación de los Coeficientes. | * Estudiar la relación que hay entre los parámetros y la regresión lineal vía una transformación apropiada. * Construir un modelo de regresión logística y poisson, comprendiendo la necesidad de generalizar el concepto de regresión. * Ajustar y seleccionar de manera apropiada los modelos de regresión general. * Interpretar los coeficientes de manera correcta con respecto al modelo seleccionado. * Estudiar un problema real y realizar un modelo apropiado. |

**E. Estrategias de Enseñanza**

Durante el desarrollo del curso, se procederá a impartir el contenido teórico de la asignatura a través de clases expositivas y la resolución de ejercicios. El desarrollo de dichas clases estará basado fundamentalmente en la lección magistral, motivando el análisis y discusión de los conceptos fundamentales ilustrados en los ejemplos, desarrollando sus consecuencias y mostrando sus aplicaciones. Lo anterior será complementado con el uso de un software apropiado y pertinente a los estándares pedidos por la industria hoy en día. Se usará un software libre y gratuito, de fácil instalación. Todas las clases se complementarán con el uso de este como un medio facilitador del aprendizaje del estudiante, El estudiante deberá evidenciar y contrastar el conocimiento teórico con el practico a través de del. Se realizarán ejemplos mostrando la importancia del uso correcto y eficiente de este como medio de reporte de resultados. En el uso del software se hará hincapié en la importancia de la interpretación por sobre la belleza del código. El estudiante tendrá acceso a una variada bibliografía proporcionada por los docentes con código limpio y repetible.

Talleres de resolución de problemas y talleres con uso de TIC donde los estudiantes podrán trabajar de manera individual y/o grupal. Se impartirán también ayudantías en las que se resolverán ejercicios y problemas indicados y supervisados por el profesor.

**F. Estrategias de Evaluación**

Las instancias de evaluación son las siguientes:

**Controles**: 5 controles con la posibilidad de eliminar 1.

**Certámenes**: 2 certámenes con una parte presencial y otra para la casa en software con una entrega de 48 horas máximo.

**Talleres:** Se realizarán talleres complementarios a las clases, en donde se tratarán de detalle algunos ejemplos usando software y algunos modelos clásicos de la literatura.

**Examen**: se realizará 1 examen, al término del semestre, en la fecha establecida por la Facultad, y exigiéndose nota mínima de 3.0, para todos los estudiantes, según R.A.A.R.

**G. Recursos de Aprendizaje**

**Obligatorio**

* Probabilidades y Estadística, Mendenhal-Beaver-Beaver
* Apunte generado por los docentes UDD.
* Codigo R generado por profesores UDD.

**Complementario**

* Estadística matemática con aplicaciones Séptima Edición Wackerly, Dennis D.-William Mendenhall III-Richard L. Scheaffer
* Foundations of Linear and Generalized Linear Models – Alan Agresti. WILEY SERIES IN PROBABILITY AND STATISTICS.