



## 1. Datos Generales de la asignatura

<b>Nombre de la asignatura:</b>	Estadística para ciencia de datos
<b>Clave de la asignatura:</b>	CDD-2406
<b>SATCA<sup>1</sup>:</b>	2-3-5
<b>Carrera:</b>	Ingeniería en Ciencias de Datos

## 2. Presentación

<b>Caracterización de la asignatura</b>
Esta asignatura contribuye al perfil del ingeniero en ciencia de datos proporcionándole herramientas metodológicas para procesar y analizar una cantidad muy grande de datos, de tal manera que identifique patrones, relaciones o tendencias que le permitan modelar fenómenos o procesos de interés, con un ajuste y nivel de confianza satisfactorios, garantizando una eficaz toma de decisiones.
<b>Intención didáctica</b>
El contenido de esta asignatura es una continuación del último tema abordado en estadística inferencial (regresión lineal y logística), el cual puede considerarse como un primer acercamiento para modelar la relación existente entre una o más variables regresoras con respecto a una variable de respuesta (o variable de interés). En estadística para ciencia de datos, se abordarán otro tipo de métodos que pueden servir para modelar esta relación. En los primeros cuatro temas, se mantiene el enfoque de evaluar el efecto que una o más variables (o factores) pueden tener sobre una variable de interés. Sin embargo, en la última unidad, se abre la posibilidad a incluir más de una variable de respuesta y se dan a conocer, además, herramientas de estadística multivariada, que permiten analizar de manera sistémica la relación existente entre todas las variables que forman parte del proceso o fenómeno de estudio, específicamente, aquellas en las que se tiene especial interés.
En el tema 1, se estudia el comportamiento de una variable con respecto al tiempo, empleando diferentes modelos de series de tiempo y pronósticos que, a través de la descomposición de la serie, nos permite identificar el modelo que mejor se ajusta a los datos.
En el tema 2, se introduce al estudiante en la teoría y metodología del diseño estadístico experimental, bajo el enfoque de diseño factorial; empezando con el caso más simple en el que se varían los niveles de un factor de interés, para observar y analizar el efecto que estos cambios tienen en la variable de respuesta. En este tema se presenta la metodología de análisis estadístico, propio del diseño experimental, la cual se seguirá aplicando en los tres temas siguientes.
En el tema 3, se continúa con el diseño factorial, incluyendo dos o más factores de interés, los cuales se varían sistemáticamente, siguiendo los principios fundamentales de experimentación estadística, para evaluar los cambios que las diferentes combinaciones de prueba producen en la variable de respuesta.

<sup>1</sup> Sistema de Asignación y Transferencia de Créditos Académicos



En el tema 4, se limitan las condiciones de prueba de cada factor de interés, de manera que cada factor se varía únicamente en dos condiciones diferentes y se prueban todas las posibles combinaciones de todos los factores, para los cuáles nos interesa estudiar su efecto sobre la variable de respuesta.

En el último tema, se presentan al estudiante las técnicas fundamentales de análisis estadístico multivariado, que permitirán analizar aquellos casos en los que nos interesa conocer el efecto que uno o más factores tienen sobre más de una variable de respuesta, así como la forma en que interactúan todas las variables entre sí.

Es importante mencionar, que cada uno de los temas de esta asignatura, se abordarán con un nivel de profundidad mínimo, suficiente para que el estudiante conozca el alcance y característica de cada uno de los métodos y el uso que se le puede dar para resolver casos de aplicación de ciencia de datos, sin profundizar en todos los posibles casos que se pueden presentar dentro de cada tema, ni pretender lograr un completo dominio de los conceptos e implicaciones teóricas de estos métodos.

Al igual que en estadística 1, el docente que imparta esta asignatura, debe introducir al estudiante en el uso de software estadístico especializado, de mayor uso a nivel internacional, considerando los diferentes campos en que puede desempeñarse el ingeniero en ciencia de datos, sin limitarse al uso de un software en particular. En este contexto, se considera indispensable el uso de R y R-Studio y, adicionalmente, cualquier otro que el docente considere pertinente.

Finalmente, las competencias adquiridas en esta asignatura, son indispensables para cursar asignaturas como Investigación de operaciones, aprendizaje automático, big data y visualización de datos.

### 3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto Tecnológico Superior de Alvarado del 21 al 23 agosto de 2023.	Representante del Instituto Tecnológico Superior de Alvarado.	Propuesta inicial.
Tecnológico Nacional de México 30 octubre 2023	Representante del Instituto Tecnológico de: Querétaro y del Instituto Tecnológico Superior de Alvarado.	Presentación de la propuesta de la carrera de Ingeniería en Ciencia de Datos.
Instituto Tecnológico de Querétaro Campus Norte del 19 al 22 de marzo 2024.	Representantes de los Institutos Tecnológicos de: Morelia, Puebla, Querétaro, Tehuacán. Instituto Tecnológico Superior de Alvarado. CENIDET. Representante de Ciencias Básica de los Institutos de: Celaya, Morelia y CIIDET.	Diseño y/o desarrollo curricular de la carrera de Ingeniería en Ciencia de Datos.



Tecnológico Nacional de México del 22 al 24 de abril del 2024	Representante del Instituto Tecnológico de Querétaro e Instituto Tecnológico Superior de Alvarado.	Contraste y ajuste de las asignaturas de Ingeniería en Ciencia de Datos con respecto a las de Ing. en Inteligencia Artificial, Ing. en Desarrollo WEB e Ing. en Ciberseguridad
Tecnológico Nacional de México del 27 al 31 de mayo del 2024.	Representantes de los Institutos Tecnológicos de: Morelia, Querétaro. Instituto Tecnológico Superior de Alvarado. CENIDET.	Consolidación curricular de la carrera de Ingeniería Ciencia de Datos

#### 4. Competencia(s) a desarrollar

Competencia(s) específica(s) de la asignatura
Identifica las fuentes de variación controlables o no controlables que influyen significativamente sobre una o más variables de respuesta, para la obtención de un modelo estadístico que permita explicar la relación e interacción entre estas variables; con un ajuste satisfactorio, un nivel de confianza aceptable y un efecto mínimo del error aleatorio.

#### 5. Competencias previas

<ul style="list-style-type: none"> <li>● Seleccionar una muestra representativa de un proceso o caso de interés, usando el método probabilístico adecuado, de acuerdo a las características poblacionales.</li> <li>● Obtención e interpretación de medidas y gráficos descriptivos.</li> <li>● Estimación puntual y por intervalos.</li> <li>● Planteamiento y prueba de hipótesis estadísticas.</li> <li>● Análisis, interpretación e informe de los resultados obtenidos en un estudio de inferencia estadística.</li> <li>● Obtener un modelo regresión lineal o logística que se ajuste satisfactoriamente a un conjunto de datos.</li> <li>● Verificación de supuestos estadísticos.</li> <li>● Conceptos de análisis de varianza, fuente de variación y error aleatorio.</li> <li>● Uso y aplicación de funciones de densidad y distribuciones de muestreo, para inferencia estadística.</li> <li>● Trabajo en equipo.</li> <li>● Formulación y solución de problemas de ingeniería, a través de matrices y sistemas de ecuaciones.</li> <li>● Comunicación oral y escrita.</li> <li>● Uso de Software Estadístico R y R-Studio, entre otros.</li> </ul>
---



## 6. Temario

No.	Temas	Subtemas
1	Series de tiempo.	1.1. Introducción a las Series de tiempo. 1.1.1. Características de una serie de tiempo. 1.1.2. Descomposición de una serie de tiempo. 1.1.3. Series de tiempo estacionarias. 1.2. Modelos para series de tiempo. 1.2.1. Correlación y autocorrelación. 1.2.2. Modelos AR y ARIMA. 1.3. Pronóstico para series de tiempo. 1.3.1. Suavizamiento exponencial. 1.3.2. Método Holt-Winters.
2	Diseño de experimentos con un factor	2.1. Introducción al diseño de experimentos. 2.2. Diseño factorial de “una vía”. 2.3. Diseño de bloques. 2.4. Análisis de residuales. 2.5. Uso de software estadístico.
3	Modelo lineal general	3.1. Diseño de experimentos con dos factores. 3.2. Diseño de experimentos con tres factores. 3.3. Uso de software estadístico
4	Diseños $2^k$	4.1. Diseños completos. 4.2. Diseños fraccionados. 4.3. Análisis superficies de respuestas. 4.4. Uso de software estadístico.
5	Estadística multivariada	5.1. Introducción a la estadística multivariada. 5.2. Distribución normal multivariada. 5.3. Análisis de varianza multivariado. 5.4. Análisis de componentes.



## 7. Actividades de aprendizaje de los temas

1.-Series de tiempo	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i> Conoce y aplica los conocimientos de pronósticos y series de tiempo para obtener el modelo que mejor se ajusta al comportamiento de una variable con respecto al tiempo, para una eficaz toma de decisiones.</p> <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> <li>● Capacidad de abstracción, análisis y síntesis.</li> <li>● Capacidad para identificar, plantear y resolver problemas</li> <li>● Búsqueda de información.</li> <li>● Recopilación de datos.</li> <li>● Trabajo en equipo.</li> <li>● Interpretación de resultados.</li> <li>● Comunicación oral y escrita.</li> <li>● Uso de TIC's.</li> </ul>	<ul style="list-style-type: none"> <li>● Buscar información en diferentes fuentes, sobre series de tiempo.</li> <li>● Explicar los componentes de una serie de tiempo.</li> <li>● Elaborar un mapa conceptual para los diferentes métodos de pronósticos.</li> <li>● Recopilar datos de un caso real y hacer el pronóstico del comportamiento de una variable con respecto al tiempo.</li> <li>● Exponer ante el grupo el resultado de su caso práctico.</li> <li>● Usar software estadístico para descomponer series de tiempo, obtener pronósticos e identificar características de estacionalidad.</li> </ul>
2. Diseño de experimentos con un factor.	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i></p> <ul style="list-style-type: none"> <li>● Conoce y usa los conceptos fundamentales del análisis de varianza para probar la significancia del efecto de un factor sobre una variable de respuesta.</li> <li>● Conoce y aplica los métodos de pruebas de hipótesis para comparaciones de medias, para identificar diferencias significativas entre los niveles del factor.</li> <li>● Identifica factores perturbadores y prueba la significancia de su efecto sobre la variable de respuesta.</li> <li>● Uso de software estadístico para desarrollar el análisis de varianza y pruebas de significancia.</li> </ul> <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> <li>● Capacidad de abstracción, análisis y síntesis.</li> <li>● Capacidad para identificar, plantear y resolver problemas.</li> </ul>	<ul style="list-style-type: none"> <li>● Elaborar un mapa conceptual para representar los diferentes modelos de diseño estadístico experimental.</li> <li>● Investigar en diversas fuentes, los elementos de los diseños completamente al azar y los diseños con bloques aleatorizados.</li> <li>● Formular y describir las diversas pruebas de comparación de medias que existen.</li> <li>● Investigar aplicaciones del diseño factorial en casos de estudio de ciencias de datos.</li> <li>● Trabajar en equipo para plantear un caso de aplicación a través de un diseño experimental con un solo factor, hacer las pruebas correspondientes, siguiendo los principios básicos del diseño experimental y exponer ante el grupo los resultados de su caso práctico.</li> <li>● Analizar los resultados de su caso práctico, con por lo menos dos opciones de software estadístico.</li> </ul>



<ul style="list-style-type: none"> <li>● Búsqueda de información.</li> <li>● Recopilación de datos.</li> <li>● Trabajo en equipo.</li> <li>● Interpretación de resultados.</li> <li>● Comunicación oral y escrita.</li> <li>● Uso de TIC's.</li> </ul>	
<b>3. Modelo lineal general</b>	
<b>Competencias</b>	<b>Actividades de aprendizaje</b>
<p><i>Específica(s):</i></p> <ul style="list-style-type: none"> <li>● Conoce y aplica los fundamentos teóricos de diseño de experimentos con dos, tres o más factores.</li> <li>● Prueba la significancia del efecto de cada factor y de sus interacciones, sobre la variable de respuesta.</li> <li>● Verifica los supuestos estadísticos correspondientes al modelo lineal general, para probar la adecuación del modelo.</li> <li>● Gráfica y analiza los efectos individuales y de interacción de los factores de interés.</li> </ul> <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> <li>● Capacidad de abstracción, análisis y síntesis.</li> <li>● Capacidad para identificar, plantear y resolver problemas</li> <li>● Búsqueda de información.</li> <li>● Recopilación de datos.</li> <li>● Trabajo en equipo.</li> <li>● Interpretación de resultados.</li> <li>● Comunicación oral y escrita.</li> <li>● Uso de TIC's.</li> </ul>	<ul style="list-style-type: none"> <li>● Describir los conceptos básicos y notación usados en diseño factorial.</li> <li>● Investigar y explicar en grupos de tres, la diferencia entre modelo factorial de efectos fijos y modelo de efectos aleatorios.</li> <li>● Usar software estadístico para analizar experimentos factoriales con dos o más factores incluidos en las fuentes de información de la asignatura.</li> </ul>
<b>4. Diseños <math>2^k</math></b>	
<b>Competencias</b>	<b>Actividades de aprendizaje</b>
<p><i>Específica(s):</i></p> <ul style="list-style-type: none"> <li>● Conoce y aplica las características de los diseños factoriales <math>2^k</math> para probar la significancia de cada efecto individual de los factores y de sus interacciones.</li> <li>● Verifica los supuestos estadísticos correspondientes a los diseños factoriales <math>2^k</math>, para probar la adecuación del modelo.</li> <li>● Grafica y analiza los efectos individuales y de interacción de los "K" factores incluidos en el modelo.</li> </ul>	<ul style="list-style-type: none"> <li>● Describir los conceptos básicos y notación usados en diseño factorial <math>2^k</math>.</li> <li>● Investigar y explicar en grupos de tres, la diferencia entre diseños <math>2^k</math> completos y fraccionados.</li> <li>● Usar software estadístico para analizar experimentos factoriales <math>2^k</math> completos y fraccionados incluidos en las fuentes de información de la asignatura.</li> </ul>



<ul style="list-style-type: none"> <li>● Con base en las gráficas de efectos principales y de interacción, identifica los niveles óptimos de cada factor significativo incluido en el modelo.</li> <li>● Conoce y aplica los conceptos fundamentales del diseño fraccionado, tales como resolución, confusión y alias.</li> </ul> <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> <li>● Capacidad de abstracción, análisis y síntesis.</li> <li>● Capacidad para identificar, plantear y resolver problemas</li> <li>● Trabajo en equipo.</li> <li>● Interpretación de resultados.</li> <li>● Comunicación oral y escrita.</li> <li>● Uso de TIC's.</li> </ul>	
<b>5. Estadística multivariada</b>	
<b>Competencias</b>	<b>Actividades de aprendizaje</b>
<p><i>Específica(s):</i> Aplica métodos de estadística multivariada para identificar variables que interactúan entre sí y hace inferencias respecto al tipo y fuerza de la relación entre ellas.</p> <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> <li>● Capacidad de abstracción, análisis y síntesis.</li> <li>● Capacidad para identificar, plantear y resolver problemas</li> <li>● Trabajo en equipo.</li> <li>● Interpretación de resultados.</li> <li>● Comunicación oral y escrita.</li> <li>● Uso de TIC's.</li> </ul>	<ul style="list-style-type: none"> <li>● Analizar mediante mapas conceptuales, con base en lecturas sugeridas por el profesor, los elementos teóricos fundamentales del análisis multivariado y la clasificación de los métodos más comúnmente usados en ciencia de datos.</li> <li>● Investigar casos de aplicación de estadística multivariada en ciencia de datos. Identificando el método de análisis multivariado aplicado en cada caso.</li> <li>● Usar un software estadístico para resolver los casos de aplicación identificados en la investigación anterior y presentar los resultados ante el grupo.</li> </ul>



## 8. Práctica(s)

- Para los temas 1, 2, 3 y 4, resolver al final del tema, un ejercicio sugerido por el profesor, en Excel, Minitab y R-Studio y comparar los resultados obtenidos. A partir de estos resultados, elaborar una tabla comparativa de ventajas y desventajas del uso de cada software para el análisis de los datos.
- Generar una carpeta en la nube de drive, con las cuentas institucionales de los estudiantes, donde puedan compartir sus tablas comparativas del uso de software en los temas 1, 2, 3 y 4.
- Generar una carpeta en la nube de drive, con las cuentas institucionales de los estudiantes, donde puedan compartir el análisis estadístico de sus casos prácticos de estadística multivariada, desarrollados en el tema cinco.
- Al final de la asignatura, subir a la nube de drive un ensayo sobre las fortalezas y debilidades de cada método estudiado, así como de cada software estadístico empleado para el análisis de los datos.

## 9. Proyecto de asignatura

El objetivo del proyecto que planteé el docente que imparta esta asignatura, es demostrar el desarrollo y alcance de la(s) competencia(s) de la asignatura, considerando las siguientes fases:

**Fundamentación:** marco referencial (teórico, conceptual, contextual, legal) en el cual se fundamenta el proyecto de acuerdo con un diagnóstico realizado, mismo que permite a los estudiantes lograr la comprensión de la realidad o situación objeto de estudio para definir un proceso de intervención o hacer el diseño de un modelo.

**Planeación:** con base en el diagnóstico en esta fase se realiza el diseño del proyecto por parte de los estudiantes con asesoría del docente; implica planificar un proceso: de intervención empresarial, social o comunitario, el diseño de un modelo, entre otros, según el tipo de proyecto, las actividades a realizar los recursos requeridos y el cronograma de trabajo.

**Ejecución:** consiste en el desarrollo de la planeación del proyecto realizada por parte de los estudiantes con asesoría del docente, es decir en la intervención (social, empresarial), o construcción del modelo propuesto según el tipo de proyecto, es la fase de mayor duración que implica el desempeño de las competencias genéricas y específicas a desarrollar.

**Evaluación:** es la fase final que aplica un juicio de valor en el contexto laboral-profesional, social e investigativo, ésta se debe realizar a través del reconocimiento de logros y aspectos a mejorar se estará promoviendo el concepto de “evaluación para la mejora continua”, la metacognición, el desarrollo del pensamiento crítico y reflexivo en los estudiantes.

## 10. Evaluación por competencias

- Las técnicas, herramientas y/o instrumentos sugeridos que permiten obtener el producto del desarrollo las actividades de aprendizaje: mapas conceptuales, reportes de prácticas, estudios de casos, exposiciones en clase, ensayos, problemas resueltos, reportes de visitas, portafolio de evidencias, proyecto de asignatura o integrador y cuestionarios.
- Las técnicas, herramientas y/o instrumentos sugeridos que me permiten constatar el logro o desempeño de las competencias del estudiante: listas de cotejo, listas de verificación, matrices de valoración, guías de observación, coevaluación y autoevaluación.





## 11. Fuentes de información

1. Gutiérrez Pulido, Humberto; De la Vara Salazar, Román. (2012). Análisis y Diseño de Experimentos. 3ª edición McGraw-Hill/Interamericana. México.
2. Montgomery, Douglas C. (2019). Design and Analysis of Experiments. 10th edition. John Wiley & Sons. New Jersey, U.S.A.
3. Toutenburg H., Shalabh. (2009). Statistical Analysis of Designed Experiments. Third Edition. Springer Texts in Statistics. New York, USA.
4. Walpole, Roland E.; Raymond H. Myers, Sharon L. Myers, Keying Ye (2012). Probabilidad y Estadística. 9ª. Edición. Pearson Educación de México. México.
5. Box, George E. P., J. Stuart Hunter, William G. Hunter. (2005). Statistics for Experimenters, Design, Innovation, and Discovery. 2<sup>nd</sup>.
6. Díaz, Luis G., Morales, Mario A. (2012). Análisis estadístico de datos multivariados. 1ª edición. Universidad Nacional de Colombia. Bogotá D. C., Colombia.
7. Everitt, B., & Hothorn, T. (2011). An Introduction to Applied Multivariate Analysis with R. Springer Science & Business Media. USA.