



**Facultad de  
Ciencias**  
UNAM



## **Diplomado en Técnicas y Modelos de la Estadística para el Análisis de Datos**

### **Instructores**

Profesor Francisco Sánchez Villarreal.  
Profesor Benjamín Figueroa Solano.  
Profesor Miguel Ángel Chong Rodríguez.  
Profesora Susana Barrera Ocampo.

### **Estructura académica**

Coordinador Académico del Diplomado: Profesor Francisco Sánchez Villarreal.

El Diplomado tiene una duración de 160 horas distribuidas de la siguiente forma:

Módulo I: Probabilidad nivel básico (32 horas).

Profesor Benjamín Figueroa Solano

Módulo II: Fundamentos y métodos de la Estadística (32 horas).

Profesor Miguel Ángel Chong Rodríguez

Módulo III: Métodos estadísticos no paramétricos (32 horas).

Profesor Francisco Sánchez Villarreal

Módulo IV: Análisis de regresión y series de tiempo (32 horas).

Profesora Susana barrera Ocampo

Módulo V: Análisis estadístico (32 horas).

Profesor Francisco Sánchez Villarreal

## **TEMARIO**

### **MÓDULO I: PROBABILIDAD NIVEL BÁSICO.**

#### **Tema 1. Espacios de probabilidad.**

- 1.1. Espacio muestral, eventos y su interpretación.
- 1.2. Panorama histórico de la probabilidad, interpretación frecuentista, definición clásica.
- 1.3. Definición axiomática de probabilidad (sin énfasis en sigma-álgebras).
- 1.4. Técnicas de conteo.
- 1.5. Probabilidad condicional e independencia.
- 1.6. Fórmulas de probabilidad total y de Bayes.
- 1.7. Teorema de continuidad de la probabilidad.

## **Tema 2. Variables aleatorias y funcionales de distribución.**

- 2.1. Definición de variable aleatoria.
- 2.2. Distribución, esperanza y varianza de una variable aleatoria finita.
- 2.3. Función de distribución y sus propiedades.
- 2.4. Momentos, función generatriz de momentos.
- 2.5. Desigualdad de Tchebychev, Ley de los Grandes Números.
- 2.6. Funciones de distribución de variables aleatorias discretas.
  - 2.6.1. Bernoulli
  - 2.6.2. Binomial
  - 2.6.3. Poisson
  - 2.6.4. Uniforme discreta
  - 2.6.5. Geométrica
  - 2.6.6. Hipergeométrica
  - 2.6.7. Binomial Negativa
  - 2.6.8. Ejemplos. Aplicaciones y generación de valores por simulación
- 2.7. Funciones de distribución de variables aleatorias continuas.
  - 2.7.1. Uniforme continua
  - 2.7.2 Normal
  - 2.7.3. Exponencial Negativa
  - 2.7.4. Gamma
  - 2.7.5. Beta
  - 2.7.6. Cauchy
  - 2.7.7. Pareto
  - 2.7.8. Weibull
  - 2.7.9. Frechet
  - 2.7.10. Logística
  - 2.7.11. Gausiana Inversa
  - 2.7.12. Ejemplos. Aplicaciones y generación de valores de simulación

**Forma de evaluar:**

Tareas: 70%

Trabajo final: 30%

**MÓDULO II: FUNDAMENTOS Y MÉTODOS DE LA ESTADÍSTICA.****Descripción.**

Como resultado de los avances tecnológicos, las organizaciones disponen de grandes volúmenes de información relacionada con el desarrollo de su actividad, un siguiente paso es realizar su análisis que ponga al descubierto hallazgos que se conviertan en los insumos para la toma de decisiones y diseño de estrategias que contribuyan a alcanzar los objetivos. A pesar de que se cuentan con herramientas para analizar de forma ágil y vistosa grandes volúmenes de datos, continúa siendo necesario el uso de la Estadística Descriptiva e Inferencial para entenderlos y transformarlos en información, conocimiento y sabiduría.

**Objetivo del Módulo:**

- Capacitar al participante en las técnicas y métodos más usuales de la Estadística Descriptiva e Inferencial.
- Capacitar al participante en analizar información con apoyo de R, SPSS y Hoja de Cálculo.

**Tema 1. Introducción al proceso de análisis estadístico.**

- 1.1 Historia y estado actual de la Estadística.
- 1.2 Conceptos de población, muestra, censo y muestreo.
- 1.3 Tipos de estadística y proceso de análisis estadístico.

**Tema 2. Análisis exploratorio de datos.**

- 2.1 Tablas y gráficas.
- 2.2 Medidas descriptivas.

### **Tema 3. Estimación puntual.**

- 3.1 Estadística y estimadores.
- 3.2 Criterios de evaluación de estimadores.
  - 3.2.1 Sinsesgamiento
  - 3.2.2 Eficiencia
  - 3.2.3 Consistencia
  - 3.2.4 Suficiencia, estadísticas suficientes
- 3.3 Métodos de construcción de estimadores.
  - 3.3.1 Momentos
  - 3.3.2 Máxima verosimilitud
  - 3.3.3 Mínimos cuadrados
- 3.4 Estimación inesperada
  - 3.4.1 Cota inferior de Cramér-Rao
  - 4.4.2 Suficiencia y completez

### **Tema 4. Estimación por intervalos.**

- 4.1 Intervalo aleatorio e intervalo de confianza
- 4.2 Métodos para construir un intervalo de confianza.
  - 4.2.1 Método pivotal.
  - 4.2.2 Método general

### **Tema 5. Pruebas de hipótesis.**

- 5.1 Hipótesis estadística: simple y compuesta
- 5.2 Región crítica
- 5.3 Errores tipo I y II
- 5.4 Función potencia
- 5.5 Lema de Neyman-Pearson
- 5.6 Prueba uniformemente más potente
- 5.7 Prueba del cociente de verosimilitud

## MÓDULO III: MÉTODOS ESTADÍSTICOS NO PARAMÉTRICOS.

### Descripción.

La Estadística No Paramétrica es una rama de la estadística inferencial que estudia los modelos y pruebas a partir de muestras cuyas distribuciones no pueden ser definidas a priori y cuyos niveles de medición suelen ser del tipo categórico u ordinal. Recientemente se han incorporado a su estudio temas de muestreos repetidos mediante diversas técnicas de Montecarlo como Bootstrap. La Estadística No Paramétrica tiene como ámbitos de aplicación la Sociología, Politología, Investigación de Mercados, Psicología, Clínica Social, etc.

### Objetivo del Módulo:

- Capacitar al participante en las técnicas y métodos más usuales de la Estadística No Paramétrica en diversos ámbitos de aplicación.
- Capacitar al participante en técnicas y procedimientos de cálculo con apoyo de Hoja de Cálculo, Paquete SPSS y Lenguaje de Programación R.

### Tema 1. Introducción y Pruebas para Muestras Relacionadas.

- 1.1 Concepto de medición y Niveles de medición.
- 1.2 Prueba de Signos para 2 Muestras Relacionadas.
- 1.3 Prueba de Wilcoxon para 2 muestras Relacionadas.
- 1.4 Prueba de Friedman para K Muestras Relacionadas.
- 1.5 Coeficiente de Correlación de Spearman.

### Tema 2. Introducción y Pruebas para Muestras Independientes.

- 2.1 Prueba U de Mann Whitney para 2 Muestras Independientes
- 2.2 Prueba de la Mediana para 2 Muestras Independientes
- 2.3 Prueba de Kruskal-Wallis para K Muestras Independientes (ANOVA)

### Tema 3. Análisis de Tablas de Contingencia.

- 3.1. Pruebas de Independencia entre 2 variables. Tablas RxC
- 3.2. Estadísticos de Relación. Coeficiente de Contingencia, Phi de Pearson, V de Cramer, Tau B y Tau C
- 3.3. Pruebas de Igualdad de K Proporciones

3.4. Pruebas de Mediana para K Muestras

3.5. Prueba de McNemar

#### **Tema 4. Pruebas de Bondad de Ajuste**

4.1 Prueba Chi Cuadrada de Bondad de Ajuste

4.2 Prueba de Kolmogorov Smimov para 1 y 2 Muestras

4.3 Prueba de Normalidades Jarque Bera y Lilifors

4.4 Prueba de Rachas para Verificar Aleatoriedad de Series

#### **Tema 5. Técnicas basadas en Remuestreo**

5.1 Método de Jackknife

5.2 Método Bootstrap para Intervalos de Confianza (Mediana, Varianza)

#### **Operación del Módulo.**

Clases a distancia con exposición de fundamentos teóricos del Tema y ejemplos de aplicaciones con apoyo de Hoja de Cálculo de Excel, Lenguaje R y SPSS

#### **Auxiliares Didácticos.**

Notas de los temas en archivos PDF

Ejercicios y ejemplos en Excel y formato separado por comas CSV

Archivos de Programas Fuente en R (Script)

#### **Evaluación.**

Tareas: 30%

Trabajo y Ejercicios Finales: 70%

## MODULO IV: ANÁLISIS DE REGRESIÓN Y SERIES DE TIEMPO.

### Objetivos del Módulo:

- El participante utilizará modelos de regresión lineal para explicar el comportamiento de una variable en función de otras existentes en el mismo contexto.
- El Participante aplicará modelos de Series de Tiempo para el análisis y predicción del comportamiento de una variable que es observada durante un periodo de tiempo.

### Tema 1. Modelo Lineal General.

- 1.1 Definición y supuestos del modelo lineal general
- 1.2 Estimadores de mínimos cuadrados y sus propiedades
- 1.3 Teorema de Gauss Marcov
- 1.4 Distribución de los estimadores
- 1.5 Inferencia sobre los parámetros
- 1.6 Coeficientes de correlación simple parcial y múltiple y de determinación.
- 1.7 Análisis de varianza en regresión.
- 1.8 Predicción
- 1.9 Transformaciones, polinomios y variables dummy
- 1.10 Violaciones a los supuestos del modelo multicolinealidad
- 1.11 Violaciones a los supuestos del modelo heteroscedasticidad
- 1.12 Análisis de residuales y valores influyentes

### Tema 2. Análisis de Series de Tiempo.

- 2.1 Definición de una serie de tiempo
- 2.2 Procesos estocásticos y series de tiempo
- 2.3 Operadores y polinomios
- 2.4 Filtros lineales
- 2.5 Procesos estacionarios
- 2.6 Modelos autorregresivos AR(p)
- 2.7 Modelos de medias móviles
- 2.8 Modelos de autorregresivos y de medias móviles ARMA(p,d,q)
- 2.9 Modelos integrados autorregresivos y medias móviles ARIMA (p,d,q)

2.10 Construcción de modelos (identificación, estacionarización, estimación)

2.11 Modelos estacionales ARIMA (p,d,q) (P,D,Q)

## MÓDULO V: ANÁLISIS ESTADÍSTICO MULTIVARIADO.

### Objetivos del Módulo:

- El participante conocerá las bases metodológicas del análisis multivariado.
- Conocerá los conceptos de distancia y similitud asociados a la clasificación numérica del análisis de conglomerados y escalas multidimensionales.
- Aplicará el análisis discriminante clásico para la clasificación de elementos a partir de vectores multidimensionales.

### Tema 1. Análisis de Conglomerados.

1.1 Introducción

1.2 Concepto de Clasificación Numérica y Conglomerados

1.3 Medidas de Distancia

Euclidiana

Cuadrado de la Euclidiana

Manhatan

Coseno de Variables

Chevychev

Potencia de Valor Absoluto

Distancias Binarias

1.4 Técnicas de Clasificación Jerárquica

Encadenamiento Simple Vecino más cercano

Encadenamiento Completo Vecino más alejado

Método del Centroide

Método de la Mediana

Método de Ward

1.5 Técnicas No Jerárquicas

Método de K medias

1.6 Ejemplos y Aplicaciones

## **Tema 2. Análisis factorial.**

2.1 Introducción

2.2 Fases de Cálculo

2.3 Pruebas de Factibilidad

2.4 Obtención de Factores

2.5 Puntuaciones Factoriales

2.6 Matriz de Correlación Reproducida

2.7 Ejemplos y aplicaciones

## **Tema 3. Escalas Multidimensionales.**

3.1 Introducción

3.2 Matrices de Distancias y Similaridades

3.3 Modelos de Escalas Multidimensionales

3.4 Valores y Vectores Característicos

3.5 Medidas de Stress

3.6 Representación Gráfica R2

## **Tema 4. Análisis Discriminante**

4.1 Introducción

4.2 El caso de 2 poblaciones normales

4.3 Probabilidades a posteriori

## **Tema 5. Introducción a las Redes Neuronales Artificiales**

5.1 Introducción

5.2 Estructura de una RNA

5.3 Funciones de Activación

5.4 Aprendizaje

5.5 Aplicación de RNA en clasificación

5.6 Aplicación de RNA en Regresión

#### **Evaluación.**

Tareas: 30%

Trabajo y Ejercicios Finales: 70%

#### **Formato de Desarrollo.**

Sesiones de exposición a distancia, por parte del instructor en donde se abordan temas metodológicos y aplicaciones, apoyadas por presentaciones y cálculos de ejemplos con apoyo de hoja de cálculo y software específico.

#### **Material Didáctico para participantes.**

- Archivos PDF con notas del curso y lecturas adicionales
- Archivos de datos y código fuente para ejemplos y ejercicios.