

**Universidad de Puerto Rico
Recinto de Río Piedras
Facultad de Ciencias Naturales
Departamento de Matemáticas**

Título:

Introducción a la Estadística con Computadoras

Codificación:

MATE 3026

Número de horas crédito:

3 créditos

Número de horas contacto:

5 horas (3 horas de conferencia y 2 horas de laboratorio).

Requisitos:

MATE 3024 o MATE 3018 o equivalente.

Descripción del catálogo:

MATE 3026. *Introducción a la Estadística con Computadoras*. Tres créditos. Tres horas de conferencia y dos horas de laboratorio a la semana. Requisito: MATE 3024.

Estadística descriptiva. Conceptos básicos de probabilidad. Variables aleatorias continuas y discretas usuales. Distribuciones muestrales. Intervalos de confianza y pruebas de hipótesis para una y dos poblaciones. Análisis de varianza de una vía. Pruebas de chi-cuadrado. Regresión lineal simple y correlación. Sesiones semanales de resolución de problemas usando la computadora.

Descriptive statistics. Basic concepts of Probability. Common continuous and discrete random variables. Sample distributions. Confidence intervals and hypothesis testing for one and two populations. One-way ANOVA. Chi square tests. Simple linear regression and correlation. Weekly sessions of data analysis using statistical software.

Objetivos del curso:

Al finalizar el curso el estudiante podrá

1. Realizar análisis exploratorios de conjuntos de datos.
2. Usar los axiomas y las propiedades básicas de la Probabilidad para el cálculo de probabilidades de eventos.
3. Calcular probabilidades de eventos asociados con variables aleatorias con distribución Binomial y Normal.

4. Aplicar las técnicas básicas de inferencia estadística, incluyendo regresión lineal simple y análisis de varianza con un criterio de clasificación.
5. Usar un programa estadístico para realizar análisis descriptivos o inferencia.
6. Comunicar los resultados del análisis estadístico de un conjunto de datos.

Bosquejo de contenido

Temas:

1. Análisis Exploratorio de Datos (10 horas)

- a) Introducción al Programa Estadístico R
- b) Gráficos descriptivos para variables categóricas: Diagramas de barra.
- c) Gráficos descriptivos para variables cuantitativas: Diagramas de puntos, diagramas de tallo y hoja, histogramas
- d) Medidas descriptivas numéricas: Medidas de centralidad (media, mediana), medidas de dispersión (rango, rango intercuartil, variancia, desviación estándar). Resumen de cinco números, diagrama de caja.

2. Análisis Exploratorio de Relaciones entre Datos (5 horas)

- a) Descripción de asociación entre variables categóricas: Diagramas de barra lado a lado, diagramas de mosaico.
- b) Descripción de asociación entre variables cuantitativas: Diagramas x vs y , correlación.

3. Producción de Datos (5 horas)

- a) Elementos de Diseño Experimental. Aleatorización
- b) Diseño de muestras. Muestreo aleatorio simple

4. Introducción a la Probabilidad (10 horas)

- a) Aleatoriedad vs Determinismo
- b) Reglas básicas de la probabilidad.
- c) Probabilidad Condicional. Teorema de Bayes.
- d) Modelos probabilísticos.
- e) Variables Aleatorias.
- f) Medias y Varianzas de Variables Aleatorias.
- g) Distribuciones de probabilidad: Distribución Binomial, distribución Normal.
- h) Ley de Grandes Números.

5. Distribuciones muestrales (10 horas)

- a) Distribución de Medias Muestrales. Teorema del Límite Central.
- b) Distribución de proporciones muestrales para muestras grandes

6. Introducción a la Inferencia Estadística (10 horas)

- a) Estimadores. Propiedades básicas
- b) Intervalos de Confianza para medias y proporciones.
- c) Pruebas de hipótesis: Hipótesis nula e hipótesis alternativa, errores de tipo I y II, potencia, p-valor. Precauciones en el uso de las pruebas de significancia estadística

7. Inferencia sobre distribuciones (10 horas)

- a) Inferencia sobre una Media
- b) Comparación de dos Medias
- c) Inferencia sobre una proporción
- d) Comparación de proporciones
- e) Estadística Chi-Cuadrado para Tablas con 2 variables

8. Regresión lineal simple (10 horas)

- a) Estimación de los parámetros.
- b) Pruebas de hipótesis sobre los parámetros: prueba t .
- c) Comprobación de suposiciones: análisis de residuos.
- d) Intervalos de confianza para una predicción.

9. Análisis de Varianza con un criterio de clasificación (5 horas)

- a) El modelo y Tabla Anova. Prueba F para la igualdad de las medias de los grupos.
- b) Análisis de residuos
- c) Comparación de grupos. Problema de pruebas de hipótesis múltiples, ajuste de p-valores: Métodos de Bonferroni y de Holm (Bonferroni secuencial)

Estrategias instruccionales:

Discusión, lectura, conferencias, sesiones de laboratorio. En estas últimas, se usará programado estadístico para el análisis de datos, tanto descriptivo como inferencial, y se discutirán los resultados obtenidos.

Recursos de aprendizaje mínimos:

Equipo audiovisual, computador portátil (para uso del instructor), aula equipada con computadoras con servicio de Internet.

Estrategias de evaluación:

La evaluación consistirá en tareas y trabajos de laboratorio (15 %) , tres exámenes parciales (30 %), dos proyectos (35 %), y un examen final (20 %). Los exámenes parciales se realizarán fuera de horas de clase.

Acomodo Razonable:

Los estudiantes que reciban servicios de Rehabilitación Vocacional deben comunicarse con el instructor al inicio del semestre para planificar el acomodo razonable y equipo asistivo necesario conforme a las recomendaciones de la Oficina de Asuntos para las Personas con Impedimentos (OAPI) del Decanato de Estudiantes. También aquellos estudiantes con necesidades especiales que requieren de algún tipo de asistencia o acomodo deben comunicarse con el instructor. (Ley 51 del 7 de junio de 1996).

Integridad académica:

La Universidad de Puerto Rico promueve los más altos estándares de integridad académica y científica. El Artículo 6.2 del Reglamento General de Estudiantes de la UPR (Certificación Nm. 13, 2009-2010, de la Junta de Síndicos) establece que la deshonestidad académica incluye, pero no se limita a: acciones fraudulentas, la obtención de notas o grados académicos valiéndose de falsas o fraudulentas simulaciones, copiar total o parcialmente la labor académica de otra persona, plagiar total o parcialmente el trabajo de otra persona, copiar total o parcialmente las respuestas de otra persona a las preguntas de un examen, haciendo o consiguiendo que otro tome en su nombre cualquier prueba o examen oral o escrito, así como la ayuda o facilitación para que otra persona incurra en la referida conducta. Cualquiera de estas acciones estará sujeta a sanciones disciplinarias en conformidad con el procedimiento disciplinario establecido en el Reglamento General de Estudiantes de la UPR vigente.

Sistema de calificación:

A, B, C, D y F.

Bibliografía:

- Libro de texto:

1. Agresti, A. y Franklin, C. (2013) *Statistics: The Art and Science of Learning from Data* (3a. Edición). Pearson, ISBN-13: 9780321755940.

- Otras referencias:

1. Daalgard, P. (2008). *Introductory Statistics with R*. (2nd ed). Springer.
2. Irizarry, R and Love, M. (2015). *Data Analysis for the Life Sciences*. Leanpub.

- Referencias Electrónicas

1. **R**: Disponible en <http://cran.r-project.org> (Visitado 18/04/2016)
Referencia: R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

2. RStudio Team (2012). R Studio: Integrated Development for R. RStudio, Inc., Boston, MA. Software disponible en <http://www.rstudio.com/> (Visitado 18/04/2016).
3. El texto de Irizarry y Love (2015) se encuentra disponible en <https://leanpub.com/dataanalysisforthelifesciences>. Datos y material adicional pueden encontrarse en <http://genomicsclass.github.io/book/>