



PROGRAMA DE POSGRADO EN ESTADÍSTICA

NP-1601: MÉTODOS, REGRESIÓN Y DISEÑO DE EXPERIMENTOS

PROGRAMA II Semestre 2017

Docente: Deiby Solano Cambronero

Correo electrónico: deiby.solano@ucr.ac.cr

Horario de clases: K: 17:00-19:50
V: 17:00-19:50

Horas de consulta: K: 16:00-16:50
V: 16:00-16:50

Créditos: 5

No tiene requisitos ni correquisitos.

1. Descripción

Curso de **nivelación de la Maestría Académica y Profesional de Estadística**. La primera parte del curso reviste gran importancia para el desempeño del futuro profesional en Estadística, pues le proporciona herramientas básicas de la práctica inferencial. En la segunda parte se estudian los modelos de regresión que son un tema central en la práctica de la estadística y forman la base de un amplio rango de métodos estadísticos. Se presentan las bases de los modelos de regresión múltiple y los procedimientos desarrollados para asegurar una aplicación correcta de los mismos. En la tercera parte se estudian los principios de los diseños experimentales clásicos más importantes.

Se brindará al estudiante la posibilidad de aplicar los métodos mediante el uso del lenguaje de programación (R).

2. Objetivo General

Ofrecer los conceptos básicos de los métodos paramétricos clásicos de la estadística, las técnicas fundamentales de la regresión lineal, y de los diseños experimentales básicos.

3. Objetivos Específicos

Al finalizar el curso el estudiante tendrá criterio y conocimiento básico para:

- Entender y aplicar los conceptos y procedimientos básicos asociados a la estimación y contraste de hipótesis, referidos específicamente al caso de inferencias sobre medias y proporciones.
- Entender y aplicar la técnica del análisis de variancia y los procedimientos de comparaciones múltiples para comparar varios promedios.
- Identificar, calcular, e interpretar algunas de las medidas de asociación más utilizadas.
- Reconocer las situaciones donde se pueden aplicar las técnicas de regresión lineal.
- Utilizar un modelo con fines de predicción (un valor estimado individual, un valor estimado promedio).
- Usar diferentes técnicas de selección de variables para formar un modelo apropiado a partir de un conjunto de predictores disponibles.

- Llevar a cabo los diagnósticos del modelo de regresión ajustado.
- Reconocer las diferentes formas de agrupar las unidades experimentales y de asignarlas a los tratamientos.
- Reconocer los propósitos y alcance de las técnicas y métodos del diseño estadístico.
- Usar adecuadamente el lenguaje de programación R para llevar a cabo los análisis estadísticos estudiados durante el curso.

4. Contenidos

MÉTODOS ESTADÍSTICOS

Módulo I. Estimación.

- 1.1 La distribución muestral y el error estándar de un estimador.
- 1.2 Estimación puntual y por intervalos.
- 1.3 Error estándar e intervalos de confianza para:
 - a. La media y la proporción de una población.
 - b. La diferencia de dos medias o dos proporciones para muestras independientes o pareadas.
- 1.4 Determinación del tamaño de muestra para las estimaciones.

Módulo II: Prueba de hipótesis.

- 2.1 Posibles errores en las pruebas y su impacto.
- 2.2 Prueba de hipótesis para:
 - a. La media y la proporción de una población.
 - b. La diferencia de dos medias o dos proporciones para muestras independientes o pareadas.
- 2.3 Potencia o poder de la prueba.
- 2.4 Determinación del tamaño de muestra para pruebas de hipótesis.
- 2.5 Uso de bootstrap.

Módulo III: Análisis de Varianza.

- 3.1 Análisis de variancia de una vía:
 - a. Fundamentos: suma de cuadrados, cuadrados medios.
 - b. Tabla del ANDEVA.

Módulo IV: Asociación.

- 4.1 Asociación entre variables métricas (correlación de Pearson).
- 4.2 Asociación entre características ordinales (correlación por rangos de Spearman, Gamma).
- 4.3 Asociación entre características nominales (tablas de contingencia, prueba exacta de Fisher para tablas 2x2, Chi cuadrado, coeficiente de contingencia, phi, V de Cramer).

MODELOS DE REGRESIÓN

V. Modelos de regresión lineal.

- 5.1 Ideas básicas.
- 5.2 El modelo y los supuestos.
- 5.3 Estimación de los coeficientes (notación matricial).
- 5.4 Valores ajustados y residuales.
- 5.5 Inferencias. El cuadrado de medio de error. Varianzas de los estimadores. Intervalos de confianza para los coeficientes. Intervalos de confianza para la respuesta media y predicción de nuevas observaciones.
- 5.6 Coeficientes de determinación (múltiple y parcial). Suma de cuadrados de regresión adicional (condicional a predictores incluidos previamente).
- 5.7 Pruebas de hipótesis. Comparación de modelos anidados con la distribución F. Uso de la distribución t.

VI. Selección de predictores.

- 6.1 Coeficiente de determinación (R^2) ajustado.
- 6.2 Selección hacia adelante y hacia atrás.
- 6.3 Criterios de información: Akaike y Bayes.
- 6.4 Estadístico de Mallow (C_p).

VII. Diagnósticos y medidas remediales.

- 7.1 Verificación de los supuestos:
 - a. Normalidad vs no-normalidad: QQ plot.
 - b. Homoscedasticidad vs heteroscedasticidad: gráfico de residuales y prueba formal de Breush Pagan.
 - c. Independencia lineal vs multicolinealidad: factor de inflación de variancia (VIF), análisis de suma de cuadrados de regresión adicional, coeficientes de correlación entre predictores.
 - d. Linealidad vs no-linealidad entre predictor y respuesta: gráficos de regresión parcial.
 - e. Aleatoriedad vs autocorrelación: prueba de Durbin-Watson.
- 7.2 Medidas para detectar valores extremos:
 - a. En la respuesta: residuales estudentizados externamente.
 - b. En los predictores: valores de "leverage".
- 7.3 Medidas para detectar casos de influencia:
 - c. Sobre un valor ajustado: DFFITs.
 - d. Sobre el ajuste general: distancia de Cook.
 - e. Sobre los coeficientes: DFBETAs.
- 7.4 Validación del modelo.
- 7.5 Medidas remediales.

DISEÑOS EXPERIMENTALES

VIII. Diseños con un solo factor.

- 8.1 Introducción al diseño de experimentos:
 - a. Principios y conceptos del diseño experimental: unidad, experimental, factor, tratamiento, aleatorización.
 - b. Consideraciones en la elección de los factores, niveles y rangos.
- 8.2 Verificación de los supuestos del modelo:
 - a. Independencia de los errores.
 - b. Normalidad.
 - c. Homoscedasticidad.
- 8.3 Estimaciones de medias y efectos:
 - a. Contrastes.
 - b. Intervalos simultáneos: Scheffé y Bonferroni.
 - c. Comparaciones múltiples entre medias: pruebas de Duncan, Tukey, etc.

IX. Diseños factoriales.

- 9.1 El modelo de dos factores sin interacciones.
- 9.2 El modelo de dos factores con interacciones.
- 9.3 El análisis de varianza en el diseño factorial.
- 9.4 Efecto sobre la variabilidad del error al agregar o eliminar factores.
- 9.5 Estimación de parámetros.
- 9.6 Temas relacionados con el tamaño de muestra en un diseño factorial.
- 9.7 El diseño factorial general.

X: Diseños de bloques aleatorizados.

- 10.1 El caso de pruebas pareadas como un caso sencillo de bloques.
- 10.2 El concepto de bloques.
- 10.3 El análisis de varianza en el diseño de bloques.
- 10.4 Prueba de la relevancia del uso de los bloques – eficiencia relativa.
- 10.5 El diseño de medidas repetidas como un enfoque de diseño de bloques.
- 10.6 Combinación de bloques con un diseño factorial.
- 10.7 Bloques anidados.

5. Metodología

- Presentaciones teóricas: se impartirán lecciones magistrales por parte del docente donde se explicarán los conceptos y sus aplicaciones.
- Prácticas: se usará el laboratorio de computadoras de la escuela de estadística para el desarrollo de lecciones prácticas durante el semestre. Durante las sesiones de laboratorio se utilizará el lenguaje de programación R para realizar ejercicios de la materia vista en clase. Se realizarán ejercicios y sus respectivas prácticas.
- Trabajo con datos reales:
 - Los estudiantes realizarán un trabajo de análisis de datos reales utilizando lo aprendido en el curso.
 - Las aplicaciones podrán ser tomadas de investigaciones en áreas como la economía, medicina, psicología, biología, población, etc.

- Los estudiantes presentarán un anteproyecto donde expondrán los objetivos del estudio, el cual será discutido con el profesor y devuelto para que se realicen las mejoras sugeridas.
 - Al final de curso los estudiantes deberán escribir un documento con formato de artículo científico con la investigación realizada.
 - Los resultados deberán ser expuestos en forma oral en una presentación con tiempo limitado.
- Se usará como texto el libro ***Applied Linear Statistical Models*** de ***Neter et al.***
 - La asistencia a las lecciones será deseable, además, en las mismas se brindará materia para complementar los textos y las diapositivas que también puede ser evaluada en exámenes o tareas.

6. Cronograma de la materia

Se indica entre paréntesis el número estimado de semanas para cada capítulo:
I(1), II(1), III(1), IV(1), V(2), VI(1), VII(2.5), VIII(1), IX(1.5), X(1.5)

7. Evaluación

Se realizarán tres exámenes parciales, además, los estudiantes presentarán un trabajo de análisis de datos y varias tareas cortas.

Para la evaluación del trabajo de análisis de datos se proporcionará un desglose de la distribución de criterios, en los cuales se tomará en cuenta la revisión bibliográfica, la redacción, el planteamiento, y la profundidad en el análisis estadístico.

Si un estudiante faltase a algún examen o a la entrega de alguna tarea por **causa justificada**, debe solicitar por escrito la reposición indicando las razones de la ausencia, acompañada de los documentos justificantes. La misma debe entregarse ante el profesor **a más tardar en cinco días hábiles del reintegro a lecciones**.

En el siguiente cuadro se presenta el desglose porcentual de la nota:

Primer examen parcial	20%
Segundo examen parcial	30%
Tercer examen parcial	20%
Trabajo de análisis de datos	20%
Tareas	10%
Total	100%

8. Cronograma de exámenes

Examen	Temas	Fecha tentativa
Primer parcial	I, II, III y IV	Semana 5 (8 de setiembre)
Segundo parcial	V, VI y VII	Semana 12 (27 de octubre)
Tercer parcial	VIII, IX y X	Semana 17 (1 de diciembre)

9. Referencias bibliográficas

Box, G.E.P., W.G.Hunter y J.S.Hunter. 1978. *Statistics for experimenters*. John Wiley & Sons. New York.

Davies (editor). 1977. *The Design and Analysis of Industrial Experiments*. Longman Group Ltd., London.

Faraway, J. (2004). *Linear Models with R*. Chapman & Hall/CRC.

Faraway, J. (2005). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC.

Gutiérrez, Edgar. *Métodos estadísticos para las Ciencias Biológicas*. Editorial EUNA. Año 2000.

Montgomery, D.C. 2005. *Diseño y análisis de experimentos*. México, D.F. Editorial LIMUSA, S.A. de C.V. 2ª. Ed.

Fox, J. (2002). *An R ans S-Plus Companion to Applied Regression*. Thousand Oaks, Cal.: Sage Publications, Inc.

Gujarati, D. (1990). *Econometría básica*. McGraw-Hill.

Neter, J et al. (1996). *Applied Linear Statistical Models*. 5a. ed. WCB, McGraw-Hill.