



Universidad de Costa Rica
Sistema de Estudios de Posgrado
Programa de Posgrado en Estadística

NP-1601 Métodos, regresión y diseño de experimentos

INFORMACIÓN GENERAL

| | |
|--|--|
| Plan de estudios al que pertenece el curso: | 730501 Maestría Académica en Estadística y 730506 Maestría Profesional en Estadística |
| Tipo de curso: | Nivelatorio para la Maestría Académica en Estadística y Maestría Profesional en Estadística |
| Modalidad: | Teórico- Bimodal Se utilizará la plataforma de mediación virtual |
| Número de créditos: | 5 créditos |
| Horas presenciales: | 6 horas semanales |
| Horario: | <i>Martes: 18:00-20:50 y Viernes: 18:00-20:50</i> |
| Aula: | Martes: CE-300A y Viernes: <i>Virtual, plataforma ZOOM</i> |
| Horas de atención de estudiantes: | 2 horas semanales (<i>Martes: 17:00-17:50 CE-300A</i> y <i>V: 17:00-17:50 Virtual</i>) |
| Requisitos: | No tiene |
| Correquisitos: | No tiene |
| Ciclo lectivo: | <i>II-2023</i> |
| Profesor: | Deiby Solano Cambroner. <i>Máster.</i> |



PROGRAMA

Justificación

Curso de nivelación para estudiantes de la Maestría Académica o Profesional de Estadística que provienen de disciplinas distintas a la Estadística o bien, que nunca hayan aprobado un curso similar previamente. Se pretende con este ofrecer el conocimiento mínimo necesario para nivelarse con los estudiantes que provienen de la carrera de Estadística. Con este curso, y el resto de cursos de nivelación, los estudiantes meta se integrarán exitosamente los cursos regulares del programa. La primera parte del curso le proporciona al estudiante herramientas básicas de la práctica inferencial. En la segunda parte se estudian los modelos de regresión, que son la base de un amplio rango de métodos estadísticos y 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).

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.

Objetivos específicos

Al final del curso se espera que el estudiante haya aprendido a:

1. Entender y aplicar los conceptos y procedimientos básicos asociados a la estimación y contraste de hipótesis, referidos específicamente al caso de inferencias sobre medias.
2. Identificar, calcular, e interpretar algunas de las medidas de asociación más utilizadas.
3. Reconocer las situaciones donde se pueden aplicar las técnicas de regresión lineal.
4. Utilizar un modelo con fines de predicción (un valor estimado individual, un valor estimado promedio).
5. Usar diferentes técnicas de selección de variables para formar un modelo apropiado a partir de un conjunto de predictores disponibles.
6. Llevar a cabo los diagnósticos del modelo de regresión ajustado.
7. Reconocer las diferentes formas de agrupar las unidades experimentales y de asignarlas a los tratamientos.
8. Entender y aplicar la técnica del análisis de variancia y los procedimientos de comparaciones múltiples para comparar varios promedios.
9. Reconocer los propósitos y alcance de las técnicas y métodos del diseño estadístico.
10. Usar adecuadamente el lenguaje de programación R para llevar a cabo los análisis estadísticos estudiados durante el curso.



Contenido del curso

MÉTODOS ESTADÍSTICOS

- I. Estimación
 - a. La distribución muestral y el error estándar de un estimador
 - b. Estimación puntual y por intervalos
 - c. Error estándar e intervalos de confianza para la media de una población y la diferencia de dos medias para muestras independientes o pareadas.
- II. Prueba de hipótesis
 - a. Posibles errores en las pruebas y su impacto
 - b. Prueba de hipótesis para la media de una población y la diferencia de dos medias para muestras independientes o pareadas
 - c. Potencia o poder de la prueba
- III. Asociación
 - a. Asociación entre variables métricas (correlación de Pearson)
 - b. Asociación entre características ordinales (correlación por rangos de Spearman, Gamma)
 - c. 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

- IV. Modelos de regresión lineal
 - a. Ideas básicas
 - b. El modelo y los supuestos
 - c. Estimación de los coeficientes (notación matricial)
 - d. Valores ajustados y residuales
 - e. 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
 - f. Coeficientes de determinación (múltiple y parcial). Suma de cuadrados de regresión adicional (condicional a predictores incluidos previamente)
 - g. Pruebas de hipótesis. Comparación de modelos anidados con la distribución F. Uso de la distribución t.
- V. Selección de predictores
 - a. Coeficiente de determinación (R^2) ajustado
 - b. Selección hacia adelante y hacia atrás
 - c. Criterios de información: Akaike y Bayes
 - d. Estadístico de Mallows (C_p).
- VI. Diagnósticos y medidas remediales
 - a. Verificación de los supuestos: normalidad vs no-normalidad (QQ plot), homoscedasticidad vs heteroscedasticidad (gráfico de residuales y prueba formal de



- Breush Pagan), independencia lineal vs multicolinealidad (factor de inflación de variancia (VIF), análisis de suma de cuadrados de regresión adicional, coeficientes de correlación entre predictores), linealidad vs no-linealidad entre predictor y respuesta (gráficos de regresión parcial), aleatoriedad vs autocorrelación (prueba de Durbin-Watson)
- Medidas para detectar valores extremos en la respuesta (residuales estudentizados externamente) y en los predictores (valores de “leverage”)
 - Medidas para detectar casos de influencia: sobre un valor ajustado (DFFITs), sobre el ajuste general (distancia de Cook) y sobre los coeficientes (DFBETAs)
 - Validación del modelo
 - Medidas remediales: transformaciones, Box-Cox, mínimos cuadrados ponderados.

DISEÑOS EXPERIMENTALES

VII. Diseños con un solo factor

- Introducción al diseño de experimentos: Principios y conceptos del diseño experimental (unidad, experimental, factor, tratamiento, aleatorización) y consideraciones en la elección de los factores, niveles y rangos
- Análisis de variancia de una vía y sus fundamentos: suma de cuadrados, cuadrados medios.
- Tabla del ANDEVA
- Verificación de los supuestos del modelo: independencia de los errores, normalidad, homoscedasticidad
- Estimaciones de medias y efectos: contrastes, Intervalos simultáneos (Scheffé y Bonferroni) y comparaciones múltiples entre medias (pruebas de Duncan, Tukey, etc).

VIII. Diseños factoriales

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

IX. Diseños de bloques aleatorizados

- El concepto de bloques
- El análisis de varianza en el diseño de bloques
- Combinación de bloques con un diseño factorial

X. Modelo general

- Predictores nominales y predictores continuos. Necesidad de inclusión de interacciones factor nominal y variable explicativa continua
- El análisis de covarianza. Introducción de covariables para controlar el error. Diferencia entre una covariable y un factor continuo a evaluar.



Metodología

El curso es teórico-práctico y donde se realizarán presentaciones teóricas, se impartirán lecciones magistrales por parte del docente donde se explicarán los conceptos y sus aplicaciones. Además, se realizarán laboratorios en computadora que se desarrollarán mediante el programa R durante el semestre. Se resolverán los laboratorios y se brindarán algunas tareas y/o quices evaluados como práctica.

Los estudiantes realizarán un trabajo de análisis de datos reales del tema regresión, donde las aplicaciones podrán ser tomadas de investigaciones en áreas como la economía, medicina, psicología, biología, población, etc. Los estudiantes presentarán un anteproyecto con los objetivos del estudio, el cual será discutido con el profesor y devuelto para que se realicen las mejoras sugeridas. Los estudiantes deberán escribir un documento con formato de artículo científico con la investigación realizada en la fecha brindada por el profesor. 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 será evaluada en exámenes o tareas.

Se utilizará la plataforma de Mediación Virtual para colocar documentos, presentaciones, presentación de tareas y exámenes. Así mismo, cuando se requiera, se utilizará también el correo electrónico. Así mismo las clases virtuales serán brindadas por la plataforma Zoom, para lo cual se enviarán con anterioridad el acceso correspondiente y las físicas en el aula asignada.

Evaluación

Se realizarán tres exámenes parciales, en ellos se evaluarán los conceptos, el uso del software para obtener resultados y la forma de interpretar resultados. Además, los estudiantes presentarán un trabajo de análisis de datos y varias tareas cortas y/o quices.

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. Entre los entornos a utilizar en el curso, el mismo será bimodal: con mezcla de utilización de la plataforma mediación virtual, el correo electrónico y la plataforma Zoom. Para las evaluaciones se utilizará una mezcla de las plataformas mencionadas anteriormente.



La evaluación del curso se desglosa según los siguientes rubros:

| | |
|------------------------|-------------|
| Primer examen parcial | 15% |
| Segundo examen parcial | 30% |
| Tercer examen parcial | 25% |
| Artículo sobre técnica | 25% |
| Tareas | 5 % |
| Total | 100% |

Cronograma

Se indica entre paréntesis el número estimado de semanas para cada capítulo:

I(1), II(1), III(1), IV(2), V(1), VI(2), VII(1), VIII(1), IX(1), X(1)

Bibliografía

La siguiente bibliografía es la recomendada para el curso:

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.

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.

Gutiérrez, Edgar (2000). Métodos estadísticos para las Ciencias Biológicas. Editorial EUNA

Matloff, Norman (2011). The art of R programming

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

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

Santana, Julio y Farfan, Efraín (2014). El arte de programar en R, un lenguaje para la estadística. México : Instituto Mexicanode Tecnología del Agua. UNESCO. Comité Nacional Mexicano del Programa Hidrológico Internacional, ©2014.