

Carta al estudiante

Curso: Estadística Actuarial II (CA-403).

Requisito: Estadística Actuarial I (CA-303).

Número de créditos: 5.

Horas de clase: 5 horas semanales.

Horas de estudio extraclase: 10 horas semanales.

Horario de clase: Martes: 11h–13h (217 FM) / Viernes: 10h–13h (404 FM).

Horario de consulta: Martes: 14h–16h y Jueves: 13h–14h

Profesor: Maikol Solís.

Oficina: Ciudad de la investigación, Edificio de Matemática, Segundo piso, Oficina # 5.

Casillero: # 86.

Correo electrónico: maikol.solis@ucr.ac.cr

Teléfono: (+506) 25 11 66 07.

1. Descripción del curso

Este curso es una continuación de los temas de Estadística Actuarial I (CA-303) y pretende proveer una introducción a temas avanzados de la estadística. En este curso veremos problemas relacionados con la estadística asintótica, la estimación no paramétrica y el aprendizaje de máquinas (machine learning).

En general, se pretende que el estudiante amplíe su horizonte de conocimiento hacia otras ramas de la estadística que son frecuentemente utilizadas en la aplicación de modelos actuariales modernos.

2. Objetivos

2.1. General

Estudiar las principales técnicas de la estadística asintótica, no paramétrica y el aprendizaje estadística a través de:

1. La generalización de la teoría inferencial aprendida en Estadística Actuarial I.
2. La aplicación del lenguaje estadístico R en problemas reales y simulados.

2.2. Específicos

1. Analizar el comportamiento de las herramientas de inferencia estadística aprendidas en CA303 cuando el tamaño de muestra va a infinito, en particular los estimadores MLE y pruebas de cociente de verosimilitud.

2. Introducir algunos conceptos básicos de la estadística no paramétrica, con especial énfasis en la técnica Bootstrap y la estimación de funciones de densidad utilizando kernels.
3. Introducir técnicas básicas de Aprendizaje Estadístico supervisado y no-supervisado, junto con algunas herramientas de selección de variables y regularización.

3. Contenidos

Primer parcial (Semana 1 a la 5):

1. Repaso de convergencia de variables aleatorias.
2. Teoremas de Slutsky. Método Delta. Distribución del coeficiente de correlación muestral.
3. Prueba Pearson de χ^2 . Normalidad asintótica del MLE.
4. Normalidad asintótica de distribuciones posteriores.
5. Distribución asintótica del estadístico LRT. Pruebas de hipótesis asintóticas.

Segundo parcial (Semana 6 a la 10):

1. Fundamentos de Estadística no-paramétrica: Estimadores de densidad por kernels.
2. Estimadores de densidad por kernels y estimadores por validación cruzada.
3. Bootstrap y Jackknife.

Tercer parcial (Semana 11 a la 16):

1. Aprendizaje supervisado cuantitativo. Regresión lineal simple y múltiple. Estimación de parámetros y prueba F.
2. Diagnóstico de supuestos y extensiones.
3. Clasificación binaria: Regresión logística simple y múltiple.
4. Regresión logística. Curvas ROC. Clasificación a través de análisis discriminante lineal.
5. Métodos de selección de variables y regularización.
6. Clasificación a través de máquinas de soporte vectorial. Clasificador de margen maximal.
7. Máquinas de soporte vectorial.
8. Introducción al aprendizaje no-supervisado: Análisis en componentes principales.

4. Metodología

Clases magistrales en donde se combinará la justificación teórica de los conceptos teóricos junto con la participación de los estudiantes en los ejercicios desarrollados en clase.

5. Evaluación

La nota final será calculada ponderando los siguientes rubros:

Promedio final	
3 exámenes parciales	25 % cada uno
Proyecto final	20 %
Presentación de noticias	5 %
Total	100 %

5.1. Presentación de noticias

En la clase corta de 2 horas, se dará un espacio que el estudiante presente una alguna noticia o documento científico vinculado con los temas del curso.

- La presentación de cada noticia debe ser corta y concisa. Se dispone solo de **un máximo de 5 minutos por estudiante**.
- La noticia o documento pueden periódicos convencionales o electrónicos, blogs, artículos científicos, entre otros.

Cada vez que el estudiante presente su noticia ante la clase, se le marcará como completa la sesión de ese día. Al estudiante se le asignará el puntaje correspondiente según la cantidad de sesiones completadas entre la cantidad de sesiones totales.

Las sesiones de noticias no podrán ser repuestas en semanas posteriores a menos exista una justificación reglamentada en el artículo 24 del régimen académico estudiantil.

5.2. Proyecto

El proyecto de investigación será un trabajo teórico - práctico será presentado en la fecha correspondiente según el calendario. Este proyecto consistirá en la exposición y aplicación de un tema relacionado con los contenidos del curso. En general, el estudiante deberá de estudiar el tema, recolectar datos relevantes y desarrollar una serie de análisis según lo requerido.

El tema para el proyecto podrá ser propuesto por los mismos estudiantes una vez que haya sido aprobado por mi persona. En caso de que el estudiante no proponga un tema, entonces se le asignará alguno en el transcurso del semestre.

El estudiante deberá generar un reporte final donde se expongan todo el material investigado, además de su aplicación correspondiente.

Adicionalmente, el estudiante deberá de presentar este reporte en una exposición oral. Para ese día, se invitarán a otros profesores y expertos del área, de modo que sus trabajos puedan ser calificados por terceros. Al finalizar la exposición, habrá una sección de preguntas por parte de los profesores, los invitados y los otros estudiantes.

Atención: El reporte y la presentación para el proyecto deberán ser generados en \LaTeX y el lenguaje de programación oficial del curso es R. **¡Sin excepción!** Se le exigirá al estudiante el código fuente de todos su trabajo (archivos tex y R) de estos documentos. Cualquier trabajo presentado en otro procesador de texto (Microsoft Office, Open/Libre Office, etc.) o lenguaje de programación (Matlab, Mathematica, Java, etc.) será rechazado y al estudiante se le asignará nota 0 automáticamente

Todos los programas que se usarán en el curso son de acceso libre y gratuito. Además, estos programas están disponibles para las plataformas Windows, Mac OS X y Linux.

- \LaTeX : <http://www.latex-project.org/>
- TeXStudio (editor): <http://texstudio.sourceforge.net/>
- R: <https://cran.rstudio.com/>
- RStudio: <https://www.rstudio.com/>

Los laboratorios de cómputo de la Escuela de Matemática cuentan con R y \LaTeX instalado. Por favor, consulte sobre los horarios de apertura para estudiantes en la secretaría.

6. Calendario de evaluaciones

El aula para cada evaluación será informado posteriormente.

Fecha límite para presentar propuestas para el trabajo de investigación	Segunda semana de Mayo
1 ^{er} Parcial	3 de Mayo, 3 p.m.
2 ^{do} Parcial	7 de Junio, 3 p.m.
3 ^{er} Parcial	12 de Julio, 3 p.m.
Presentación del Proyecto	10 de Julio, (hora por definir)
Entrega de promedios	14 de Julio (hora por definir)
Examen de ampliación	20 de Julio (hora por definir)

7. Reglamento régimen académico estudiantil

Para conocer sus deberes y derechos como estudiante en este curso, les aconsejo revisar el reglamento de régimen académico estudiantil, el cual pueden encontrar en la dirección http://www.cu.ucr.ac.cr/normativ/regimen_academico_estudiantil.pdf.

Referencias

- Bickel, P. J. & Doksum, K. A. (1977). *Mathematical statistics: basic ideas and selected topics*. San Francisco: Holden-Day.
- Casella, G. & Berger, R. L. (2002). *Statistical inference*. Pacific Grove, CA: Duxbury.
- DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer Science & Business Media.
- Ferguson, T. S. (1996). *A course in large sample theory*. Chapman & Hall London.
- Friedman, J., Hastie, T. & Tibshirani, R. (2001). *The elements of statistical learning*. Springer series in statistics Springer, Berlin.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2014). *An introduction to statistical learning: with applications in r*. Springer Publishing Company, Incorporated.
- Robert, C. P. (2007). *The bayesian choice: from decision-theoretic foundations to computational implementation*. OCLC: 187082900. New York: Springer.
- van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- Wasserman, L. (2004). *All of statistics: a concise course in statistical inference*. New York: Springer.
- Wasserman, L. (2010). *All of nonparametric statistics* (1st). Springer Publishing Company, Incorporated.
- Young, G. A. & Smith, R. L. (2005). *Essentials of statistical inference: g.a. young, r.l. smith*. OCLC: 61410200. Cambridge, UK; New York: Cambridge University Press.

Cordialmente,
Maikol Solís