



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

SP-1629 Análisis de Datos Categóricos

INFORMACIÓN GENERAL

Plan de estudios al que pertenece el curso:	730501 Maestría Académica en Estadística y 730506 Maestría Profesional en Estadística
Tipo de curso:	Optativo para la Maestría Académica en Estadística y Obligatorio Maestría Profesional en Estadística
Modalidad:	Teórico
Número de créditos:	4 créditos
Horas presenciales:	4 horas semanales
Horas de atención de estudiantes:	2 horas semanales (<i>K</i> , de 7pm a 9pm)
Requisitos:	No tiene
Correquisitos:	No tiene
Profesor:	Tatiana Barboza Solórzano

PROGRAMA

Justificación

Las variables categóricas son muy comunes en todas las áreas de aplicación. Su análisis va más allá de modelos de regresión logística, por lo que este curso vendrá a complementar el material ya visto en ese primer semestre del programa.



Objetivo general

Capacitar al estudiante a manejar los métodos más importantes de análisis de datos categóricos en tablas de contingencia basados en las estadísticas de ji cuadrado y el cociente 'odds'.

Objetivos específicos

1. Proporcionar al estudiante las técnicas básicas de análisis categóricos para tablas de contingencia utilizando modelos de regresión Poisson, escogencia discreta y modelos loglineales.
2. Permitir al estudiante familiarizarse con el paquete STATA para analizar los datos categóricos e interpretar los resultados.

Descripción del curso

- I. Introducción a STATA.
- II. Introducción a grafos y redes sociales desde el punto de vista descriptivo.
- III. Tablas de contingencia de dos vías. Contrastes de independencia, contrastes exactos, diferencia en proporciones, cociente 'odds', riesgo relativo, sensibilidad y especificidad, contraste de McNemar, conjuntos de tablas 2×2 , el contraste de Mantel-Haenszel, medidas de asociación, conjuntos de tablas $2 \times r$ y $s \times 2$, la tabla $s \times r$: asociación, contrastes exactos de asociación, medidas de asociación, conjuntos de tablas $s \times r$: metodología de Mantel-Haenszel, aplicaciones.
- IV. Introducción a los modelos lineales generalizados. Formulación matemática, ejemplos de modelos lineales generalizados: regresión clásica, logística, regresión Poisson. Pruebas exactas.
- V. Análisis de escogencia discreta. Logito, logito condicional, logística multinomial, logística exacta, probit, logístico mixto; ejemplos, datos ordenados, aplicaciones a economía y mercadeo.
- VI. Modelos de regresión logística de efectos fijos, aleatorios y de promedio poblacional, análisis de variables respuestas binarias con medidas repetidas (también conocidas como diseños longitudinales).
- VII. Modelos log-lineales básicos. para 2×2 tablas, en más dimensiones, residuos para modelos log-lineales, contraste de significancia, la función verosimilitud.

Metodología

Clases magistrales y sesiones de laboratorio con la participación de los estudiantes. Al final del curso los estudiantes expondrán un trabajo donde aplicarán algunos de los métodos aprendidos en el curso a datos reales de nuestro medio. Se hará un esfuerzo para que parte de los informes de tareas (o proyectos pequeños) se entregan en formato de reporte o informe laboral, en lugar de artículo científico. Aún así, el trabajo final sí tiene que tener formato de artículo científico, o al menos, de monografía. El trabajo final se puede hacer en grupos de máximo 2 personas.



Este curso trata del análisis de datos cuando las variables respuesta son categóricas: cualitativas o cuantitativas con un número limitado de valores. Las variables predictoras o ‘explicativas’ pueden ser categóricas o continuas. Se estudian dos enfoques. El primero, llamado ‘clásico’, utiliza varias medidas de asociación basadas en las estadísticas de ji cuadrado y los cocientes ‘odds’. El segundo, desarrollado en los años sesentas, utiliza métodos análogos al análisis de variancia y a los modelos de regresión para variables respuestas continuas. Algunos de los modelos de este enfoque son la regresión logística, la regresión Poisson y los modelos loglineales.

Cronograma

I(1), II(2), III(2), IV(3), V(3), VI(3), VII(2), exámenes (1c/u). El número de semanas de cada tema (entre paréntesis) es un valor estimado.

Bibliografía

- Agresti, A. (2004). *Categorical Data Analysis*. Segunda edición. John Wiley, N.Y.
 Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. Segunda edición. John Wiley, N.Y.
 Hosmer, D.W. y Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley.
 Lachin JM. (2000). *Biostatistical Methods. The Assessment of Relative Risks*. New York: John Wiley and Sons, Inc.
 Allison, Paul.D. (1999). *Logistic Regression Using the SAS System*. Cary, NC:SAS Institute, Inc.
 Manuales de SPSS, SAS y STATA.
 Stokes, M.E, Davis, C.S y Koch, G.G. (2009). *Categorical Data Analysis using the SAS System*. Cary, NC:SAS Institute, Inc.
 Train, K.E. (2003). *Discrete Choice Methods with Simulation*. Cambridge Univ. Press.
 Zelterman, Daniel (2002). *Advanced Log-Linear Modes Using SAS*. Cary, NC:SAS Institute, Inc.

Evaluación

TOTAL	100%
I Examen Parcial (Temas II , III y IV)	25%
II Examen Parcial (Temas V, VI y VII)	25%
Tareas y quizzes (El valor de cada evaluación dependerá de la cantidad de tareas y quizzes que se asignen)	25%
Trabajo final: Se divide en 5% entrega de revisión, 15% artículo final y 5% preguntas al final de exposición	25%