



**Secretaría de Estudios Avanzados
Subsecretaría de Doctorado**

Programa de Seminario de Doctorado

Año:

Nombre del seminario: Análisis de Datos Categóricos

Nombre del Profesor: Patricio Solís

Área temática sugerida: Metodología

1. Fundamentación

En la investigación social frecuentemente intentamos explicar el comportamiento de variables de tipo nominal o categórico. Para este tipo de problemas se han desarrollado un conjunto de técnicas estadísticas que se suelen agrupar bajo la etiqueta de “Análisis de datos categóricos”. Esto incluye técnicas descriptivas o exploratorias como el análisis de contingencia, regresiones para variables dependientes categóricas, como la regresión logística en sus versiones binomial y multinomial, y regresiones para modelar el patrón de asociación entre dos variables categóricas, como la regresión log-lineal.

El propósito de este curso es cubrir las técnicas de análisis de datos categóricos más utilizadas en la sociología. El curso tiene una modalidad aplicada, por lo que se hará énfasis en las posibilidades y limitaciones que tiene cada técnica para analizar problemas concretos de investigación. La dinámica del curso consistirá en la exposición en clase de los fundamentos teóricos y prácticos de cada técnica, seguida por ejercicios prácticos, en los que los estudiantes podrán conocer la aplicación de las técnicas mediante el uso de software especializado (Stata).

2. Objetivos

Al finalizar el curso los alumnos serán capaces de interpretar y utilizar los modelos de regresión logística binomial, logística multinomial, y loglineal, en situaciones reales de investigación, así como identificar y resolver los problemas técnicos más comunes asociados a la aplicación de estas técnicas.

3. Contenidos (divididos en unidades temáticas)

- Análisis de contingencia
- El modelo de regresión lineal, el modelo lineal de probabilidad y sus limitaciones
- Regresión logística binomial
- Regresión logística multinomial
- Regresión log-lineal

4. Metodología de trabajo

Exposición en clase y ejercicios prácticos

5. Cronograma de clases

Sesión	Contenido
1	Introducción al curso. Revisión del análisis de contingencia en tablas de doble entrada. Pruebas de asociación en el análisis de contingencia. Riesgos. Riesgos relativos. Momios. Razones de momios.
2	Revisión de regresión lineal. El modelo lineal de probabilidad y sus limitaciones. El modelo lineal generalizado. Introducción al modelo de regresión logística binomial.
3	Interpretación de coeficientes y razones de momios en el modelo de regresión logística binomial. Probabilidades estimadas
4	Estimación por máxima verosimilitud del modelo de regresión logística binomial. Medidas de bondad de ajuste. Pseudo R cuadrada de McFadden. Prueba de Hosmer y Lemeshow. BIC.
5	Ejercicio en clase: ajuste de modelos de regresión logística binomial. Interpretación de coeficientes. Evaluaciones de bondad de ajuste
6	Regresión logística multinomial. Introducción. Presentación formal del modelo. Selección de categoría base de contraste. Supuesto de Independencia de Alternativas Irrelevantes (IIA)
7	Interpretación de coeficientes, probabilidades ajustadas y bondad de ajuste en el modelo de regresión logística multinomial
8	Ejercicio en clase: ajuste de modelos de regresión logística multinomial. Interpretación de coeficientes. Evaluaciones de bondad de ajuste
9	Regresión loglineal. Introducción. Presentación formal del modelo. Efectos marginales y efectos de interacción. Ajuste en tabla de 2 x 2
10	Regresión loglineal. Modelo nulo, modelo saturado, distintos patrones de asociación (ejemplos con tablas de movilidad social). Medidas de bondad de ajuste: Devianza, índice de disimilitud, BIC

6.

6. Evaluación

La evaluación del curso se realizará mediante la evaluación de dos ejercicios en clase (40%) y un trabajo final (60%), en el que los alumnos deberán aplicar alguna de las técnicas aprendidas durante el curso.

7. Bibliografía

Libros de referencia básicos:

Agresti, Alan. 2007. An introduction to Categorical Data Analysis. (2ª Edición). New Jersey: John Wiley & Sons.

Long, Scott y Freese, Jeremy. 2006. Regression Models for Categorical Dependent Variables Using Stata. College Station: Stata Press

Otra bibliografía de consulta:

Hosmer, David H. y Stanley Lemeshow. 2000. Applied Logistic Regression. (2ª Edición). New Jersey: John Wiley & Sons

Powers, Daniel A. & Yu Xie 2000. Statistical methods for categorical data analysis. New York: Academic Press

Treiman, D. J. (2014). *Quantitative data analysis: Doing social research to test ideas*. John Wiley & Sons.

Ejemplos de artículos que utilizan la regresión logística binomial:

Aparicio, Ricardo y David H. Corrochano. 2005. El perfil del votante clientelar en México durante las elecciones de 2000. *Estudios Sociológicos* Vol. 23, No. 68 (May - Aug., 2005), pp. 375-396

Cerrutti, Marcela y Rene Zenteno, 2000 “Cambios en el Papel Económico de las Mujeres entre las Parejas Mexicanas”. *Estudios Demográficos y Urbanos*, 15/1

CORTES, F. Y R. RUBALCAVA. (1993), "Algunas determinantes de la inserción laboral en la industria maquiladora de exportación de Matamoros", en: *Estudios Sociológicos*, Vol. XI, No 31, enero - abril, México, Ed. El Colegio de México.

García, Brígida y Edith Pacheco (2000), “Esposas, hijos e hijas en el mercado de trabajo de la ciudad de México en 1995”, *Estudios demográficos y Urbanos*, El Colegio de México, vol. 15, no. 1, January-April, pp. 35-64

Ejemplos de artículos utilizan la regresión logística multinomial:

López Ramírez, Adriana. 2008. "Migración, remesas y arreglos residenciales de los adultos mayores en México". *Estudios Demográficos y Urbanos*, Vol. 23, Núm. 3, septiembre-diciembre, pp. 513-541

Andrés Peri e Ignacio Pardo. 2006. "Nueva evidencia sobre la hipótesis de la doble insatisfacción: ¿Cuán lejos estamos de que toda la fecundidad sea deseada?", en III Congreso de la Asociación Latinoamericana de Demografía, Guadalajara (versión en CD-ROM).

Ejemplos de artículos que utilizan la regresión log-lineal:

Erikson, Robert and John H. Goldthorpe. 1987. "Commonality and variation in social fluidity in industrial nations. Part I: A model for evaluating the 'FJH hypothesis'." *European Sociological Review* 3:54-77.

Erikson, Robert and John H. Goldthorpe. 1987. "Commonality and variation in social fluidity in industrial nations. Part II: The model of core social fluidity applied." *European Sociological Review* 3:145-166.

Solís, Patricio y Fernando Cortés. 2009. "La movilidad ocupacional en México: rasgos generales, matices regionales y diferencias por sexo" pp. 395-433 en: Cecilia Rabell Romero (Ed.): *Tramas familiares en el México contemporáneo. Una perspectiva sociodemográfica*. México, D.F.: Instituto de Investigaciones Sociales, UNAM - El Colegio de México.