ANALYSIS OF DISCRETE DATA USING LINEAR REGRESSION.

Robert J. Flowers
Universidad Juárez Autónoma de Tabasco.
Instituto de Ingeniería,

ABSTRACT

Maximum likelihood estimates are obtained for a linear regressio. model where the dependent variable is a linear transformation of mutinomially distributed random variables. Comparisons are made with minimum modified chi-square estimates.

Since an iterative re-weighted least squares procedure is used, estimates may be obtained even in the presence of zero cells. For models of marginal homogeneity, it is shown that maximum likelihood cell estimates can differ sharply from minimum modified chi-square cell estimates.

Key words: Multinomial distribution, maximum likelihood estimation, minimum modified chi-square estimation, models of marginal homogeneity, binomial regression.

RESUMEN

Se obtienen estimadores de máxima verosimilitud para un modelo de regresión lineal donde la variable dependiente es una transformación lineal de un vector de variables aleatorias que siguen una distribución multinomial. Se hacen comparaciones con estimadores de mínima ji-cuadrada modificada. Dado que se calculan iterativamente los estimadores usando el método de cuadrados mínimos ponderados, se pueden obtener los estimadores no obstante que haya casillas en la tabla de contingencia que contengan ceros. Para modelos de homogeneidad marginal se muestra que los estimadores de máxima verosimilitud pueden diferir bastante de los estimadores de mínima ji-cuadrada modificada.

Palabras clave: Distribución multinomial, estimadores de máxima verosimilitud, estimadores de mínima ji-cuadrada modificada, modelos de homogeneidad marginal, regresión binomial.

* The work for this article was done while the author was employed by Texas Transportation Institute, Texas A&M University

1. Introduction.

Currently there are two major methodologies for performing linear regression analysis of discrete data. One is a categorical regression procedure in which weighted least square estimates are obtained, and the other is Poisson regression for which maximum likelihood estimates are obtained. While the categorical regression procedure is highly versatile, certain problems result from the use of the sample variance-covariance matrix, particularly when zero cells are present. Poisson regression avoids the problem with the zero cells by an iterative re-weighted least sqares procedure using ordinary least squares estimates as starting values. Unfortunately, Poisson regression is only