

Chapter Title: Introducción

Book Title: Algunos problemas de formalización y estimación en modelos de regresión con variables cualitativas, aplicadas a la investigación social

Book Author(s): Fernando Cortés

Published by: Colegio de Mexico

Stable URL: <https://www.jstor.org/stable/j.ctv233mcp.2>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



This content is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>.



JSTOR

Colegio de Mexico is collaborating with JSTOR to digitize, preserve and extend access to *Algunos problemas de formalización y estimación en modelos de regresión con variables cualitativas, aplicadas a la investigación social*

1. Introducción

Los objetivos centrales de este trabajo son dos: i) establecer un conjunto de principios que nos permitan traducir a modelos matemáticos las proposiciones teóricas que originan variables cualitativas y ii) analizar los problemas de ajuste que se encuentran involucrados en su estimación.

Veamos un poco más de cerca ambos propósitos. El primero dice relación directa con un cambio de lenguaje. En efecto, la formalización puede ser vista como un problema de traducción a términos matemáticos de las hipótesis teóricas, las que normalmente en las ciencias sociales son expresadas en lenguaje natural. Pero no nos interesa cualquier tipo de formalización sino aquel que origina un modelo matemático en que los factores explicativos son cualitativos. Otra restricción a este primer objetivo del trabajo se refiere a que sólo nos preocuparemos por aquel tipo de proposiciones teóricas que dan pie a la construcción de modelos uniecuacionales y lineales o susceptibles de ser linealizados.

En consecuencia nuestro interés es bastante limitado por cuanto se restringe a sólo aquellas hipótesis teóricas que: i) sean susceptibles de ser traducidas a un lenguaje matemático (aquellas que no se puedan llevar al campo de las matemáticas no tienen cabida en este escrito), ii) que el modelo así construido sea uniecuacional y posible de ser llevado a la forma lineal y iii) que todos o algunos de los factores explicativos sean de naturaleza cualitativa.

Para estimar el modelo matemático (segundo objetivo central de este trabajo) es necesario asociarle un modelo estadístico que incorpore los parámetros característicos del primero. Los requisitos que hemos perfilado para el tipo de ecuación que nos preocupa, tienen su correspondencia estadística inmediata en el modelo de regresión lineal, pero no se trataría del modelo ordinario sino de uno que tiene la particularidad de incorporar variables cualitativas.

Lo anterior entra en abierta contradicción con lo sostenido en algunos libros de texto que han tenido amplia repercusión¹ los cuales han difundido la noción que postula que una de las limitaciones más importantes en la aplicación del modelo de regresión a los problemas característicos de las ciencias sociales radicaría en el nivel de medición de las variables: sólo se podría estimar una ecuación de regresión si todas y cada una de sus variables han sido expresadas a nivel métrico.

Dentro de esta línea argumental es que se promueve el uso del análisis de asociación y de covarianzas de Lazarsfeld en aquellos casos en que las variables son nomi-

N. del A.: Agradezco los comentarios y sugerencias realizados por Rosa María Rubalcava, que se han traducido en claridad de exposición y en rigor conceptual. Las imprecisiones que aún subsistan así como las deficiencias que encuentre el lector son de responsabilidad exclusiva del autor.

¹ Un ejemplo conspicuo se encuentra en el libro de Sydney Siegel, *Nonparametric Statistics for the Behavioral Science*, McGraw Hill, 1956.

nales u ordinales. La construcción y estimación de modelos de regresión que incorporen variables cualitativas (normalmente denominadas mudas o ficticias y que corresponden al vocablo inglés *dummy*) pueden ser pensadas como una manera alternativa de analizar el grado de relación que une a dos o más variables. El intento de reemplazar un tipo de instrumento estadístico por otro encuentra su justificación en el mayor poder analítico que caracteriza al modelo de regresión.

En cuanto a la estrategia de exposición hemos decidido destacar en las primeras secciones el enlace entre el pensamiento teórico y su expresión matemática. Para ello supondremos la existencia de hipótesis sustantivas que orienten la descomposición de la variable explicada² en términos de los factores explicativos y a partir de ella preguntarnos por la ecuación matemática y el modelo de regresión que mejor la refleja. Desde el séptimo apartado en adelante nos dedicamos al estudio de los problemas de estimación que son característicos de los modelos de regresión con variables mudas.

Si bien en la primera parte destacamos los problemas de formalización y en la segunda los de estimación, esto no quiere decir que la separación sea taxativa, lo que acontece es que la mezcla en un momento adquiere un tono mayor de formalización y en otro una coloración menor. Además, el orden que seguiremos tiene que ver con una gradación en el orden de dificultad para acceder a la comprensión del material que se entrega. Las últimas secciones requieren que el lector posea un cierto nivel de formación previa en inferencia estadística y en algunos tópicos econométricos, en tanto que las primeras sólo exigen una lectura atenta y detenida.

2. Estimación de la función escalonada

Con el objeto de darle contenido al problema que abordaremos, supongamos que un analista social, después de una serie de operaciones teóricas, llega a establecer que existe una relación entre el grado de urbanización y el volumen de producción industrial. La vinculación entre estas variables descansa en la noción de que las calidades asociadas al espacio físico condicionarían la actividad industrial. En este sentido pareciera ser evidente el papel que juegan las disponibilidades de servicios fundamentales como por ejemplo, energía eléctrica, agua, etc., así como las facilidades de comunicación física con el ambiente exterior, en todo lo que tiene que ver con el movimiento de los insumos y los productos terminados.

Esta aseveración de carácter general se especifica en el momento en que el investigador sostiene que, si bien esta relación existe, ella no es continua. Vale decir, entre determinados niveles de urbanización, la producción industrial es relativamente constante, pero a partir de cierto punto, o de cierto grado de urbanización en adelante, la producción industrial sufre un aumento manteniéndose constante hasta que alcanza otro punto, donde nuevamente experimenta una modificación importante. Este proceso se repite un gran número de veces.

² En lugar del término variable dependiente usaremos los vocablos variable a explicar o explicada, y en lugar de variable independiente el de variable explicativa. Ello se debe, por una parte, a que si bien en las matemáticas es posible plantearse la existencia de una variable que se mueva independientemente, lo mismo difícilmente puede ser sostenido para las variables sociales. Y por otra parte, la diferencia terminológica también alude al hecho que el modelo trata de explicar un fenómeno (representado por una variable) sobre la base del comportamiento de otros fenómenos que son usados como explicativos del primero.