

Fodein

por Heyder Camilo Sanchez Gonzalez

Fecha de entrega: 09-ago-2019 11:23a.m. (UTC-0500)

Identificador de la entrega: 1158912778

Nombre del archivo: 21009_Heyder_Camilo_Sanchez_Gonzalez_Fodein_150775_1408983846.docx (293.52K)

Total de palabras: 3136

Total de caracteres: 17413

DECIMOQUINTA CONVOCATORIA PARA EL FOMENTO DE LA INVESTIGACIÓN Y LA INNOVACIÓN 2020

Estimadores de regresión beta para la media poblacional en muestreo probabilístico: teoría y aplicaciones

Campo de acción	Transdisciplinariedad - Aporte al PIM
<p>Apuestas académicas ya que se pretende realizar una nueva metodología de estimación de proporciones en muestreo probabilístico que supere, en términos de la calidad de la estimación, a las ya existente en este campo.</p>	<p>En el marco de planeación de la Universidad el proyecto aporta a la proyección social e investigaciones pertinentes, numeral 3 PIM, dado que busca focalizar y articular la investigación y la proyección social USTA Colombia con visibilidad e impacto nacional y global; esto es proporcionar espacios y herramientas que permitan generar un mejor proceso del entorno y por ende aportar al conocimiento y divulgación de estrategias que mejoren el desarrollo económico.</p>

Articulación con funciones sustantivas y el sector social y productivo

El proyecto propuesto se ajusta a lo propuesto en el Plan General de Desarrollo Bogotá, en el numeral 3.2, garantizando que el programa y la investigación se articule con las funciones sustantivas de la Universidad.

Grupo de investigación	Línea de investigación en la que se inscribe el proyecto
<p>Ustadística Alternativas de estimación en muestreo probabilístico</p>	

Nombre del Investigador principal	Enlace CvLAC	Enlace ORCID	Enlace Google Académico
<p>Mario José Pacheco Lopez</p>	<p>http://scienti.colciencias.gov.co:8081/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0000775479</p>		
División	Facultad	Programa	Grupo de investigación



Ciencias económicas y administrativas	Estadística	Estadística	USTAdística
Nombre del Co-investigador	Enlace CvLAC	Enlace ORCID	Enlace Google Académico
Cristian Fernando Tellez Piñerez	http://scienti.colciencias.gov.co:8081/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0000016463	https://orcid.org/0000-0003-3869-1831	https://scholar.google.es/citations?user=PJ6JUFwAAAAJ&hl=es
División	Facultad	Programa	Grupo de investigación
Ciencias económicas y administrativas	Estadística	Estadística	USTAdística

Resumen de la propuesta	Palabras clave
<p>En muestreo probabilístico es común tener información auxiliar disponible para los elementos de la población de interés. Los estimadores de regresión son un tipo de estimadores que hacen uso eficiente de dicha información auxiliar. En este trabajo se considera la estimación de la media de variables beta distribuidas. Se propone la construcción de un estimador de regresión beta (BREG) para situaciones en las que la variable de interés está restringida a un intervalo (0,1) y está relacionada con variables auxiliares a través de una estructura de regresión lineal o no lineal. También se tratan los casos importantes de modelos de regresión beta 0-1 inflados y de</p>	<p>Información auxiliar, distribución beta, modelos de regresión beta, máxima pseudo verosimilitud, estimadores de regresión, encuestas, paridad de vida.</p>

Nit. 860.012.357-6

SEDE PRINCIPAL BOGOTÁ - PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9.ª n.º 51-11 / contactenos@usantotomas.edu.co
www.usta.edu.co

DIVISIÓN DE EDUCACIÓN ABIERTA Y A DISTANCIA
PBX: (571) 595 00 00 ext. 2044 / Carrera 10.ª n.º 72-50 / admisiones@ustadistancia.edu.co
www.ustadistancia.edu.co



dispersión variable. Ilustraremos la metodología a través de una aplicación con información proveniente del DANE de las encuestas de hogares, particularmente, la encuesta de calidad de vida.

Problema de investigación

En la práctica, es común encontrar variables de interés que se encuentren restringidas a intervalos de tipo $[a,b]$ tales como índices o proporciones (porcentajes). Particularmente en Colombia, el índice de pobreza multidimensional y la tasa de desempleo son dos de los temas más consultados en la actualidad que cumplen con la característica de estar restringidos en un intervalo. Ahora bien, cuando se dese hacer estimaciones de parámetros poblaciones, como los mencionados anteriormente, dentro de la teoría de muestreo probabilístico, es posible construir estimadores de regresión (estimadores que utilicen modelos estadísticos de regresión para su construcción) que empleen de manera eficiente la información auxiliar disponible para los elementos de la población. Entiéndase como información auxiliar, toda aquella información relacionada directamente con la variable de interés; por ejemplo, para calcular el índice de pobreza multidimensional, se puede tomar como variables auxiliares el número de empleados en cada hogar, tipo de suelo de la casa, entre otros. Sin embargo, la teoría disponible para la estimación de totales o medias de poblaciones finitas mediante estimadores de regresión no está disponible para variables restringidas a intervalos $(0,1)$ o para el caso general de variables restringidas a intervalos del tipo $[a,b]$, con a y b arbitrarios. Este tipo de variables de interés presentan algunos problemas adicionales que tampoco han sido abordados desde la teoría de estimación mediante estimadores de regresión en muestro probabilístico. Tal es el caso de las variables cero (a) infladas y uno (b) infladas. Es por esto y por ser variables de mucho interés práctico, que se hace necesario abordar la teoría de estimación por medio de estimadores de regresión beta, que permiten la construcción de estimadores de regresión generalizados para este tipo de variables. Este tipo de variables se encuentran presentes en, por ejemplo, la encuesta de calidad de vida (ECV) hecha por el DANE, y es común que los estimadores empleados para la estimación de los parámetros de interés no sean los más adecuados, provocando sesgos mayores en las estimaciones. Por tanto, es necesario un cambio metodológico en las herramientas de estimación que generen un impacto positivo en cuanto a eficiencia de las estimaciones lo que conlleva a una mejor toma de decisiones, particularmente, de política pública.

Justificación

En la teoría de muestreo probabilístico es necesario construir estimadores de parámetros de poblaciones finitas para variables de diferente naturaleza, sean continuas o discretas. El desarrollo de estos estimadores puede depender de la disponibilidad de información auxiliar acerca



de los elementos de la población de interés y de la relación funcional entre la variable de interés y las variables auxiliares. El aprovechamiento de esta relación ha permitido la construcción de los denominados estimadores de regresión (ver por ejemplo Cassel, Särndal, y Wretman (1976)). Los estimadores de regresión disponibles en la literatura se basan principalmente en el modelo normal y más generalmente en los modelos lineales generalizados que permiten modelar variables de tipo normal y una amplia gama de distribuciones no normales, por ejemplo, para variables asimétricas o de conteo. Pero no tienen en cuenta otro tipo de modelos para variables restringidas a un intervalo o las variables de proporciones que están restringidas a un intervalo (0,1) o de manera más general a un intervalo de tipo [a,b]. Es por todo esto que se hace necesario desarrollar la teoría relacionada con la construcción de estimadores de regresión para variables de este tipo, siendo el candidato natural los modelos de regresión Beta y sus extensiones.

En la práctica, ¹ las encuestas dirigidas a hogares constituyen una de las principales fuentes de datos socioeconómicos con las que cuentan los países. A partir de la información obtenida de ellas, se calculan indicadores para la medición de variados aspectos económicos y sociales. Además, facilitan el conocimiento y explicación los determinantes o factores causales del comportamiento de dichos aspectos, lo cual es de gran importancia para el diseño, monitoreo y medición de resultados de política pública.

Una de las principales encuestas de hogares es ² la encuesta de calidad de vida (ECV). Dicha encuesta es una investigación que el DANE realiza con el objeto de recoger información sobre diferentes aspectos y dimensiones del bienestar y las condiciones de vida de los hogares, incluyendo temas como: el acceso a bienes y servicios públicos, privados o comunales, salud, educación, atención integral de niños y niñas menores de 5 años, entre otros. La consideración de estos aspectos hace posible realizar posteriores análisis a los factores que explican los diferentes niveles de vida existentes en la sociedad. Sin embargo, en 2017, esta encuesta, desde la parte técnica, tuvo falencias ya que la muestra que se seleccionó no fue representativa para los niveles de desagregación que se tenía, además, las medidas de calidad de las estimaciones que se realizaron fueron deficientes. Claramente, el impacto que tuvo esto a nivel nacional, fue grande. El estimador que se propone en esta metodología permitirá generar estimaciones con medidas de calidad muy superiores si se tiene en cuenta que es posible contar con buenas variables auxiliares (altamente correlacionadas) lo que conlleva a un cambio metodológico en las estimaciones de proporciones, medias y totales en empresas como el DANE y cualquier otra empresa que realice este tipo de encuestas.

Objetivo general

Desarrollar la teoría relacionada para la construcción de estimadores de regresión para variables de tipo razones y proporción mediante los modelos de regresión Beta y sus extensiones.

Objetivos específicos

1. Construir un estimador de regresión beta (BREG) para situaciones donde la variable de interés está restringida a un intervalo de tipo $(0, 1)$ o (a, b) y está relacionada a variables auxiliares a través de una estructura de regresión lineal o no lineal.
2. Ampliar la teoría de estimación en muestreo probabilístico asistido por modelos a estimadores de regresión beta generalizados (GBREG) para casos de variables 0-1 o a-b infladas y de dispersión variable.
3. Aplicar el estimador propuesto a los datos de la encuesta de calidad de vida y comparar la eficiencia con las estimaciones ya publicadas.

Estado del arte y marco conceptual

En muestreo probabilístico, es posible tener disponible información auxiliar para elementos de la población de interés en la forma de una o más variables ³ auxiliares. Los estimadores de regresión son un tipo de estimadores que hacen un uso eficiente de la información auxiliar sobre la población. El estimador ³ de regresión más popular es el estimador de regresión general (GREG) (Casel et al., 1976; Särndal, 1980), que puede derivarse cuando la ³ relación entre la variable de interés y las variables auxiliares puede representarse mediante un modelo lineal normal. Otros tipos de estimadores de regresión generalizados son los estimadores de calibración (Deville y Särndal, 1992) y los estimadores de verosimilitud empírica (Chen y Qin, 1993).

Además del estimador de regresión lineal, existen otros tipos de estimadores de regresión, que consideran diferentes ³ tipos de ³ relación entre la variable de interés y las variables auxiliares, así como diferentes escalas de medición de la variable de interés. Por ejemplo, Lehtonen y Veijanen (1998) desarrollaron un estimador de regresión logística generalizada (LGREG) cuando la variable de interés es categórica, Breidt y Opsomer (2000) proponen una clase de estimadores basados en regresión polinómica local, y Rondon et al. (2012) introdujeron un estimador de regresión asistidos por el modelo lineal generalizado en el que la variable de interés es vista como si fuera generada por un miembro de la familia exponencial.

Aunque el estimador propuesto por Rondón et al. (2012) permite construir estimadores para totales de características de una población finita cuando la variable de interés pertenece a la familia exponencial, esto puede no ser apropiado para situaciones donde la respuesta está restringida a un intervalo de la forma $(0, 1)$ ya que puede producir valores ajustados para la variable de interés que exceden sus límites inferior y superior. En muchos problemas de muestreo, es común encontrar variables respuesta que tienen soporte en el intervalo $(0, 1)$ como tasas, proporciones, concentraciones, etc. Una posible solución es emplear una transformación en la variable respuesta continuando empleando el supuesto de normalidad; sin embargo, esta no es la forma natural de trabajar con tales variables. Otra posible solución es trabajar la variable respuesta como beta distribuida y ajustar los modelos de regresión beta como se propone en Ferrari y Cribari-Neto (2004), que también es útil en situaciones donde la respuesta está restringida a un intervalo arbitrario (a, b) , junto con las extensiones de Smithson y Verkuilen (2006), Simas et al. (2010) y Ospina y Ferrari (2012).

La aplicación que queremos hacer con la metodología propuesta tiene en cuenta que las encuestas dirigidas a hogares constituyen una de las principales fuentes de información para los gobiernos, las cuales surgen como respuesta a la necesidad de caracterizar la población en relación al bienestar de los hogares, el DANE frente a la necesidad de profundizar en la caracterización de la población pobre, a través de aspectos del bienestar como la salud, educación, actividades laborales, tenencia de bienes en el hogar, etc., y dadas las restricciones de los instrumentos utilizados exclusivamente para la medición de la pobreza, identificó la necesidad de diseñar una encuesta de calidad de vida que captara información sobre estos y otros aspectos. Es así que desde 1991 se viene aplicando la Encuesta de Calidad de Vida (ECV) y los procesos de estimación de parámetros de interés se han ido mejorando en busca de una mayor eficiencia estadística en todo este proceso. Estos procedimientos se aplican para estimar con la muestra los parámetros poblacionales, teniendo en cuenta los factores básicos de expansión, las variaciones del marco muestral, los ajustes por no respuesta, la calibración a valores poblacionales, la generación de los estimadores y la determinación de la precisión de estos. En particular la metodología de ajuste a las proyecciones de población se basa en la calibración por regresión (DANE, 2019).

Metodología

La metodología para desarrollar este proyecto contempla los siguientes pasos:

1. La revisión bibliográfica de los trabajos clásicos y recientes de la teoría de estimación en muestreo probabilístico mediante estimadores de regresión y de la teoría de modelos beta para variables de razón y proporciones.
2. Construir del estimador de regresión beta para variables de interés en el intervalo (0,1) empleando métodos de máxima verosimilitud para datos provenientes de estudios por muestreo con probabilidades distintas.
3. Obtener las expresiones para la varianza de los estimadores y construir los estimadores de dicha varianza, estableciendo las restricciones y propiedades correspondientes.
4. Desarrollar la teoría de un estimador de regresión beta generalizado para variables 0-1 infladas y de dispersión variable, su varianza y sus estimadores.
5. Desarrolla un estudio de simulación para evaluar el comportamiento estadístico de los estimadores propuestos
6. Finalmente, se realizará una aplicación de la metodología propuesta en la que se generarán estimaciones junto con sus medidas de calidad que permitan demostrar la ganancia en eficiencia al incorporar variables auxiliares del cambio metodológico propuesto para las estimaciones en el caso particular de la encuesta de calidad de vida.

Los pasos anteriores describen de manera completa y clara cómo se llevará a cabo este proyecto y cómo se abordará, desde la teoría y estudios de simulación, cada paso de la metodología.

FINANCIACIÓN	RECURSO	DESCRIPCIÓN	Valor partida	Valor contrapartida (Externa)	Total (\$)
RUBROS	Servicios Técnicos				\$ 0
	Salidas de campo				\$ 0
	Equipos				\$ 0
	Materiales, insumos y software				\$ 0
BOLSAS	Papelería	Se requiere papelería para imprimir el documento, poster, entre otros.	\$ 100.000		\$ 100.000
	Fotocopias	Se requiere fotocopias de la documentación, posters, etc.	\$ 50.000		\$ 50.000
	Material bibliográfico	Libros, suscripciones de revista, etc.	\$ 500.000		\$ 500.000
	Auxilio de transporte				\$ 0
	Movilidad	Se requiere este rubro para movilidad en eventos académicos nacionales o internacionales	\$ 8.000.000		\$ 8.000.000
	Publicaciones (Artículos, proceso editorial y traducción)	Se requiere este rubro para posibles costos de publicación del artículo	\$1.000.000		\$ 1.000.000
TOTAL DEL PROYECTO:					\$ 9.550.000

Referencia bibliográficas

- Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review* 51, 279-292.
- Breidt, F. J., Opsomer, J. D. (2000). Local polynomial regression estimators in survey sampling. *Annals of Statistics* 28, 1026-1053.
- Cassel, C. M., Särndal, C. E., Wretman, J. (1976). *Foundations of Inference in Survey Sampling*. Wiley, New York, United States.
- Chen, J., Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective use of auxiliary information. *Biometrika* 80, 107-116.
- Departamento Administrativo Nacional de Estadística (DANE) (2019). *Metodología general encuesta nacional de calidad de vida - ECV..* Bogotá D.C.
- Deville, J. C., Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* 87, 376-382.
- Ferrari S. L. P. and Cribari-Neto F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31, 7, 799-815.



Kennel, T. L., Valliant, R. (2010). Logistic generalized regression (LGREG) estimator in cluster samples. Proceedings of the section on survey research methods, Washington D.C., American Statistical Association, 4756-4770.

Lehtonen, R., Veijanen, A. (1998). Logistic generalized regression estimators. Survey Methodology 24, 51-55.

Ospina, R., Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. Computational Statistics and Data Analysis 56, 1609-1623.

Rondon, L. M., Vanegas, L. H., Ferraz, C. (2012). Finite population estimation under generalized linear model assistance. Computational Statistics and Data Analysis 56, 680-697.

Särndal, C. E. (1980). On π -inverse weighting versus best linear unbiased weighting in probability sampling. Biometrika 67, 639-650.

Simas, A., Barreto-Souza, W., Rocha, A. V. (2010). Improved estimators for a general class of beta regression models. Computational Statistics and Data Analysis 56, 1609-1623.

Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological Methods 11, 54-71.

Fodein

INFORME DE ORIGINALIDAD

10%

INDICE DE SIMILITUD

10%

FUENTES DE INTERNET

2%

PUBLICACIONES

8%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1

microdatos.dane.gov.co

Fuente de Internet

3%

2

formularios.dane.gov.co

Fuente de Internet

3%

3

bdigital.unal.edu.co

Fuente de Internet

2%

4

dane.gov.co

Fuente de Internet

2%

Excluir citas

Apagado

Excluir coincidencias

< 2%

Excluir bibliografía

Apagado

Fodein

INFORME DE GRADEMARK

NOTA FINAL

/100

COMENTARIOS GENERALES

Instructor

PÁGINA 1

PÁGINA 2

PÁGINA 3

PÁGINA 4

PÁGINA 5

PÁGINA 6

PÁGINA 7

PÁGINA 8

PÁGINA 9

PÁGINA 10
