Anexo 1. Formato de presentación de proyecto

Título: Una aplicación de estimación de áreas pequeñas utilizando imputación múltiple con modelos logísticos de tres parámetros en pruebas estandarizadas

Duración (en meses):9

Lugar de ejecución: Bogotá

Investigador principal: Cristian Fernando Tellez Piñerez

Co-investigador(es): Andrés Gutiérrez Rojas

Programa(s)	Facultad(es)	Datos generales			
		Línea activa	Línea medular	Campos de acción institucional (Selecione)	Grupo(s) de investigación
Facultad de Estadística	Facultad de Estadística	Alternativas de estimación en muestreo probabilístico	Línea Alberto Magno	Derechos humanos, ciudadanía y construcción de política pública en y para escenarios de paz () Desarrollo tecnológico con apuesta social () Desarrollo ambiental sostenible (X) Cambio educativo y social desde la multi e interculturalidad ()	USTAdística

Equipo de investigación requerido: 2 Docentes internos.

Alianza estratégica: N.A

BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

Resumen de la propuesta (máximo 300 palabras)

La generación de estadísticas de alta calidad y bajo costo es una necesidad para los tomadores de decisiones hoy en día, en educación estos datos son necesarios para decidir acerca de la creación de políticas, la continuidad de programas existentes y la asignación de recursos año a año. Se propone una metodología para la estimación del rendimiento de los establecimientos educativos en Colombia a partir de la observación de una muestra de ellos y de información auxiliar pertinente. La propuesta combina la construcción de un modelo de Fay Herriot basado en las metodologías de estimación en áreas pequeñas con la estimación múltiple del desempeño de cada estudiante vía valores plausibles. Al final, se presentan los resultados de la aplicación de la metodología propuesta para los establecimientos de Colombia

Palabras clave (máximo 5)

Teoría de respuesta al ítem, Estimación en áreas pequeñas, pruebas estandarizadas.

Planteamiento del problema y pregunta de investigación

Existe una demanda creciente en el uso del análisis estadístico en cualquier área del conocimiento. Una de las razones por las cuales se emplea tal análisis se debe a que su objeto es extraer la información relevante para deducir propiedades de la población y otra es la toma de decisiones con argumentos científicos.

La generación de estadísticas de alta calidad y bajo costo es una necesidad para la toma de decisiones hoy en día; por ejemplo, en el sector educativo sería útil generar dichas estadísticas que permitan decidir acerca de la creación de políticas educativas a todas las escuelas, la continuidad de programas existentes y la asignación de recursos año a año solo con una muestra de escuelas. Otro ejemplo se da en las encuestas de hogares donde, por razones de seguridad no se pueda llegar a todas las viviendas, sin embargo, se deben reportar estadísticas, inclusive a aquellas en donde no se pudo llegar. Estos subconjuntos de la población donde no se recolectó información (o la que se tiene es insuficiente para poder realizar una estimación directa precisas y confiables) se conoce como área pequeña.

Según Pfeffermann (2002) y Rao (2003) para obtener estimaciones precisas y confiables en áreas pequeñas, es necesario utilizar otro tipo de estimadores que no solo se basen en el diseño de muestreo, sino que también utilicen, entre otras, valores de la variable de interés que se encuentran en áreas vecinas. Esta metodología se conoce como estimación en áreas

BOGOTÁ - PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA - Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

Battese et al. (1988) y Ghosh & Rao (1994). Existen otros trabajos en este campo que también han aportado significativamente al desarrollo de SAE como por ejemplo Chambers & Tzavidis (2006), Falorsi & Righi (2008), Chambers et al. (2009), Pfeffermann & Correa (2012).

Según Australian Bureau of Statistics (2006), la estimación de áreas pequeñas son métodos que producen estimaciones suficientemente confiables para áreas geográficas demasiado pequeñas y en las cuales no se podría obtener estimaciones con precisión, usando estimación de muestreo directa. Por estimación directa, se refiere a métodos clásicos de estimación de encuestas basados en diseño de muestreo que utilizan sólo las unidades de muestra contenidas en cada área pequeña. SAE se utiliza para superar el problema de tamaños pequeños o nulos de muestra y con esto mejorar la calidad de las estimaciones directas.

Las metodologías de SAE son sofisticadas puesto que se basan en modelar las diversas relaciones que hay en los datos observados utilizando un modelo estadístico y con esto poder predecir en subgrupos donde no hubo muestra o la que hay es insuficiente para realizar una estimación directa de calidad. En esta dirección, los estimadores de regresión proporcionan predicciones más precisas para las áreas pequeñas teniendo en cuenta la información auxiliar que se pueda obtener e incluyéndolas como predictores al modelo.

Los estimadores directos se basan en la muestra seleccionada; Sin embargo, tienen baja precisión en áreas donde el tamaño de la muestra es pequeño. El modelo SAE combina los estimadores de regresión y la estimación directa equilibrando sesgo y precisión. Para que esta técnica tenga éxito, Longford (2005) afirma que las áreas que tienen características similares deberán tener estimaciones parecidas.

Como bien se ha mencionado, SAE requiere de información completa en las áreas donde se va a tomar información para predecir las áreas pequeñas, sin embargo, en ocasiones es factible encontrar datos faltantes en dichas áreas haciendo más complejo la utilización de SAE. Uno de los primeros acercamientos que hubo alrededor de este problema fue trabajado por Longford (2004) en el cual toman la metodología propuesta por Rubin para realizar imputación múltiple, a través de valores plausibles, y con estos poder obtener estimaciones en las áreas pequeñas de interés. Sin embargo, no se ha encontrado literatura que implemente SAE y teoría de respuesta al ítem (TRI por sus siglas) utilizando imputación múltiple y dada la importancia que tendría esta metodología en el sector educativo y en la generación de insumos para política pública, esto se convierte en el objetivo principal de este trabajo.

BOGOTÁ - PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA - Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

entre otras que utilizan para la generación de los resultados modelos de teoría de respuesta al ítem. Particularmente, en Colombia un incentivo para los estudiantes de educación media, es el programa Ser Pilo Paga.

Luego de una breve introducción al problema, surge la siguiente pregunta de investigación ¿será posible encontrar una metodología que incorpore los modelos de TRI con imputación múltiple y la metodología SAE?

Marco teórico:

En esta sección se mostrarán, de manera general, las definiciones más relevantes que permitan el desarrollo de esta propuesta. Primero se abordarán los conceptos relacionados con muestreo y posterior a eso, los conceptos relacionados con teoría de respuesta al ítem.

Concepto básico de muestreo

El muestreo es un procedimiento que responde a la necesidad de información estadística precisa sobre la población y los conjuntos de elementos que la conforman; el muestreo trata con investigaciones parciales sobre la población que apuntan a inferir a la población completa. Según Gutiérrez (2009), una muestra bien seleccionada de unos cuantos miles de individuos puede representar con gran precisión una población de millones. Es requisito fundamental de una buena muestra que las características de interés que existen en la población se reflejen en la muestra de la manera más cercana posible, para esto se necesitan definir los siguientes conceptos.

Muestra aleatoria

Una muestra aleatoria es un subconjunto representativo de la población que ha sido extraído mediante un mecanismo estadístico de selección.

Subgrupos poblacionales

En general, cuando se habla de subgrupos poblacionales se está haciendo referencia a dominios de interés, estratos o postestratos. Cuando el investigador se enfrenta a una encuesta que tiene en cuenta algunos subgrupos poblacionales, es indispensable conocer en qué se diferencian cada uno de ellos pues de esto depende que la investigación arroje resultados confiables mediante el planteamiento de la mejor estrategia de muestreo. Las

BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

Estos requerimientos se planean en la etapa de diseño para asegurar que el diseño de la muestra sea tal que al momento de la recolección de la información exista una buena cobertura en cada uno de los dominios de interés. Lo anterior sólo se puede lograr ampliando el tamaño de muestra n puesto que el marco de muestreo no informa acerca de la pertenencia de los individuos a los dominios de interés. La estimación por dominios se caracteriza por el desconocimiento de la pertenencia de las unidades poblacionales al dominio.

Estratos. Cuando el marco de muestreo permite conocer la pertenencia de todos los individuos de la población a un subgrupo poblacional, se dice que esta clase de subgrupos se llaman estratos. Más aun, cuando se sabe que la característica de interés tiene un comportamiento distinto en cada uno de los estratos y se planea un diseño de muestreo que tenga en cuenta este aspecto mediante la selección aleatoria de unidades en cada uno de los estratos, se dice que el diseño de muestreo es estratificado. El aspecto fundamental de esta clase de subgrupos poblacionales es que el conocimiento de la pertenencia de los individuos a los estratos se incorpora en la etapa de diseño de la muestra. Nótese que, a diferencia de los dominios, en los estratos se conoce tanto N_h como n_h antes de la etapa de estimación.

Postestratos. La propiedad que caracteriza a este tipo de subgrupos poblacionales es que, aunque en la etapa de diseño el tamaño del postestrato N_g es conocido, se desconoce el número de individuos que pertenecerán al postestrato n_g en la muestra realizada.

Estimadores

En este apartado se abordan los distintos tipos de estimadores que existen para la evaluación de los totales o las medias de las variables en áreas pequeñas, realizando una somera descripción de los distintos tipos. Existen tres grandes clases de estimadores: los estimadores directos, que están basados en el diseño de muestreo, los estimadores indirectos, basados en posibles relaciones estructurales con variables auxiliares y los compuestos, definidos como una combinación de los dos anteriores. Para los primeros, la escasez de muestra hace que tengan grandes varianzas, mientras que las otras dos clases de estimadores necesitan de una buena información auxiliar que permita relacionar la información disponible a nivel de las áreas pequeñas.

BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

Estimadores directos

Los estimadores directos son aquellos donde solo se tiene en cuenta la muestra seleccionada para estimar los parámetros, entre ellos el más utilizado es el estimador de Horvitz-Thompson.

Estimador Indirecto

Es frecuente en el muestreo usar información adicional para mejorar la precisión de las estimaciones, esto se conoce como estimadores indirectos. Los métodos indirectos requieren el poder observar una variable auxiliar relacionada con las variables en estudio de interés, además conocer el total de la variable auxiliar. A continuación, definiremos los estimadores sintéticos y estimadores compuestos que son de la familia de los estimadores indirectos.

Estimador post-estratificado

En algunas ocasiones prácticas es útil estratificar. Sin embargo, si la información para estratificar no viene incluida en el marco, para todas las unidades de la población no es posible hacer dicha estratificación. Cochran (1977) afirma que algunos ejemplos particulares donde no se conoce la pertenencia de las unidades antes de la recolección de los datos son, entre muchas otras, la edad, el género y el nivel educativo. Nótese que el total poblacional de las anteriores variables puede ser conocido recurriendo a alguna fuente externa confiable o a estadísticas oficiales. Por lo tanto, es posible construir un estimador que tenga en cuenta el conocimiento de estos totales y que incorpore la información recolectada en la muestra acerca de la pertenencia de cada unidad a un subgrupo poblacional.

El estimador postestratificado puede considerarse como estimador asistido por el uso de variable auxiliar. Dicha variable es utilizada para reducir el sesgo generado por el desconocimiento de la pertenencia de los individuos a los subgrupos poblacionales (Little 1993). Este estimador tiene muchas ventajas; entre ellas, es el estimador máximo verosímil ofrece protección contra configuraciones muestrales desfavorables siendo considerado como un estimador robusto.

Estimador sintético

Gonzalez (1973) describe los estimadores sintéticos como aquellos estimadores que utilizan

BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

puede suponer que las áreas pequeñas tienen las mismas características que las áreas grandes con respecto a la variable de interés.

Estimador compuesto

Una forma de equilibrar el sesgo potencial del estimador sintético y la inestabilidad del estimador directo es tomar una media ponderada de ambos, el cual se denomina estimador compuesto.

Estimación en áreas pequeñas

La estimación en áreas pequeñas hace referencia a la estimación para dominios que tienen un tamaño de muestra efectivo relativamente pequeño, el cual hace que las estimaciones directas calculadas tengan errores de muestreo altos y por lo tanto inhabilitan su utilización. Los métodos para llevar a cabo esta clase de estimaciones están basados en el ajuste de modelos mixtos, que toman en cuenta las variaciones existentes dentro de cada dominio (establecimiento educativo).

El ajuste de este tipo de modelos permite aumentar la información disponible en el área pequeña, integrando la información provista por variables auxiliares o por la misma variable analizada en un periodo de tiempo distinto, este aumento de información viene acompañado de una disminución en el error de muestreo y una mejora sustancial en la calidad de las estimaciones. A continuación, se presenta el modelo de Fay Herriot como una de las herramientas más utilizadas para la estimación en áreas pequeñas.

Aspectos generales de los modelos logísticos de tres parámetros

Otro aspecto a considerar en el desarrollo de este trabajo, aparte de la estimación en áreas pequeñas son los conceptos básicos de teoría de respuesta al ítem que tienen su representación más general en los modelos de tres parámetros. Con la finalidad de modelar variables que no son observables directamente, o más conocidas como variables latentes, aparecen los modelos de rasgos latentes o más conocidos como modelos de respuesta al ítem. En estos modelos se busca representar con claridad la relación entre los diversos reactivos, ítems o conductas observadas con la variable no observable de interés, también llamada variable latente (Cervantes et al. 2008). Lord (1980) establece que estos modelos son una relación funcional entre la probabilidad de que un sujeto dé cierta respuesta a un

BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

La interpretación de la escala que proveen los modelos de respuesta al ítem es uno de los principales asuntos prácticos que deben abordar el constructor de instrumentos, el profesional que aplica e interpreta cuestionarios de evaluación y la persona que responde a los mismos. El primer paso para lograr esta interpretación es la estimación de los parámetros involucrados en el modelo. El segundo paso es dar cuenta de qué representan los diferentes valores de la escala de medición y que posteriormente serán asignados a un sujeto a quien se evalúan sus capacidades o preferencias mediante la aplicación del instrumento. Una herramienta interesante que ha adquirido importancia en la interpretación de evaluaciones nacionales de competencia es el procedimiento de anclaje de la escala. Al emplear un instrumento específico, por ejemplo, una prueba psicológica, no es posible dar cuenta de todos los posibles valores de la escala, por lo cual en el procedimiento propuesto por Beaton & Allen (1992) se procede a la selección a priori de algunos valores a los cuales se intenta asignar un significado respecto al constructo evaluado.

Por su parte, Lord et al. (1968), Hambleton et al. (1991) y de Andrade et al. (2000) entre otros autores afirman que la teoría de respuesta al ítem tiene dos supuestos fundamentales; el primero es que existe una relación funcional entre la magnitud en la característica latente y la probabilidad de que un individuo dé o presente cierta respuesta en las variables observadas y el segundo supuesto de los modelos de variables latentes es la independencia local entre los ítems; es decir que para un individuo con cierto nivel de habilidad, la respuesta que da a un ítem es estadísticamente independiente de la respuesta que da a otro ítem .

Adicional a los supuestos anteriores, se pueden imponer otros supuestos sobre la magnitud del atributo (θ), las posibles respuestas a los ítems (U) y cómo se relacionan estas dos mediante una función f , se obtienen distintos modelos de teoría de respuesta al ítem. Según de Andrade et al. (2000), en los modelos más utilizados el atributo $\theta \in (-\infty, \infty)$, los ítems pueden tomar los valores correcto (1) o incorrecto (0), es decir, son modelos en los que el atributo es unidimensional y los ítems son dicotómicos. Adicional a lo anterior, el gráfico de la función f se conoce como curva característica del ítem (CCI).

De acuerdo con la forma y parametrización específica de la función, surgen diferentes modelos. Los modelos más utilizados en la evaluación de atributos unidimensionales, para los cuales se han observado variables dicotómicas, son los modelos logísticos; los

BOGOTÁ - PBX: (571) 587 87 97 **Línea gratuita nacional:** 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA - Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

Modelo logístico de tres parámetros (3PL)

Este modelo fue desarrollado por Birnbaum (1958) y asume que la función característica está dada por un modelo logístico que contempla la dificultad del ítem (b_i), la discriminación (a_i) y el azar (c_i). Por su parte, este modelo ha recibido mayor aceptación por parte de los especialistas, dado que se ajusta de una mejor forma a las situaciones reales y se define:

$$P_i(U_{ik} = 1 | \theta_k, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{Da_i(\theta_k - b_i)}}{1 + e^{Da_i(\theta_k - b_i)}}$$

con $i=1,2,\dots,I$ y $k=1,2,\dots,n$ donde U_{ik} es una variable dicotómica y toma los valores uno cuando el individuo j responde correctamente al ítem i y cero si no responde de forma correcta. θ_k y b_i están en la misma escala y toman valores en el intervalo $(-\infty, \infty)$, c_i indica la probabilidad de que un individuo responda correctamente al ítem i cuando su valor de atributo tiende a menos infinito, y D es un factor de escala constante para todos los ítems y toma valores 1 ó 1.7.

Metodología:

En esta sección se presenta una propuesta previa de cómo se abordará el problema de la estimación en áreas pequeñas con datos faltantes utilizando un modelo de 3PL tomando como insumo las herramientas teóricas descritas en el capítulo anterior. En primer lugar, el problema de los datos faltantes se trabajará con los valores plausibles y la estimación en áreas pequeñas se hará con el modelo de Fay-Herriot y finalmente, se aplicará esta metodología a la prueba de Saber 3°, 5° y 9° que aplica el Icfes.

La prueba Saber 3°, 5° y 9° es una evaluación externa estandarizada que desde el año 2013 es aplicada anualmente a los estudiantes de grado tercero, quinto y noveno de los establecimientos educativos oficiales y no oficiales, rurales y urbanos en Colombia. La construcción, aplicación, calificación y divulgación de resultados de esta prueba está a cargo del Instituto Colombiano para la Evaluación de la Educación Icfes. Esta prueba fue diseñada con el fin de alinear los Estándares Básicos de Competencias establecidos por el Ministerio de Educación Nacional y la evaluación de la educación básica primaria y secundaria como componente fundamental de la estrategia a favor de la calidad de la educación en el país.

BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9, n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10, n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

asociados a la estimación del área en estudio, así como el parámetro asociado a la variabilidad de los resultados al interior de cada establecimiento educativo.

Información requerida a solicitar

De acuerdo a la información consignada en las bases de datos del Icfes, existen 11 226 establecimientos educativos a los cuales se les reportó la calificación de matemáticas para la prueba Saber 5; en el año 2013. Esta cantidad excluye a los establecimientos con 5 o menos estudiantes a las cuales no se les reporta la calificación y no es objetivo de esta metodología cambiar esta realidad.

Sobre estos 11 226 establecimientos, se cuenta con la información de todos los estudiantes que presentaron las pruebas Saber en 2013 y las siguientes variables:

- Zona, sector y entidad territorial a la que pertenece el establecimiento.
- Resultados para matemáticas y lenguaje en las pruebas Saber 3°, 5° y 9° de los establecimientos.
- Resultados de los establecimientos en las áreas de matemáticas, lenguaje, ciencias sociales, filosofía, biología, química, física e inglés en las pruebas Saber 11° de 2012.
- Sexo, educación, ocupación de los padres y algunas posesiones del hogar como carro, calentador, lavadora, que permitan diferenciar el nivel socioeconómico de los estudiantes.

Adicionalmente, de los 11 226 establecimientos con información reportada en el área de matemáticas de grado 5°, 477 hacen parte de la muestra controlada, la cual es una muestra seleccionada por el Icfes para presentar las pruebas bajo la observación de un veedor contratado por la entidad para la vigilancia de la prueba. El ejercicio planteado, pretende lograr estimaciones para los 11 226 establecimientos a partir de la muestra de 477.

Ejercicio a nivel de entidad territorial y establecimiento

Antes de analizar el procedimiento de estimación para los establecimientos de Colombia, se plantea la construcción de un ejercicio donde se busque predecir el resultado del área de matemáticas de la prueba Saber 5°, para todas las 94 entidades territoriales en las cuales se divide el país. Para esta estimación, se propone el ajuste de un modelo de Fay-Herriot. Recordemos que el ajuste de este tipo de modelos requiere de dos insumos principales.

1. Un modelo para el resultado en matemáticas de la prueba Saber 5° a nivel de entidad territorial (el nivel superior).
2. Una varianza del resultado en matemáticas de la prueba Saber 5° para cada entidad

BOGOTÁ - PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

Para el primer requerimiento, se propone ajustar un modelo que considere las siguientes variables:

- resultado de matemáticas estimado para la prueba Saber 5° en el i -ésimo establecimiento.
- resultado de matemáticas observado para la prueba Saber 3° en el i -ésimo establecimiento.
- el resultado de matemáticas observado para la prueba Saber 9° en el i -ésimo establecimiento.
- el resultado de lenguaje observado para la prueba Saber 3° en el i -ésimo establecimiento.
- el resultado de lenguaje observado para la prueba Saber 9° en el i -ésimo establecimiento.

En el cual, el resultado promedio de matemáticas en la prueba Saber 5° en la i -ésima entidad territorial es explicado por los resultados de matemáticas y lenguaje de las pruebas Saber 3° y 9° de la misma entidad. Adicional a esto, el promedio es obtenido a través de un modelo 3PL y la utilización de 5 valores plausibles, lo cual conlleva a el ajuste de 5 modelos SAE.

Como resultado de este ejercicio, se genera una estimación para el resultado de matemáticas de la prueba Saber 5° para cada entidad territorial por dos caminos, combinado con la varianza del estimador Horvitz-Thompson y con la varianza de un estimador de regresión.

Resultados esperados:

- Encontrar una metodología eficiente y confiable para estimar la habilidad promedio en las escuelas que no se observaron
- Generar una línea de investigación y poder extender esta metodología a otros ámbitos del saber
- Generar espacios de divulgación de los avances en la materia.

Productos esperados

- Postular artículo en revista indexada Tipo B
- Ponencia en eventos académicos

Contribución del proyecto al cumplimiento de la misión institucional: (Se pueden consultar

BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

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.

Con qué acciones del Plan General de Desarrollo Bogotá, se articula el proyecto:

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.

Presupuesto

Recurso solicitado FODEIN		
Concepto	Descripción	Monto
Personal científico	Cristian Fernando Tellez	270 * 36.000 9.720.000
	Hugo Andrés Gutiérrez	270 * 36.000 9.720.000
Auxilio a investigadores	Reconocimiento económico a estudiantes de pregrado	N.A
Asistentes de investigación	Reconocimiento económico a estudiantes de posgrado	
Equipos	Consultar en adquisiciones y suministros para evitar duplicidad	N.A
Software	Consultar en departamento TICS para evitar duplicidad	
Materiales		
Papelería		80.000
Fotocopias		80.000
Salidas de campo	Lugar, tiempos, actividades, investigadores	300.000
Material bibliográfico	Libros, suscripciones a revistas, etc	500.000
Publicaciones	Libros, traducciones publicación en revistas

BOGOTÁ - PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
 Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA - Tel. (571) 595 00 00
 Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

Organización de eventos	Eventos para difusión de resultados	
	Total	28.400.000
Contrapartida externa		
Para proyectos en cooperación y alianza estratégica		
Institución	Descripción	Monto
	Detalle los montos y los conceptos	
	Total	

Cronograma:

Actividad	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8
Revisión Bibliográfica	X	X						
Desarrollo teórico			X	X	X			
Consecución de la información					X	X		
Desarrollo computacional						X	X	
Elaboración del documento							X	X

Bibliografía:

- Australian Bureau of Statistics (2006), A guide to small area estimation.
- Birnbaum, A. (1958), 'On the estimation of mental ability', *Series Rep* 15, 7755-7723.
- Battese, G. E., Harter, R. M. & Fuller, W. A. (1988), 'An error components model for prediction of county crop areas using survey and satellite data', *Journal of the*

- Beaton, A. E. & Allen, N. L. (1992), 'Chapter 6: Interpreting scales through scale anchoring', *Journal of Educational Statistics* 17(2), 191-204.
- Chambers, R. & Tzavidis, N. (2006), 'M-quantile models for small area estimation', *Biometrika* pp. 255-268.
- Chambers, R., Chandra, H. & Tzavidis, N. (2009), 'On bias-robust mean squared error estimation for pseudo-linear small area estimators'.
- Cochran, W. G. (1977), *Sampling Techniques*: 3d Ed, Wiley.
- Cervantes, V. H., Cepeda, E. & Camargo, S. L. (2008), 'Una propuesta para la obtención de niveles de desempeño en los modelos de teoría de respuesta al ítem', *Avances en Medición* 6(1), 49-58.
- de Andrade, D. F., Tavares, H. R. & da Cunha Valle, R. (2000), 'Teoria da resposta ao item: conceitos e aplicacoes', ABE, Sao Paulo.
- Fay, R. E. & Herriot, R. A. (1979), 'Estimates of income for small places: an application of james-stein procedures to census data', *Journal of the American Statistical Association* 74(366a), 269-277.
- Falorsi, P. D. & Righi, P. (2008), 'A balanced sampling approach for multi-way stratification designs for small area estimation', *Survey Methodology* 34(2), 223-234.
- Gonzalez, M. E. (1973), Use and evaluation of synthetic estimates, in 'Proceedings of the social statistics section', pp. 33-36.
- Ghosh, M. & Rao, J. (1994), 'Small area estimation: an appraisal', *Statistical science* pp. 55-76.
- Gutiérrez, A. (2009), Estrategias de muestreo. Diseño de encuestas y estimación de parámetros, Hugo Andrés Gutiérrez Rojas.
- Gutiérrez, H. A. (2016), Estrategias de muestreo. Diseño de encuestas y estimación de parámetros.

• Hambleton, R. K., Swaminathan, H. & Brook, H. L. (1991), *Fundamentals of Item*

BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6

- Longford, N. T. (2005), 'On selection and composition in small area and mapping problems', *Statistical methods in medical research* 14(1), 3-16.
- Longford, N. T. (2004), 'Missing data and small area estimation in the uk labour force survey', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 167(2), 341-373.
- Little, R. J. (1993), 'Post-stratification: a modeler's perspective', *Journal of the American Statistical Association* 88(423), 1001-1012.
- Lord, F. M. (1980), *Applications of item response theory to practical testing problems*, Routledge.
- Lord, F. M. (1986), 'Maximum likelihood and bayesian parameter estimation in item response theory', *Journal of Educational Measurement* 23(2), 157-162.
- Mislevy, R., Beaton, A., Kaplan, B. & Sheehan, K. (1992), 'Estimating population characteristics from sparse matrix samples of item responses', *Journal of Educational Measurement* 20(2), 133-161.
- OECD (2009), Pisa 2006, Technical report, Organización para la Cooperación y Desarrollo Económicos.
- Pfeffermann, D. (2002), 'Small area estimation-new developments and directions', *International Statistical Review* 70(1), 125-143.
- Rao, J. (2003), 'Small area estimation', Wiley, New York.
- Pfeffermann, D. & Correa, S. (2012), 'Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation.', *Biometrika* 99(2).

BOGOTÁ • PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9. n.º 51-11 / www.usta.edu.co / contactenos@usantotomas.edu.co
VICERRECTORÍA GENERAL DE UNIVERSIDAD ABIERTA Y A DISTANCIA • Tel. (571) 595 00 00
Carrera 10. n.º 72-50 / www.ustadistancia.edu.co / admisiones@ustadistancia.edu.co
NIT: 860012357-6