Universidad Nacional de Colombia
Departamento de Estadística
Caso 01 - Consultoría Estadística

Estimación de las discontinuidades de la zona de Transición del Manto en el noroccidente de Suramérica a partir del análisis de la Función Receptora de onda P

- **Nombre del consultante**: Jorge Enrique Cubillos Gordillo.
- **Programa**: Maestría en Geofísica.
- **Facultad**: Ciencias.
- **Consultores**: Andrés Felipe Franco Valero, Luisa Fernanda Guantiva Vargas, Daniel Felipe Hernandez Montoya, Valeria Ramírez Sánchez y Anni Melissa Velandia Echeverry.
- **Fecha de inicio**: 22 de Abril de 2022
- **Palabras clave:** Estimador, boostrap, muestra, error estándar, estadística.

1. Introducción

El interés de la ciencia por los estudios geológicos de la estructura interna de la tierra ha llevado a grandes contribuciones sobre el funcionamiento interno del planeta. Determinar la estructura de la tierra no es una tarea fácil teniendo en cuenta el problema de inaccesibilidad de las zonas profundas, actualmente existen métodos directos e indirectos para determinar la estructura y composición de la tierra, pero una de sus principales desventajas es que este tipo de métodos tienen un alto costo, debido a que requieren perforaciones y/o adquisición sísmica profunda. Por esta razón se presenta una metodología llamada de función receptora, que ha dado resultados viables en diversas partes del mundo, esta técnica usa sismos con ciertas características donde se pueden realizar aproximaciones de primer orden de la estructura interna de la Tierra de una manera regional y cuyos resultados permiten inferir procesos geodinámicos de la zona estudiada.

2. Objetivos

- **Objetivo General**
  Estimar la profundidad de las discontinuidades sísmicas de 410 y 660 km de la corteza en el territorio nacional, a través de las funciones receptoras de ondas P entre los años de 1996 y 2021.

- **Objetivos Específicos**

  1. Por medio del método boostrap, encontrar los estimadores puntuales de las profundidades.
  2. Estimar la precisión de las estimaciones de las profundidades realizadas con el método boostrap.
  3. Realizar interpretaciones lógicas basadas en la literatura de las estimaciones boostrap.

3. Problema de consultoría estadística

Crear un algoritmo que genere las estimaciones puntuales de las profundidades máximas de las discontinuidades sísmicas de 410 y 660 kilómetros de profundidad que corresponden al inicio y al fin de la zona de transición del manto, la región de interés de esta investigación. El consultante desea responder a la pregunta: ¿qué tan preciso es el estimador de la profundidad media para las discontinuidades de 410 y 660 km? por lo tanto, se propone el método boostrap para responder a la pregunta y realizar la estimación deseada del parámetro, cabe resaltar que esta metodología ya ha sido probada en algunos trabajos previos.

4. Análisis de la información

El consultante suministró los datos para el análisis del estudio que fue recolectado por 98 estaciones sismológicas localizadas al rededor del territorio colombiano, que registraron sismos entre los años de 1996 y 2021. Los datos estudiados son formas de onda, a las cuales se les hace un procesamiento para calcular una función receptora, la cual es una serie de tiempo producto de la deconvolución entre una de las componentes horizontales de un sismograma y la componente vertical (Ligorría and Ammon, 1999). En total, se calcularon 66277 funciones receptoras entre las 98 estaciones, sin embargo, solo se usaron las que tenían buena calidad 15049.

4.1. Metodologías estadísticas utilizadas

Las funciones receptoras se seleccionan grupos según su ubicación espacial, cada una de grupos o ventanas tiene un número n_i de series de tiempo, donde i representa la i-ésima función receptora que está en la ventana, por lo que, para determinar las profundidades de las discontinuidades de 410 y 660, se deben apilar (superponer las series de tiempo que están en una ventana) estas n_i formas de onda en cada uno de los grupos y determinar las fases de onda correspondientes a las profundidades de las interfases sísmicas. Sin embargo, para la validez de este resultado es conveniente utilizar procedimientos estadísticos de remuestreo como lo es el Boostrap para poder hacer mejores estimaciones y obtener un error apropiado de estas.

4.1.1. Boostrap

El método boostrap es una métodología computacional que nos ayuda a responder la pregunta: ¿qué tan preciso es un estimador escogido de un parámetro particular θ?
A continuación se presenta la metodología Boostrap presentada por Efron and Tibshirani (1986), para estimar el error estándar (que es una estadística útil para medir la precisión estadística) y los intervalos de confianza del estimador puntual.
Supongamos que las observaciones $\mathbf{y} = (x_1, x_2, \cdots, x_n)$ consisten en muestras independientes e identicamente distribuidas tales que las observaciones $X_1, X_2, \cdots, X_n \sim_{\text{iid}} F$, donde F representa una distribución de probabilidad desconocida en $\mathcal{L}$, que es el espacio muestral común de las observaciones. Tenemos una estadística de interés, digamos $\hat{\theta}(\mathbf{y})$, a la que deseamos asignar un error estándar estimado. La estimación bootstrap del error estándar es

$$\hat{\sigma} = \sigma(\hat{F}),$$

donde $\hat{F}$ es la distribución empírica, que tiene probabilidad $1/n$ en cada punto de los datos observado x_i. El algoritmo de Monte Carlo nos ayuda a evaluar esta estimación de la siguiente manera:
Sea $\mathbf{y}^* = (x_1^*, x_2^*, \cdots, x_n^*)$ (*muestra boostrap*) una muestra aleatoria con remplazamiento de tamaño n de la muestra real $\{x_1, x_2, \cdots, x_n\}$, entonces el algoritmo procede en los siguientes tres pasos: (i) Genere independientemente B muestras aleatorias con remplazamiento, digamos $\mathbf{y}^*(1), \mathbf{y}^*(2), \cdots, \mathbf{y}^*(B)$; (ii) Para cada muestra boostrap $\mathbf{y}^*(b)$, evalúe la estadística de interés, digamos $\hat{\theta}^*(b) = \hat{\theta}\left(\mathbf{y}^*(b)\right)$, $b = 1, 2, \cdots, B$; y (iii) calcular la desviación estándar

de la muestra de los valores de $\hat{\theta}^*(b)$

$$\hat{\sigma}_B = \left(\frac{\sum_{b=1}^{B} \left\{ \hat{\theta}^*(b) - \hat{\theta}^*(\cdot) \right\}^2}{B-1} \right)^{1/2}, \tag{1}$$

$$\hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^{B} \hat{\theta}^*(b)}{B}.$$

Dado $B \to \infty, \hat{\sigma}_B$ se aproximará a $\hat{\sigma}' = \sigma(\hat{F})$, que es la estimación bootstrap del error estándar. Tener en cuenta que por lo general, un error estándar más grande se traducirá en una estimación menos precisa de la media de la población.
Además, se obtiene el intervalo de confianza del estimador con el método percentil. El intervalo del método percentil es solo el intervalo entre los percentiles $(100 \cdot \alpha)$ y $(100 \cdot (1-\alpha))$ de la distribución bootstrap de $\hat{\theta}^*$.

4.1.2. Procedimiento

Inicialmente, se realiza el algoritmo para una de las ventanas y posteriormente se replica el algoritmo en todas las demás.
Entonces se aplicó el algoritmo de Monte Carlo suponiendo que cada serie de tiempo en determinado grupo es una observación. Es decir, para la ventana j se tienen $(\mathbf{x}_1, \mathbf{x}_2, \cdots,$ $\mathbf{x}_{nj})$ funciones receptoras a las que se le quiere estimar la profundidad media de las discontinuidades sísmicas de 410 y 660 km. Por lo tanto;

(i) Se generaron B=1000 muestras aleatorias con remplazamiento tamaño n_j de las ondas recogidas, donde n_j es el número de ondas de la estación j.

(ii) Se apiló cada una de las muestras boostrap generada en el paso anterior, para obtener una sola serie de tiempo b y se determinaron discontinuidades sísmicas de 410 y 660km basadas en la mayor amplitud positiva para cada una de las discontinuidades.

(iii) Se extrajo la media de las B muestras, para obtener la estimación puntual boostrap del estadístico de interés y además se halló la estimación del error estándar boostrap con la fórmula (1). Y se obtienen los intervalos de confianza de la profundidad para $\alpha = 0.05$ del método percentil.

4.2. Resultados

Teniendo en cuenta que B es muy grande , se tiene que el error estándar estimado de θ, gracias a la estimación boostrap, $\hat{se}(\hat{\theta})$ es igual a $\hat{\sigma}_B$.

Por lo cual, después de realizar el procedimiento que se explicó anteriormente, con las ventanas seleccionadas correspondientes a las profundidades, se realiza un diagrama de dispersión donde el eje x representa la profundidad $d660$, el eje y la profundidad $d410$ y dentro del gráfico se presenta la correlación que existe entre ellas teniendo en cuenta los intervalos que se hallaron para cada profundidad.
Cada punto tiene las siguientes especificaciones obtenidas de acuerdo a la estación que representa :

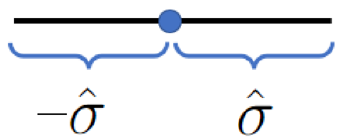

Figura 1: Intervalo de estimación boostrap para cada estación

Por lo cual el gráfico que recopila todas las ventanas de apilado es el siguiente:

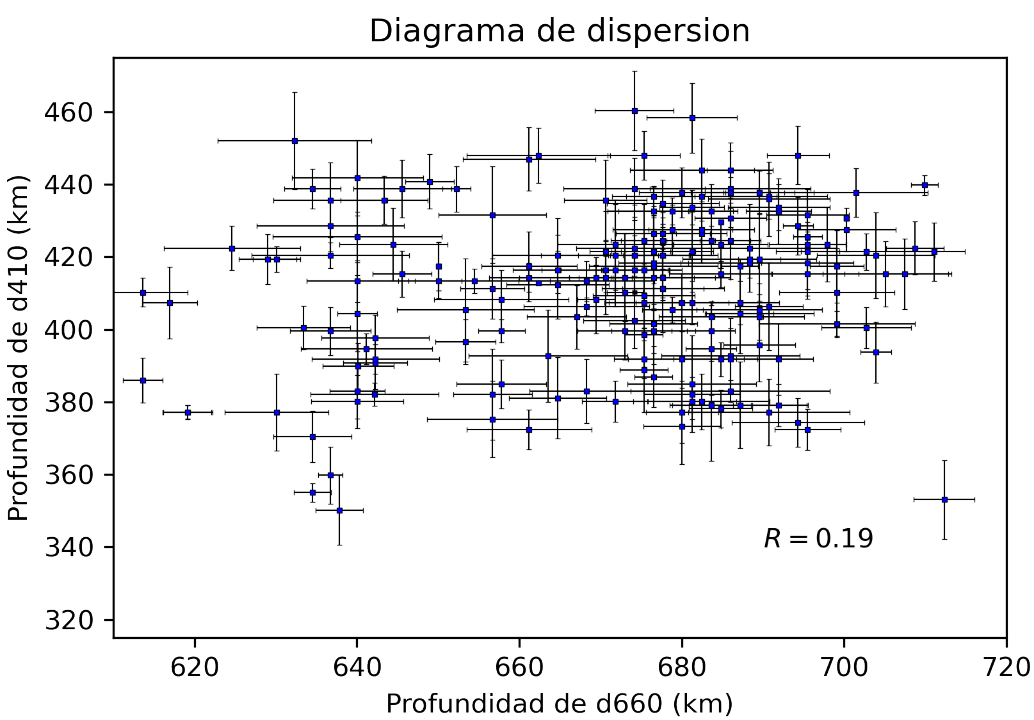

Figura 2: Gráfico de correlación cruzada entre las diferentes estaciones de d410 y d660. Las barras de error representan la desviación estándar.

El anexo 1 presenta los resultados para las discontinuidades de 410 y 660 km, las estimaciones puntuales, error estándar e intervalos de confianza de todas las estaciones del estudio.

5. Bitácora

- **Fecha de inicio**: Marzo 11 de 2022
- **Fecha de terminación**: -

	Fecha	Avances
1	Abril 22	El cliente solicita ayuda estadística para estimar la profundidad de las discontinuidades sísmicas de 410 y 660 km de la corteza en el territorio nacional, envió la presentación con la introducción, justificación, objetivos y artículos de referencia.
2	Mayo 13	El cliente se reunió con el grupo de consultoría en una videollamada por meet en donde aclaró dudas y nos brindo la información de una de las series de tiempo(trazas) para la cual se realizaría posteriormente el código en python para replicarlo para todas las series.
3	Mayo 20	Se entrega al cliente el código para la replicación del boostrap para cada una de las trazas y el cliente sugiere unos gráficos scatterplot para el análisis.
4	Mayo 23	El cliente nos envia el documento con su trabajo de grado y posteriormente se le solicita la tabla de resultados para las diferentes trazas.
4	Mayo 23	El cliente nos envia el documento con su trabajo de grado y posteriormente se le solicita la tabla de resultados para las diferentes trazas.
4	Junio 08	Se envía al cliente el el informe y el código final corregido con los últimos cambios.

Referencias

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, pages 54–75.

Ligorría, J. P. and Ammon, C. J. (1999). Iterative deconvolution and receiver-function estimation. *Bulletin of the seismological Society of America*, 89(5):1395–1400.