



# Módulo de datos - Proyecto de módulo

## Estudio de factores sociales que influyen en las pruebas Icfes usando analítica de datos

Oscar Altuve (USB), Victor Guzmán (UCV/UC)

26 de mayo de 2022

### Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Resultados y análisis</b>	<b>3</b>
2.1. Variables de interés . . . . .	3
2.2. Selección del espacio muestral . . . . .	6
2.3. Construcción de los procesos estocásticos . . . . .	7
2.4. Construcción de medidas para cada proceso estocástico . . . . .	8
2.5. Población total y muestra escogida . . . . .	8
<b>3. Discusión y Conclusiones</b>	<b>11</b>
3.1. Distribución geográfica de los estudiantes . . . . .	11
3.2. Tamaño de la muestra . . . . .	11
3.3. Acceso a internet y desempeño académico . . . . .	12
3.4. Estudios a realizar en un futuro . . . . .	12
3.4.1. Distribución geográfica . . . . .	12
3.4.2. Relación entre las pruebas Saber 11º y Saber Pro . . . . .	12
3.4.3. Asimetría de los histogramas . . . . .	13
<b>Referencias</b>	<b>14</b>

### Resumen

Los datos de las pruebas Saber 11<sup>o</sup> 2018 y 2019 y la prueba Saber Pro 2019, fueron analizados en términos del lugar de residencia de los estudiantes, su puntuación global en las pruebas y su acceso a internet. Se realizaron dos estudios, en el primero de ellos se hizo un sondeo de la cantidad y la relación de la puntuación global de los evaluados con su departamento de residencia, en este se observó que existe una relación entre la puntuación global y los departamentos de mayor competencia, por otro lado en el segundo estudio se enfocó en la relación de la variable cualitativa binaria de tener o no internet. Para este estudio se tomó una muestra representativa de 30000 estudiantes tomados de forma aleatoria para las 3 pruebas y se calculó la media y la desviación estándar del ajuste gaussiano asociado al histograma de la puntuación global de los estudiantes que tienen o no acceso a internet, obteniendo como resultado que para las pruebas Icfes aquellos con acceso a internet tienen una puntuación global media de 18,19 y 18,90 puntos más que sus pares sin acceso a internet, por otro lado del resultado de la prueba Saber Pro 2019 se observa una puntuación media de 11,28 puntos más que aquellos sin acceso a internet. Se observa que el acceso al internet es una variable considerable para el desempeño de los estudiantes en estas pruebas nacionales.

## 1. Introducción

Desde el año 2009, el Instituto Colombiano para Evaluación de la Educación (Icfes) tiene la misión de evaluar la formación que se ofrece a los distintos niveles del servicio educativo en el territorio nacional, mediante exámenes externos y estandarizados, denominados "Saber". Su propósito es valorar la calidad de la educación y brindar información útil a los evaluados en términos del desarrollo de competencias. Por otra parte, se dispone de la revisión de estrategias educativas y programas académicos, para la toma de decisiones de políticas públicas, así como para los órganos encargados de garantizar este derecho.

En este sentido, el examen Saber 11<sup>o</sup>, es presentado por estudiantes que se encuentren finalizando el grado undécimo, con el fin de obtener resultados oficiales para el ingreso a la educación superior. También, pueden presentarlo quienes ya hayan obtenido el título de bachiller o hayan superado el examen de validación del bachillerato, de conformidad con las disposiciones vigentes. Su estructura comprende la presentación de cinco pruebas genéricas sobre lectura crítica, matemática, sociales y ciudadanas, ciencias naturales e inglés, además de un cuestionario socio-económico que permite obtener información sobre los procesos de enseñanza y aprendizaje [1].

Por otro lado, la prueba Saber Pro, está dirigida de manera obligatoria a estudiantes que han aprobado 75 % de los créditos de sus respectivos programas de formación universitaria profesional. Como también, puede ser presentada de forma independiente y voluntaria por aquella persona que haya finalizado ya sus estudios de pregrado. Se estructura en dos sesiones, una en la que los inscritos presentan un examen genérico sin importar su área de formación que evalúan competencias de comunicación escrita, razonamiento cualitativo, lectura crítica, competencias ciudadanas e inglés, como también, responden un cuestionario socio-económico. En la otra sesión, los estudiantes presentan

conjuntos de exámenes específicos de acuerdo a su programa universitario [2].

En éste proyecto, se hará un análisis de los datasets resultantes de las pruebas Saber 11<sup>o</sup> 2018 y 2019, y Saber Pro 2019 en búsqueda de factores socio-económicos que influyen en los puntajes globales de los evaluados y su evolución en el tiempo.

## 2. Resultados y análisis

En este estudio se emplean los resultados públicos anonimizados de las prueba Saber 11<sup>o</sup> y Saber Pro recolectados en los años 2018-2019 [3, 4] y 2019 [5] respectivamente, de la evaluación de competencias genéricas y el cuestionario socio-económico.

Estas pruebas proveen información que permiten estimar cuáles son las diferencias en la calidad educativa y qué factores influyen en una mayor o menor competencia; estas abarcan no sólo el rendimiento educativo sino también el contexto social, cultural e histórico en el que los evaluados se desenvuelven.

Dado el abanico de posibilidades de estudio, se realizó un sondeo estadístico empleando la librería "Pandas Profiling" para identificar las variables con mayor correlación y con esto delimitar el estudio. En análisis mencionado y todos los análisis a continuación se efectuaron en Jupyter Notebook con kernel Python y está alojado en un repositorio de G-Lab [6].

### 2.1. Variables de interés

Entre la documentación publicada por Icfes, se facilita un diccionario de variables de la prueba [7], en total se tomaron dos años diferentes de la prueba Saber 11<sup>o</sup>, las variables estudiadas fueron aquellas que presentaron un grado de correlación mayor, el puntaje global y la relación con tener o no tener internet.

Un primer estudio que se consideró de importancia, fue conocer la cantidad de evaluados de acuerdo al departamento de residencia, esto se muestra en la figura 1, dado que permite tomar una idea de las regiones geográficas donde hay mayor influencia educativa y permite generar hipótesis del impacto de esta en las distintas competencias.

Como se muestra, la mayor muestra de evaluados en ambas pruebas se encuentra en los departamentos más competitivos de Colombia, entre ellos Bogotá, Antioquia y Santander, de acuerdo con [8] para el año 2019. Por su parte, Bogotá representa la mayor cantidad de evaluados, superando por miles a los demás departamentos.

En las representaciones de la figura 2 se emplearon los datasets de Saber 11<sup>o</sup> y Saber Pro en el año 2019, se observa en general que las medianas de las muestras se encuentran en el centro de las cajas, indicando que las distribuciones son simétricas. Por otra parte, las muestras de la prueba Saber

Tabla 1: Diccionario de variables empleadas en el estudio, de acuerdo a la prueba Saber 11° 2018-2019 y Saber Pro 2019. [7]

Campo	Descripción del campo	Opciones de respuesta
Información de contacto		
ESTU_DEPTO_RESIDE	Departamento de residencia	Texto
ESTU_COD_RESIDE_DEPTO	Código Dane del departamento de residencia	Númerica [99999 extranjero]
Información socioeconómica		
FAMI_TIENEINTERNET	¿Su hogar cuenta con servicio o conexión a internet?	Si No
Resultados		
PUNT_GLOBAL	Puntaje total obtenido	Númerica - Rango [0, 300]

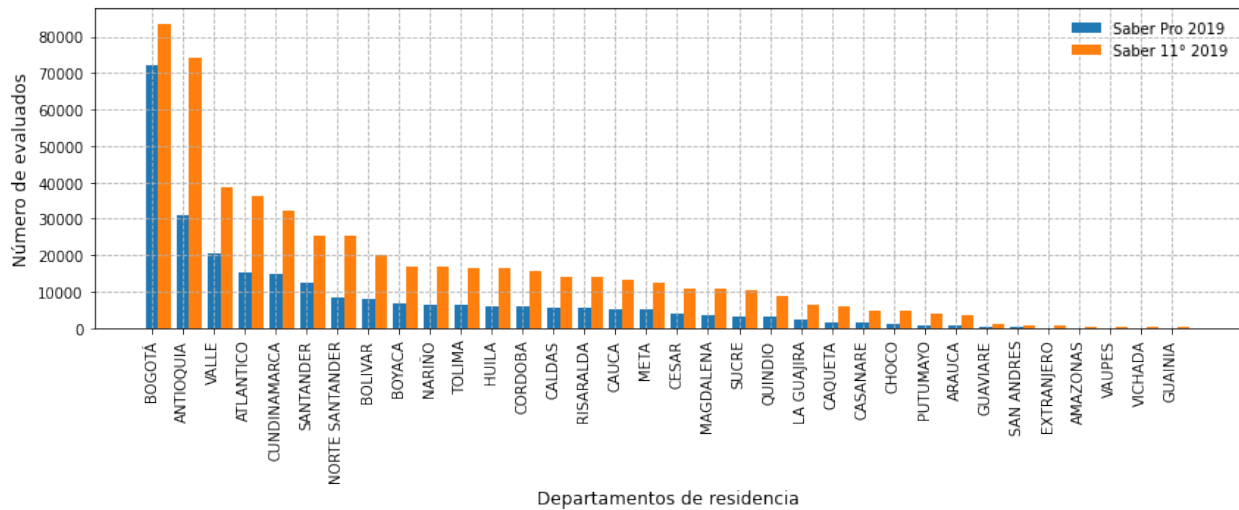
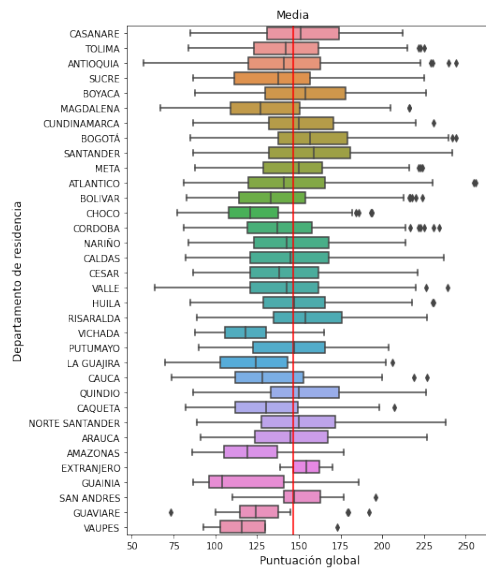


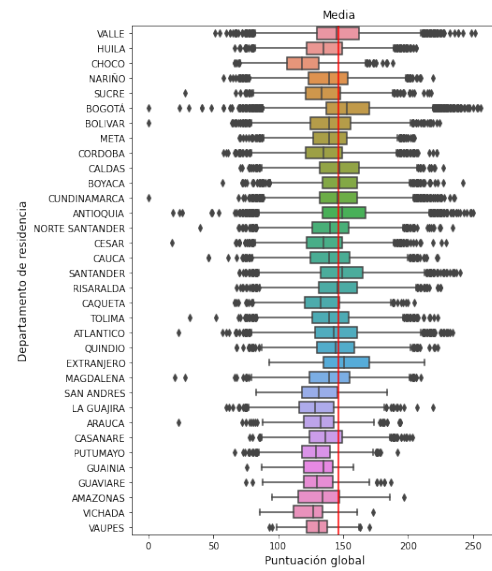
Figura 1: Muestra de estudiantes que presentaron las pruebas Saber 11° y Saber Pro en el año 2019 en función del departamento de residencia.

Pro tiene menor dispersión de los datos que la prueba Saber 11°, cabe resaltar que pudiera estar relacionado a la cantidad de muestras consideradas, esto es 260755 y 10000<sup>1</sup> respectivamente. Si damos atención a los valores atípicos, representados por los puntos negros, se observa en la prueba Saber Pro situaciones en que los evaluados tienen puntuaciones globales cero, no hay puntuaciones en 300 puntos. Para el caso de Saber 11°, la mayor parte de los valores atípicos son superiores a 200 puntos.

<sup>1</sup>El dataset de la prueba Saber 11° 2019, cuenta con 546212 muestras por variable. Por limitantes de computo no se analizaron todas las muestras. Las muestras fueron seleccionadas aleatoriamente.



(a) Saber 11°.



(b) Saber Pro.

Figura 2: Departamento de residencia en función de la puntuación global de la pruebas Saber en el año 2019.

Para la siguiente representación, dada la generalidad de la mediana, pudiera permitirse representar cada departamento por la media de la muestra. En la figura 3 se observa la media de las puntuaciones globales en cada departamento.

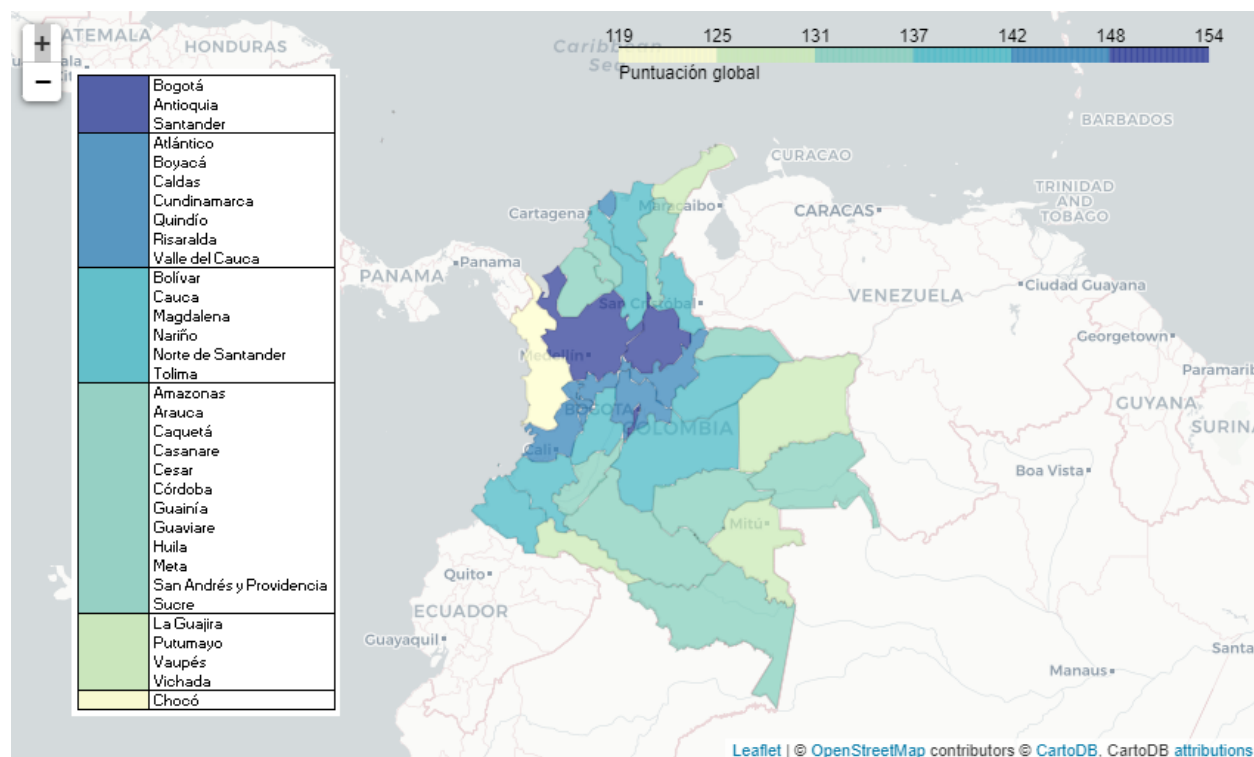


Figura 3: Promedio de la puntuación global de los evaluados de acuerdo al departamento de residencia.

Ya que las dos pruebas Saber 11<sup>o</sup> consideradas en el estudio fueron tomadas en años diferentes, las variables cada año fueron medidas de formas diferentes. Una de ellas estaba en el rango  $[0, 500]$  y la más reciente entre  $[0, 300]$ , con la finalidad de facilitar el análisis se reescaló el puntaje global promedio a la escala  $[0, 300]$ .

Adicionalmente existían algunos valores anómalos en el puntaje global, anómalos en el sentido de que podían tener un valor negativo y del orden magnitud de  $10^{18}$ . Estos valores fueron excluidos del análisis ya que la media y la desviación estándar son sensibles a valores anómalos como estos, pero adicionalmente porque un valor promedio negativo no tiene sentido por definición.

El criterio usado para la limpieza de estos datos anómalos fue:  $x - \mu_{\text{muestra}} \leq 10\sigma_{\text{muestra}}$ , donde  $x$  es el dato y  $\mu$  y  $\sigma$  son la media y la desviación estándar de la muestra.

## 2.2. Selección del espacio muestral

La cantidad de estudiantes en las tres pruebas es de 494844 para la prueba Saber 2018, 545816 para la prueba Saber 2019 y 258731 para la prueba Saber Pro 2019, las pruebas las presentaron estudiantes de los 32 departamentos de Colombia y del distrito capital. Para el análisis se tomó una

muestra de toda la población, la misma es tomada de forma aleatoria. Para identificar si el tamaño fue *suficientemente grande*, se realizó el siguiente análisis.

Sea  $\Omega$  el espacio muestral, en este caso el conjunto total de estudiantes que asistieron a alguna prueba nacional en particular, este conjunto está dotado con una noción de ocurrencia  $\mathcal{I}$ [9], tal que el mismo puede ser construido por medio de un conjunto finito de subconjuntos disjuntos  $\Omega_i$ . Es decir:

$$\Omega_i \cap \Omega_j = \emptyset, \text{ con } i, j \in \mathcal{I}, \quad \Omega = \bigcup_{i \in \mathcal{I}} \Omega_i \quad (1)$$

El cardinal es una medida que se le puede definir a este conjunto  $|\Omega|$ , es decir  $|\Omega|$  representa el número de elementos del conjunto. De la expresión matemática de arriba se sigue directamente que:

$$|\Omega| = \left| \bigcup_{i \in \mathcal{I}} \Omega_i \right| = \sum_{i \in \mathcal{I}} |\Omega_i| \quad (2)$$

Acá seccionaremos el espacio muestral bajo la noción de ocurrencia de *tener o no tener internet*, Entonces los estudiantes podrán ser clasificados teniendo o no teniendo internet y  $\Omega = \Omega_{\text{si}} \cup \Omega_{\text{no}}$ , nótese que al tomar un subconjunto arbitrario de  $\Omega$ , llámese  $\Gamma \subset \Omega$ , este heredaré la noción de ocurrencia de  $\Omega$ ,  $\Gamma = \Gamma_{\text{si}} \cup \Gamma_{\text{no}}$ .

### 2.3. Construcción de los procesos estocásticos

La intención que se tiene acá, es tomar un subconjunto del espacio muestral total, bajo la condición de que el número de elementos del conjunto ha de ser igual a  $|\Omega_i| = N_i$ , para poder identificar si este conjunto es representativo del espacio muestral total, se hará el siguiente procedimiento:

- Cálculo de la media  $\mu$  y de la desviación estándar  $\sigma$  del puntaje promedio total por medio de un ajuste gaussiano para todo el espacio muestral  $\Omega$ , nótese que la definición de una medida de probabilidad sobre el espacio muestral define de forma implícita un espacio de probabilidad.
- Construcción de un subconjunto  $\Omega_i$  de  $N_i$  elementos  $|\Omega_i| = N_i$ , los elementos de este conjunto serán tomados del espacio muestral  $\Omega$  forma aleatoria.
- Generación de  $n$  subconjuntos  $\Omega_i$  de tamaño  $N_i$ .
- Para cada subconjunto  $\Omega_i$  se le calcula las medidas de probabilidad de forma implícita y se obtiene la media y la desviación estándar, estas medidas son propias de cada subconjunto  $\Omega_i$  y representan una variable aleatoria.
- La colección parametrizada de variables aleatorias es lo que formalmente se conoce como un proceso estocástico, por ejemplo se obtendrá un proceso aleatorio de medias  $\{\mu_i\}_{i \in \mathcal{I}}$  y un proceso aleatorio de desviaciones estándar  $\{\sigma_i\}_{i \in \mathcal{I}}$ .

## 2.4. Construcción de medidas para cada proceso estocástico

Para cada proceso estocástico se calculará la media aritmética:

$$\mu_\mu = \frac{1}{n} \sum_{i=1}^n \mu_i, \quad \mu_\sigma = \frac{1}{n} \sum_{i=1}^n \sigma_i \quad (3)$$

Por medio del cálculo del error porcentual se establecerá una referencia de si el tamaño de la muestra escogido es *suficientemente grande* para poder representar a la muestra como un todo. Al comparar el error porcentual de  $\mu_\mu$  y  $\mu_\sigma$  con el  $\mu$  y  $\sigma$  obtenido de la muestra total tendremos una medida para verificar si la muestra obtenida es representativa del conjunto global. Los errores porcentuales serán calculados como:

$$\epsilon_\mu = \frac{\mu - \mu_\mu}{\mu} \times 100 \%, \quad \epsilon_\sigma = \frac{\sigma - \mu_\sigma}{\sigma} \times 100 \% \quad (4)$$

## 2.5. Población total y muestra escogida

Para cada una de las pruebas se realizó un histograma del puntaje global bajo la noción de tener (azul) o no tener (anaranjado) internet. En la figura 4 se puede identificar gráficamente como los estudiantes con acceso a internet en cualquiera de las tres pruebas han de tener un puntaje global mayor que aquellos estudiantes sin acceso a internet.

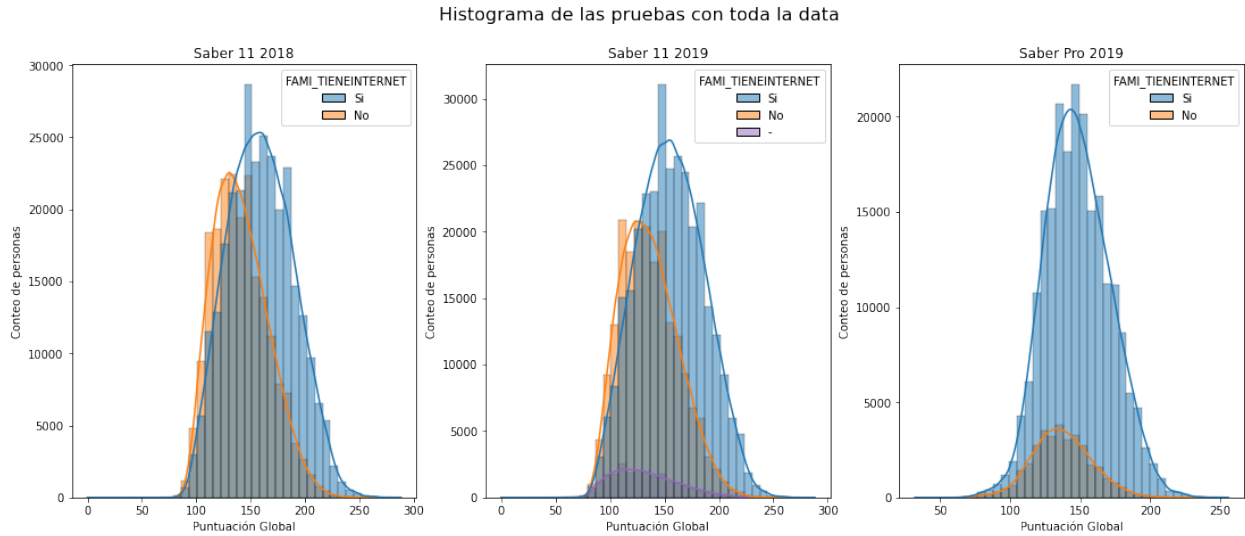


Figura 4: Histograma del puntaje global para toda la población

Nótese como el comportamiento de las distribuciones se asemeja al de una gaussiana, en el sentido de estar localizada alrededor de un valor. Debido a esta similitud se hace un ajuste gaussiano a la

data, es decir implícitamente además se asume que la muestra es simétrica.

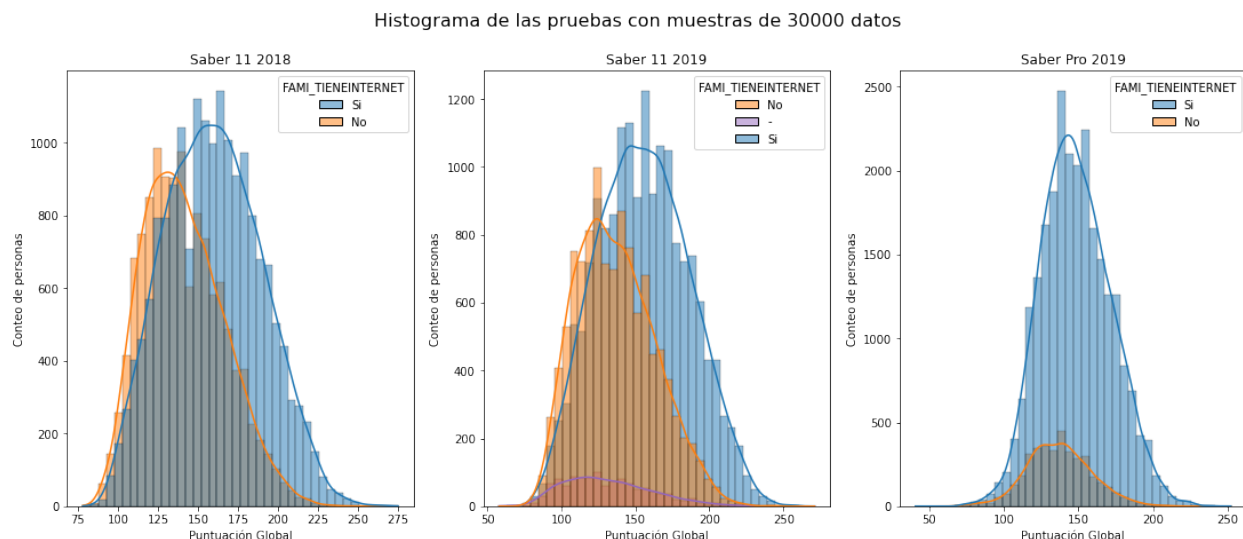


Figura 5: Histograma del puntaje global para la muestra de 30000 personas.

En la figura 5, se puede ver que para una muestra de 30000 personas tomadas de forma aleatoria aún aparece el comportamiento gaussiano. En la imagen la línea continua representa en KDE o **K**ernel **D**ensity **E**stimation arrojado directamente por seaborn para tener una idea del comportamiento de la distribución subyacente.

Con la finalidad de poder corroborar si efectivamente esta muestra es representativa de toda la población, se realizó este procedimiento 100 veces realizando un ajuste gaussiano en cada prueba al histograma de datos de personas con y sin internet, es decir de cada corrida se obtuvieron 12 datos, 4 para cada prueba, correspondientes a: las medias con y sin internet y a las desviaciones estándar con y sin internet de las muestras.

El resultado de las 100 corridas se puede ver en la imagen 6:

De acá se puede calcular el error porcentual  $\epsilon$  en relación a la población total, los valores obtenidos se pueden ver en la tabla 2, donde se observa como este error es menor al 0,007%. Así que una muestra de 30000 estudiantes es suficientemente grande como para arrojar resultados similares a los que arrojaría el análisis con la población total.

Además se ve que donde existe una mayor dispersión de valores es para las medidas de los estudiantes que no tienen internet en la prueba Saber Pro 2019, esto es debido a que la cantidad de estudiantes que no tienen internet acá es pequeña en comparación a las demás pruebas, como se puede ver en 4 y 5.

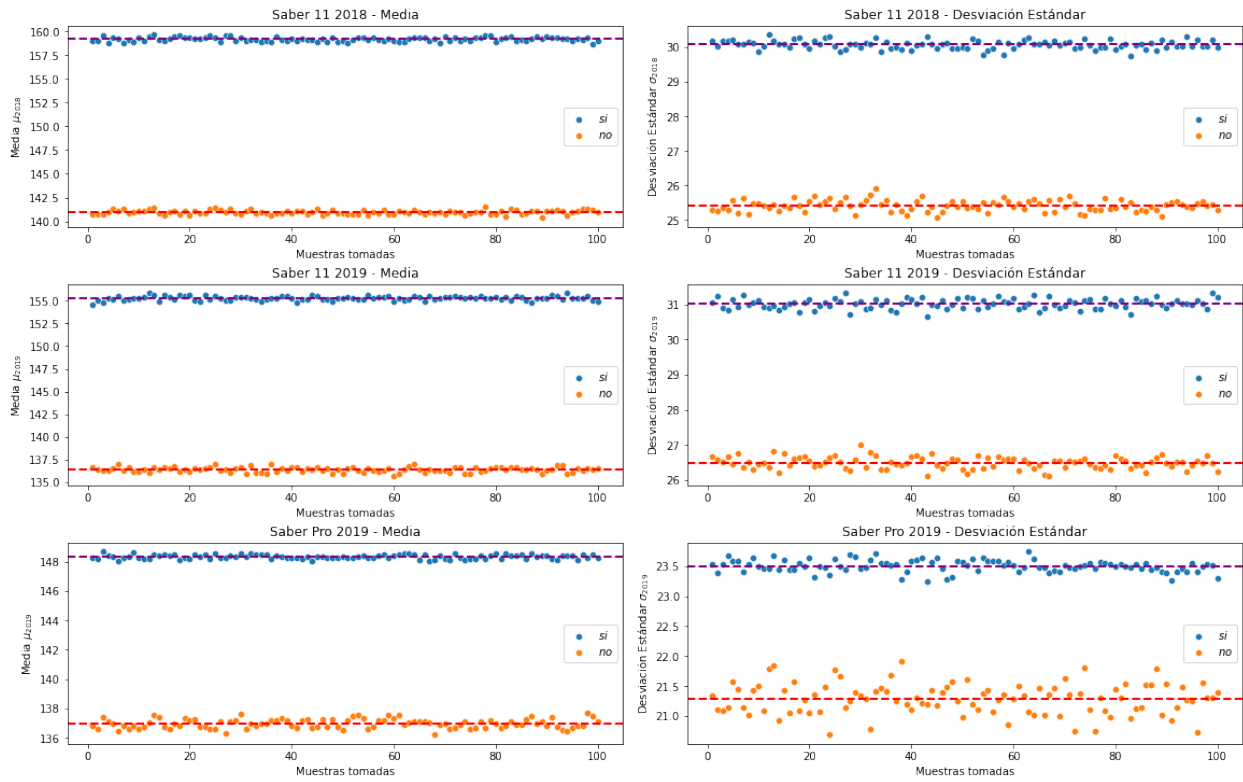


Figura 6: Valor de las medias  $\mu_i$  y las desviaciones estándar  $\sigma_i$  para cada corrida, las líneas punteadas marcan la media aritmética para cada proceso estocástico.

Tabla 2: Error porcentual  $\epsilon$  de la media tomada de la muestra para cada prueba y dependiendo de si la persona tiene o no internet.

Prueba	$\epsilon_{si}$ (%)	$\epsilon_{no}$ (%)
Saber 11° 2018	0.0050	0.0038
Saber 11° 2019	0.0005	0.0046
Saber Pro 2019	0.0063	0.0441

Otro análisis de interés que se puede realizar es el cálculo de la diferencia de las medias para las pruebas, nótese que:

$$\begin{aligned}\Delta\mu_{2018} &= \mu_{\text{si},2018} - \mu_{\text{no},2018} = 159,17 - 140,98 = 18,19 \\ \Delta\mu_{2019} &= \mu_{\text{si},2019} - \mu_{\text{no},2019} = 155,31 - 136,41 = 18,90 \\ \Delta\mu_{\text{pro},2019} &= \mu_{\text{si pro},2019} - \mu_{\text{no pro},2019} = 148,36 - 137,08 = 11,28\end{aligned}\tag{5}$$

en las pruebas Saber 11<sup>o</sup> 2018 y 2019 los estudiantes que si tienen acceso a internet obtienen aproximadamente 18,19 y 18,90 puntos mas en comparación a quienes no tienen acceso a internet, respectivamente y además que esta brecha entre quienes tienen internet y no, aún existe para aquellos estudiantes que están próximos a terminar su carrera, teniendo una media de 11,28 puntos más para aquellos que tienen acceso a internet. Algo que puede ser comentado al respecto es que al parecer la brecha de puntajes es menor en aquellos estudiantes próximos a graduarse.

### 3. Discusión y Conclusiones

#### 3.1. Distribución geográfica de los estudiantes

Los departamentos de residencia de los evaluados con mayor participación de las pruebas Saber 11<sup>o</sup> y Saber Pro 2019, fueron Bogotá, Antioquia, Valle, Atlántico, Cundinamarca, que son coincidentes con los departamentos más competitivos de Colombia. Además, la mayor cantidad de evaluados se observó en la capital del país, y supera por miles a los otros departamentos.

Al estudiar la relación de la puntuación global y los departamentos de residencia, mostró de manera general que los datos de ambas pruebas no están sesgados, por lo que emplear la media de las puntuaciones globales pudiera ser un buen estimador, que como se observa en la figura 3, hay correspondencia con los resultados mostrados en la figura 2.

Como se pudo ver en la figura 1 una gran parte de los estudiantes se encuentran residenciados en Bogotá, de manera que al momento de hacer un análisis sobre una muestra aleatoria o sobre la población total, muchos de los estudiantes que se tomarán serán de Bogotá. Esto es importante considerarlo en el caso en que uno quiera conocer el desempeño de los estudiantes en zonas como Meta, Sucre, Choco y Arauca por ejemplo que tienen una población menor que presentó la prueba.

#### 3.2. Tamaño de la muestra

Así como se pudo ver en la figura 6 y por medio de la tabla 2 una muestra de 30000 personas contiene estadísticamente la misma información que la muestra total. Nótese que los resultados del desempeño académico de los estudiantes no dependen de como ha sido seleccionada la muestra, ya que fue seleccionada de forma aleatoria. Es decir los estudiantes seleccionados son de múltiples estratos, localidades variadas, con distintas horas dedicadas al estudio y de igual manera se manifiesta el mismo comportamiento dependiendo de si tienen o no internet.

### 3.3. Acceso a internet y desempeño académico

Como se pudo observar de la figura 5 y de las ecuaciones 5, si se toman 30000 estudiantes de forma aleatoria que presentaron las pruebas Saber 11<sup>o</sup> 2018 y 2019 y la prueba Saber Pro 2019, se espera que los estudiantes con acceso a internet tenga un desempeño mayor en relación a la puntuación global.

Otro detalle interesante que se pudo observar es que en la prueba Saber Pro 2019, la cantidad de estudiantes sin acceso a internet es considerablemente menor en comparación a los casos de las pruebas Saber 11<sup>o</sup>, esto pudiese ser interpretado de múltiples maneras. Aquí se señalan algunas.

- Los estudiantes sin acceso a internet que ingresaron a la universidad no lograron avanzar hasta el 75 % de la carrera.
- Los estudiantes sin internet ingresan en menor medidas a las universidades.
- Los estudiantes que ingresan sin internet a las universidades logran adquirir internet en el proceso.

En la sección siguiente se sugieren algunas formas en las que se pudiese aclarar el por qué hay tan pocos estudiantes sin internet en la universidad.

### 3.4. Estudios a realizar en un futuro

#### 3.4.1. Distribución geográfica

Se pudiera recomendar, realizar ajustes de funciones de densidad de probabilidad PDF's sobre las puntuaciones globales para corroborar los resultados en los diagramas de caja. Como también, relacionar los histogramas sesgados con variables socio-económicas.

#### 3.4.2. Relación entre las pruebas Saber 11<sup>o</sup> y Saber Pro

Una manera de complementar el estudio sobre las pruebas Saber Pro, pudiese ser estudiando las pruebas Saber Pro de un año en relación a las pruebas Saber 11<sup>o</sup> realizadas unos 3 o 4 años antes. Para poder establecer de alguna manera la relación entre los estudiantes antes de ingresar a la universidad con los estudiantes que ya han ingresado y están próximos a graduarse.

Este estudio podría ser complementado con otras variables como la cantidad de estudiantes que reprueban por semestre/año, cuántos abandonan la carrera, a qué altura abandonan la carrera o cuántos ingresan en las carreras habiendo presentado la prueba Saber 11<sup>o</sup>. estas variables pudiesen ayudar a hacerle un *tracking* o seguimiento a los estudiantes, para así poder establecer métricas sobre la evolución del desempeño de los mismos.

Finalmente en relación a este tópico parece interesante señalar que la brecha que separa la puntuación media de estudiantes con o sin acceso a internet se reduce para la prueba Saber Pro 2019, esto

pudiéndose ser un indicador de que los estudiantes que han logrado avanzar en la carrera sin acceso a internet han conseguido medios y herramientas que le permiten mejorar su desempeño. Tomando como referencia la brecha en las pruebas Saber 11<sup>o</sup> 2018 y 2019.

### **3.4.3. Asimetría de los histogramas**

Los ajustes realizados a los histogramas para el análisis en términos de tener o no internet, fueron gaussianos, aunque se observa una cierta asimetría en los histogramas, se sugiere para futuros estudios un análisis que busque cuantificar esta asimetría, posiblemente usando alguna otra distribución que sea asimétrica o incorporar al análisis ya realizado un estudio adicional por cuartiles y analizando el comportamiento de la mediana para los diferentes conjuntos de datos.

## Referencias

- [1] ICFES, “Guía de orientación saber 11<sup>o</sup>,” 2022. [Online]. Available: <https://eservicioseducativos.com/wp-content/uploads/2022/02/Guia-de-orientacion-Saber-11-2022-1.pdf>
- [2] —, “Guía de orientación saber pro: Módulos de competencias genéricas,” jul 2019. [Online]. Available: <https://www.studocu.com/co/document/universidad-de-san-buenaventura/introduccion-a-las-relaciones-internacionales/guia-de-orientacion-modulos-de-competencias-genericas-saber-pro-2019/14138800>
- [3] J. Mantilla, H. Gonzalez, and L. Jaimes, “Estudio de factores sociales determinantes para los resultados de las pruebas icfes saber 11 2018,” 2021. [Online]. Available: <https://github.com/juandmantilla/Estudio-de-Factores-Sociales-Icfes>
- [4] ICFES, “Saber 11° 2019-2,” dic 2019. [Online]. Available: <https://www.datos.gov.co/Educaci%C3%B3n/Saber-11-2019-2/ynam-yc42>
- [5] —, “Resultados saber pro competencias genericas 2019-2,” feb 2020. [Online]. Available: <https://www.datos.gov.co/Educaci-n/Resultados-Saber-Pro-Competencias-Genericas-2019-2/2x55-9wxm>
- [6] O. A. Altuve and V. Guzmán, “Estudio de factores sociales que influyen en las pruebas icfes usando analítica de datos,” 2022. [Online]. Available: <https://gitmilab.redclara.net/altuveo/estudio-de-factores-sociales-que-influyen-en-las-pruebas-icfes-usando-analitica-de-datos>
- [7] ICFES, “Diccionario de variables saber-pro periodo 2012-2019.” [Online]. Available: <https://fddocuments.es/document/diccionario-de-variables-saber-pro-periodo-2012-2019-2020-05-27-1-diccionario.html>
- [8] Portafolio, “Los departamentos más competitivos del país,” Nov 2019. [Online]. Available: <https://www.portafolio.co/economia/gobierno/los-departamentos-mas-competitivos-del-pais-535920>
- [9] B. Øksendal, *Stochastic differential equations: an introduction with applications*, 6th ed., ser. Universitext. Springer, 2010.