

Trabajo Número 1

Introducción al Análisis de Datos Categóricos

Docente: Juan Carlos Correa Morales

Kleider Stiven Vásquez Gómez

Universidad Nacional de Colombia

Sede Medellín

2022

Índice

Introducción	1
Objetivo General	1
Objetivos Específicos	1
Marco Muestral	2
ICFES	2
Pruebas Saber 11°	2
Recolección de los datos	2
Creación de Variable Riqueza por estudiante	3
Análisis de correspondencias múltiples	3
Gráfica de sedimentación	4
Gráfico de Contribuciones de las variables en la primera dimensión	5
Gráfica de sedimentación actualizada	6
Gráfico de barras de Riqueza	7
Creación de Variable Riqueza por colegio	7
Análisis por departamento	8
Tabla Resumen de Puntaje de Colegios Oficiales y No Oficiales con respecto a su riqueza, por Departamento.	8
Gráfico de Distribución de Riqueza por colegio	9
Gráfico para la variable del resultado por colegio.	10
Boxplot para el Puntaje del estudiante en función de la riqueza del colegio .	11
Boxplot para el Puntaje del estudiante en función de la naturaleza del colegio	12
Boxplot para el Puntaje del estudiante en función de jornada del colegio . . .	13
Tablas de Contingencia del Nivel de Riqueza vs. Tipo de jornada con respecto al puntaje en las Pruebas Saber 11	14
Nivel de riqueza 5 vs Jornada de la Mañana:	14
Nivel de riqueza 8 vs Jornada de la Mañana:	15
Nivel de riqueza 5 vs Jornada de la Noche:	15

Nivel de riqueza 10 vs Jornada de la Noche:	16
Nivel de riqueza 5 vs Jornada de la Tarde:	16
Nivel de riqueza 1 vs Jornada Sabatina:	17
Nivel de riqueza 4 vs Jornada Sabatina:	18
Nivel de riqueza 10 vs Jornada Sabatina:	18
Conclusiones Generales	19
Comentarios	19
Anexos	20
Bibliografía	20

Introducción

El presente trabajo es un estudio estadístico que se desarrolla con la finalidad de analizar el nivel de riqueza de los colegios y estudiantes, mediante la información que se posee de los resultados ICFES (Saber 11) del 2018-I, los cuales se encuentran alojados en los datos abiertos de la página oficial del Gobierno de Colombia (<https://www.datos.gov.co/>).

Por tanto, por medio de este estudio, se pretende demostrar la relevancia que tiene el estudio estadístico desde el planteamiento del problema, la recolección de la información y análisis de los datos, hasta el momento de la evaluación y aplicación del buen uso de las metodologías estadísticas, ya que se quiere demostrar que con un debido manejo de los datos en función de la información que se tiene disponible de dichas Pruebas, se pueden obtener inferencias y llegar a obtener resultados eficientes para la resolución de dicha problemática planteada.

Objetivo General

Crear una nueva variable que represente el nivel de riqueza, tanto del estudiante como la correspondiente a su colegio, y una vez obtenida ésta hacer inferencia, mediante diversas técnicas estadísticas.

Objetivos Específicos

- Analizar la información obtenida de la página de datos abiertos del Gobierno de Colombia.
- Formular una interpretación de los resultados dados.
- Aplicar las técnicas y metodologías estadísticas más adecuadas para la resolución del caso de estudio.
- Utilizar el software estadístico R, y R-Studio para presentar de manera resumida la información mediante gráficos y tablas.
- Poner en práctica los conocimientos afianzados de Datos Categóricos, y las de otras ramas de la estadística para el desarrollo de dicho caso de estudio.

Marco Muestral

ICFES

El Instituto Colombiano para el Fomento y Evaluación de la Educación Superior (ICFES), es una entidad autónoma vinculada al Ministerio de Educación Nacional de Colombia. Este organismo es el encargado de promover la educación superior en Colombia.

Por consiguiente, ofrece diversas actividades de evaluación de la educación en todos sus niveles y así mismo apoya al Ministerio de Educación en la realización de los exámenes de Estado. También, realiza investigaciones sobre los factores que inciden en la calidad educativa, con el fin de ofrecer información para mejorarla.

Pruebas Saber 11°

Es un examen que evalúa el grado de desarrollo de las competencias académicas de los estudiantes que están por finalizar el grado 11 de la educación media en Colombia. Consta de dos sesiones (de 4 horas y 30 minutos de duración cada sesión), en las que se realizan pruebas de las siguientes competencias:

- Matemáticas. (50 preguntas en total)
- Lectura Crítica. (41 preguntas en total)
- Sociales y Competencias Ciudadanas. (50 preguntas en total)
- Ciencias Naturales. (58 preguntas en total)
- Inglés. (55 preguntas en total)
- Cuestionario Socioeconómico. (24 preguntas en total)

Con un total de 278 preguntas. (131 preguntas en la primera sesión y 147 preguntas en la segunda sesión).

Recolección de los datos

La información con la que se va a trabajar este caso de estudio será la hallada en los Resultados Saber 11 2018-1, que se encuentra alojada en la página de datos abiertos oficial del Gobierno de Colombia: <https://www.datos.gov.co/Educaci-n/Resultados-Saber-11-2018-1-Refinado/ptck-fi3s> La cual tiene una fecha de actualización del 17 de octubre de 2019 y no supera aún las 300 descargas del archivo.

Creación de Variable Riqueza por estudiante

La base de datos con la cual se va a trabajar consta de 19798 observaciones de 87 variables.

Como en la base de datos se puede hallar la variable ‘punt_global’, que representa la puntuación global de la prueba por cada estudiante, se crea una nueva variable categórica con 5 niveles (0,100], (100, 200], (200, 300], (300, 400], (400, 500]. Donde el rango (400, 500] representa a los estudiantes con mejores resultados en dicha prueba.

Luego, se busca crear la variable que indique la riqueza del estudiante, y es por eso que se agrupan ciertas variables con relación a los servicios que poseen en su hogar y sus horas dedicadas a lectura, internet y trabajo.

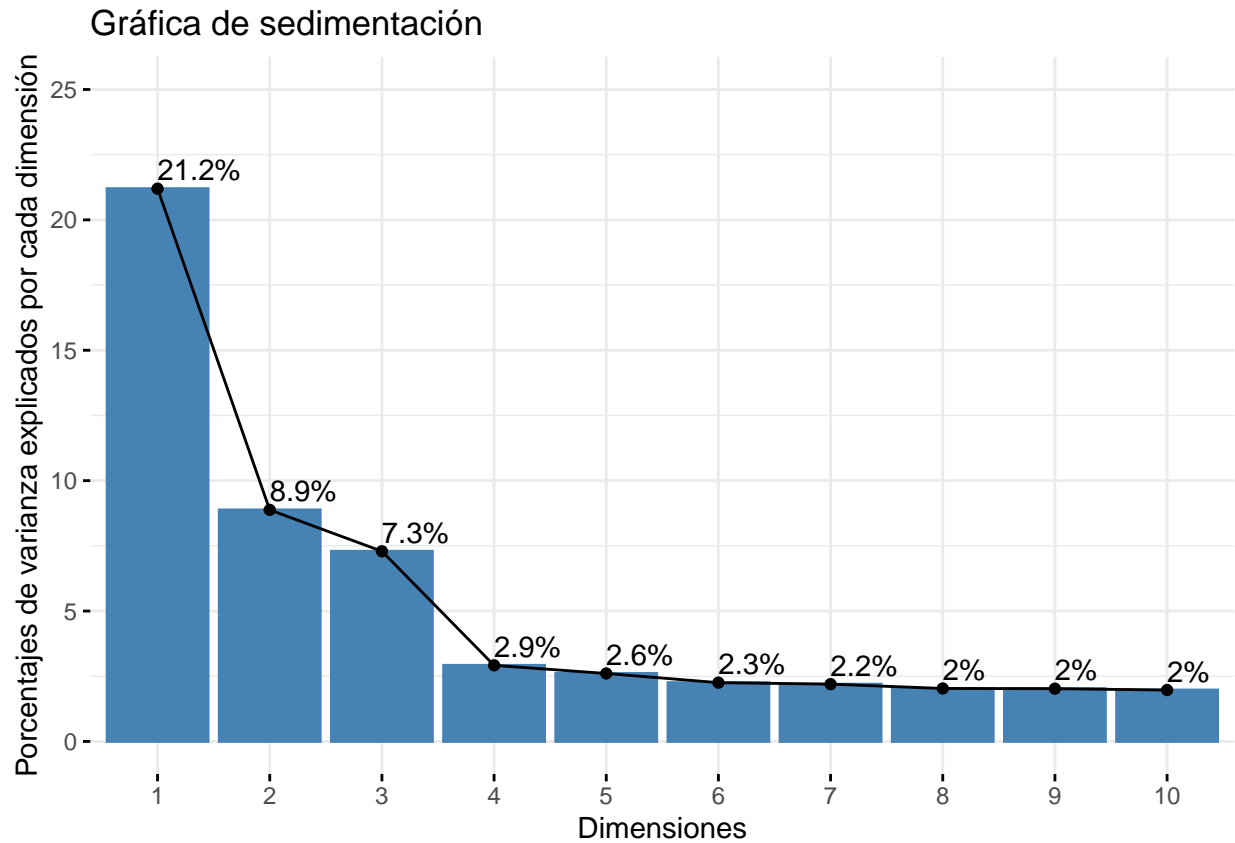
Por tanto, se transforman las 17 variables empleadas a tipo factor, con ayuda de un ciclo for.

Análisis de correspondencias múltiples

Luego, para crear la variable riqueza se hará uso del análisis de correspondencia múltiple, que es una generalización del uso del análisis de componentes principales, cuando las variables que se pretenden analizar son de tipo categórico.

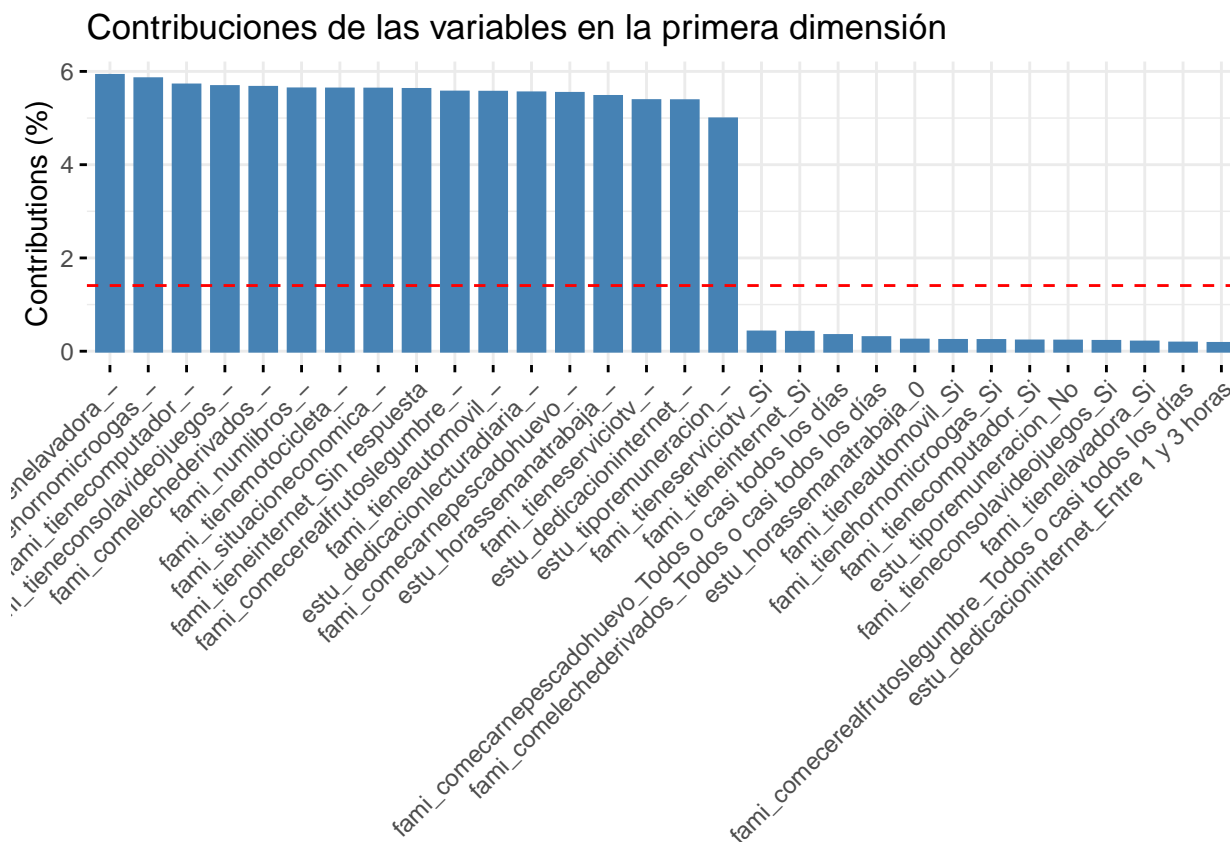
Así, se obtienen los porcentajes de varianza explicados por cada dimensión:

Gráfica de sedimentación



Donde se puede observar que la primera componente logra explicar un 21.2% la variabilidad total del conjunto de datos. Entonces, se utilizará la primera dimensión del Análisis de Correspondencia Múltiple (MCA), con la finalidad de obtener las variables más relevantes para dicha componente.

Gráfico de Contribuciones de las variables en la primera dimensión



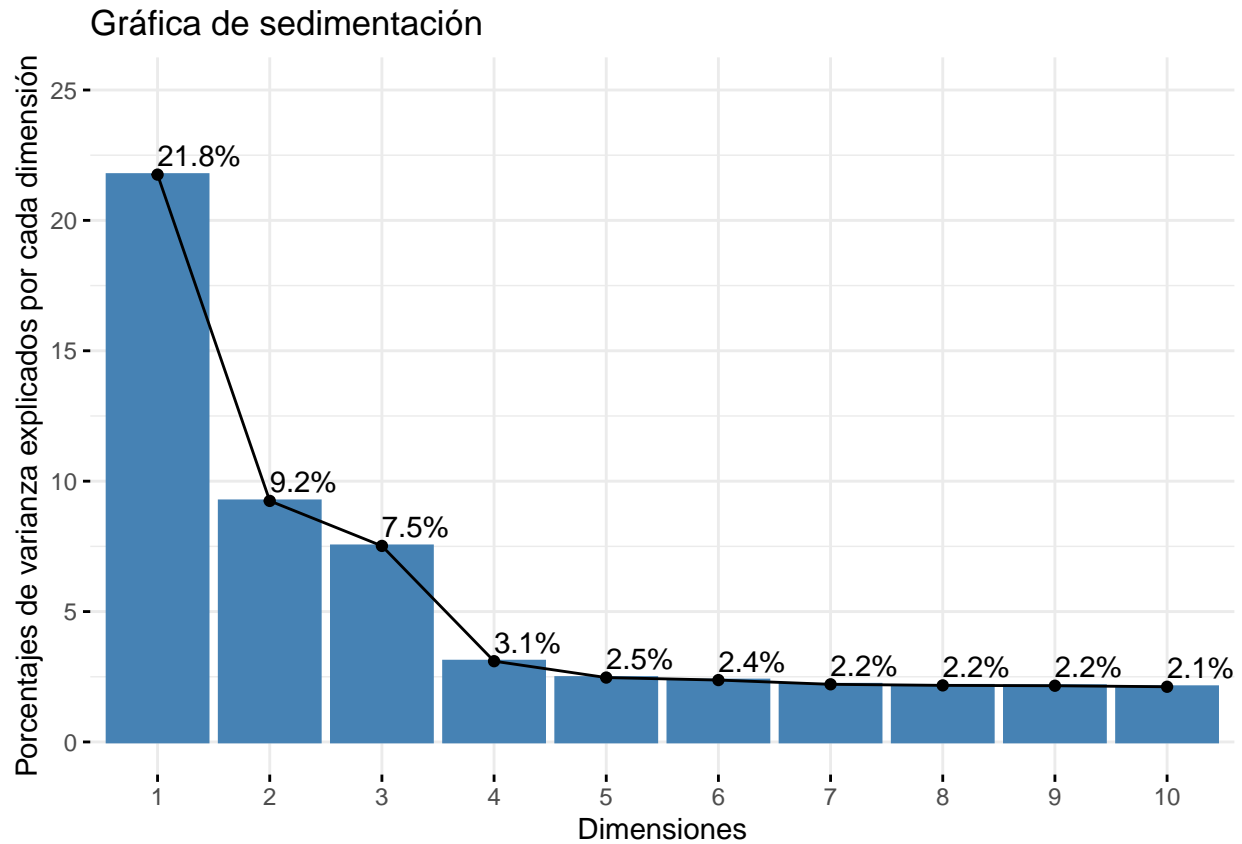
Así, se eligen las siguientes 16 variables:

- 'fami_tieneinternet'
- 'fami_tieneserviciotv'
- 'fami_tienecomputador'
- 'fami_tienelavadora'
- 'fami_tienehornomicroogas'
- 'fami_tieneautomovil'
- 'fami_tienemotocicleta'
- 'fami_tieneconsolavideojuegos'
- 'fami_numlibros'
- 'fami_comelechederivados'
- 'fami_comecarnepescadohuevo'
- 'fami_comecerealfrutoslegumbre'
- 'fami_situacioneconomica'
- 'estu_dedicacionlecturadiaria'
- 'estu_dedicacioninternet'
- 'estu_horassemanatrabaja'

Ya que son las variables más contribuyentes a la formación de la primera componente para la creación del índice de riqueza.

Luego, se utiliza nuevamente el MCA Para obtener el índice de riqueza por estudiante. Donde se visualizan los porcentajes de varianza, explicados por cada dimensión del MCA:

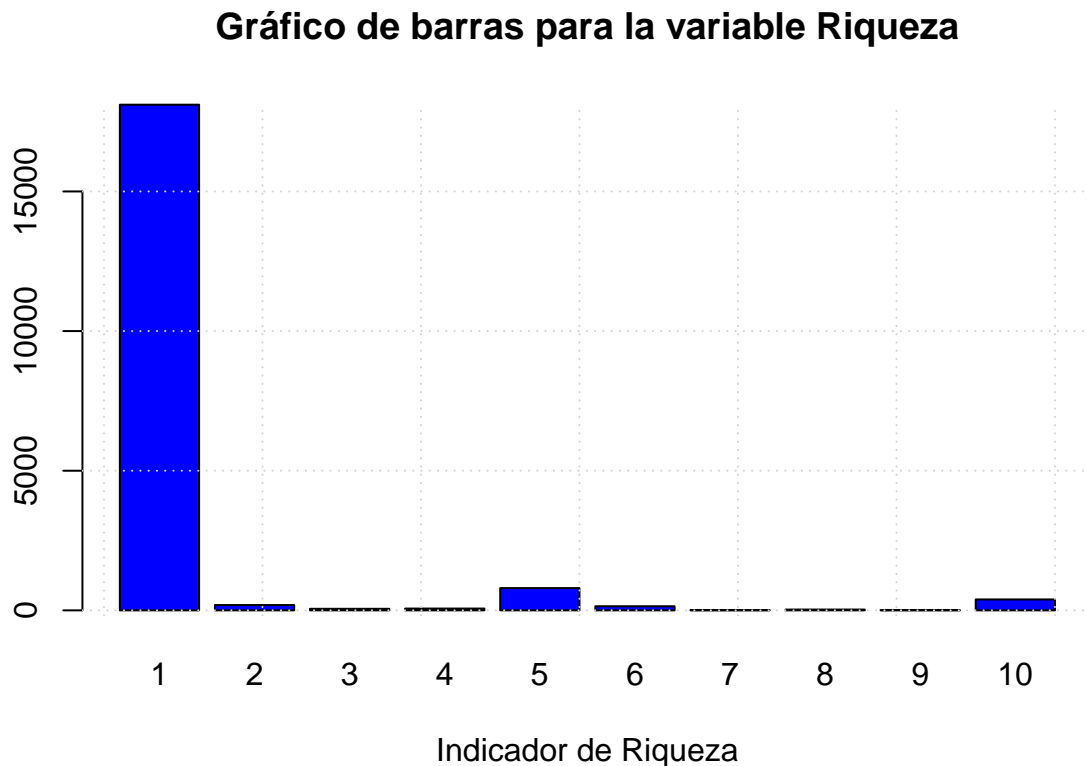
Gráfica de sedimentación actualizada



Ahora, se nota que explica un 21.8 % de la variabilidad del conjunto de datos. Entonces, la primera dimensión se utilizará como el índice de riqueza por estudiante.

Luego, se crea otra variable categórica que almacene la riqueza de los estudiantes con 10 niveles, donde el 1 va a representar al más pobre, y el nivel 10 al más rico.

Gráfico de barras de Riqueza



Del gráfico anterior, se visualiza que existen muchos alumnos en condiciones de pobreza, y que pocos gozan de tener riqueza en sus hogares.

Creación de Variable Riqueza por colegio

Anteriormente, se creó la variable de riqueza por estudiante, ahora para hacer la variable de riqueza por colegio se aprovecha la existencia de una variable de la base de datos que se denomina 'cole_cod_dane_sede', que representa un indicador que tiene cada colegio. Por tanto, se procede a agrupar por dicha variable y se sacan los promedios de riqueza por cada institución, a partir del indicador de riqueza de cada estudiante. Donde se clasifica cada colegio en 10 grupos, según su nivel de riqueza.

Entonces, como se tiene ahora por cada colegio el grupo al que pertenece en el nivel de riqueza, se procede a añadir esa información a la base de datos, creando una nueva variable donde para cada estudiante se indique el nivel de pobreza del colegio, con 10 niveles, donde 1 es el más bajo, y 10 el nivel más alto.

Análisis por departamento

Ahora se realiza un análisis comparativo de los resultados entre los colegios públicos y privados de los departamentos de Colombia.

Por tanto, se realiza una tabla con los datos, de manera resumida por departamentos:

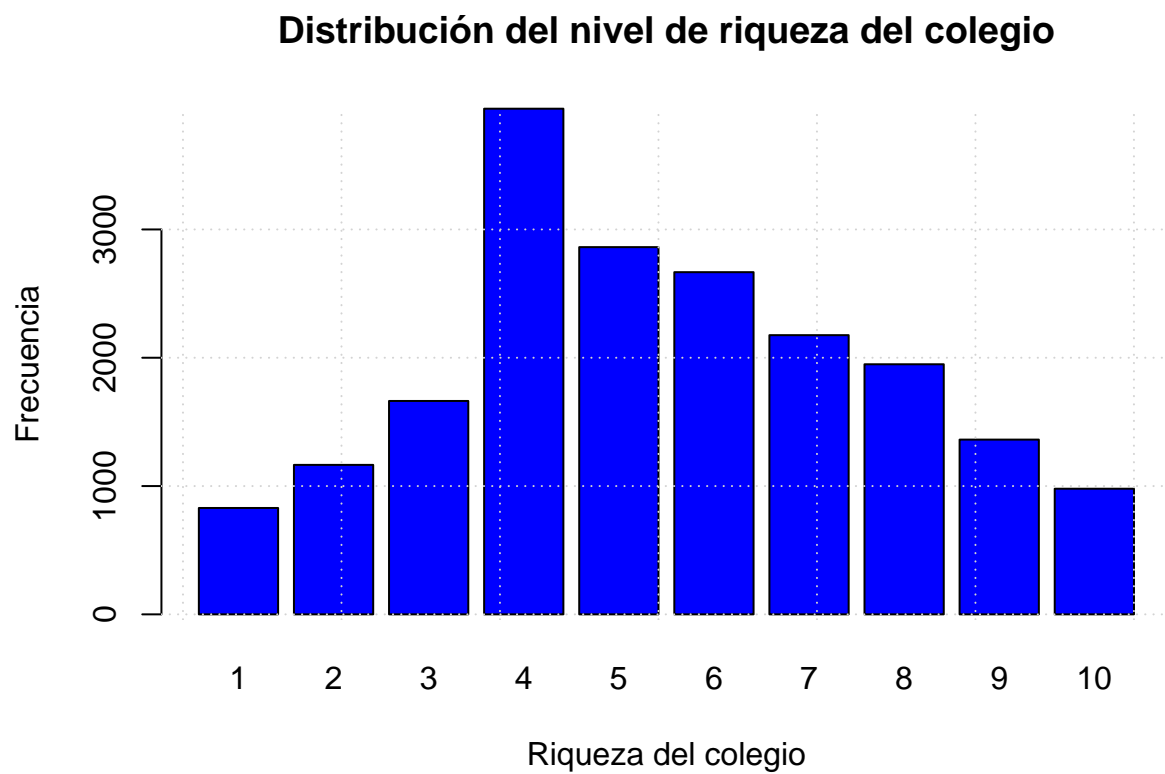
Tabla Resumen de Puntaje de Colegios Oficiales y No Oficiales con respecto a su riqueza, por Departamento.

Se muestran las primeras 10 y las 10 últimas observaciones de la tabla:

Riqueza Colegio	Ubicación del colegio	Tipo de colegio	Media de Puntaje
1	ANTIOQUIA	NO OFICIAL	306.0976
1	ANTIOQUIA	OFICIAL	247.8889
1	ARAUCA	OFICIAL	208.0000
1	ATLÁNTICO	NO OFICIAL	305.9237
1	BOGOTÁ	NO OFICIAL	312.0238
1	BOGOTÁ	OFICIAL	231.5217
1	CALDAS	OFICIAL	199.0909
1	CASANARE	OFICIAL	219.2000
1	CUNDINAMARCA	NO OFICIAL	322.9474
1	GUAINÍA	OFICIAL	212.9412
10	MAGDALENA	OFICIAL	190.4000
10	NARIÑO	NO OFICIAL	250.8636
10	NARIÑO	OFICIAL	223.6667
10	QUINDÍO	NO OFICIAL	354.4667
10	RISARALDA	NO OFICIAL	237.2353
10	RISARALDA	OFICIAL	185.0000
10	SANTANDER	NO OFICIAL	261.1304
10	TOLIMA	NO OFICIAL	171.5000
10	VALLE	NO OFICIAL	223.0317
10	VALLE	OFICIAL	234.9286

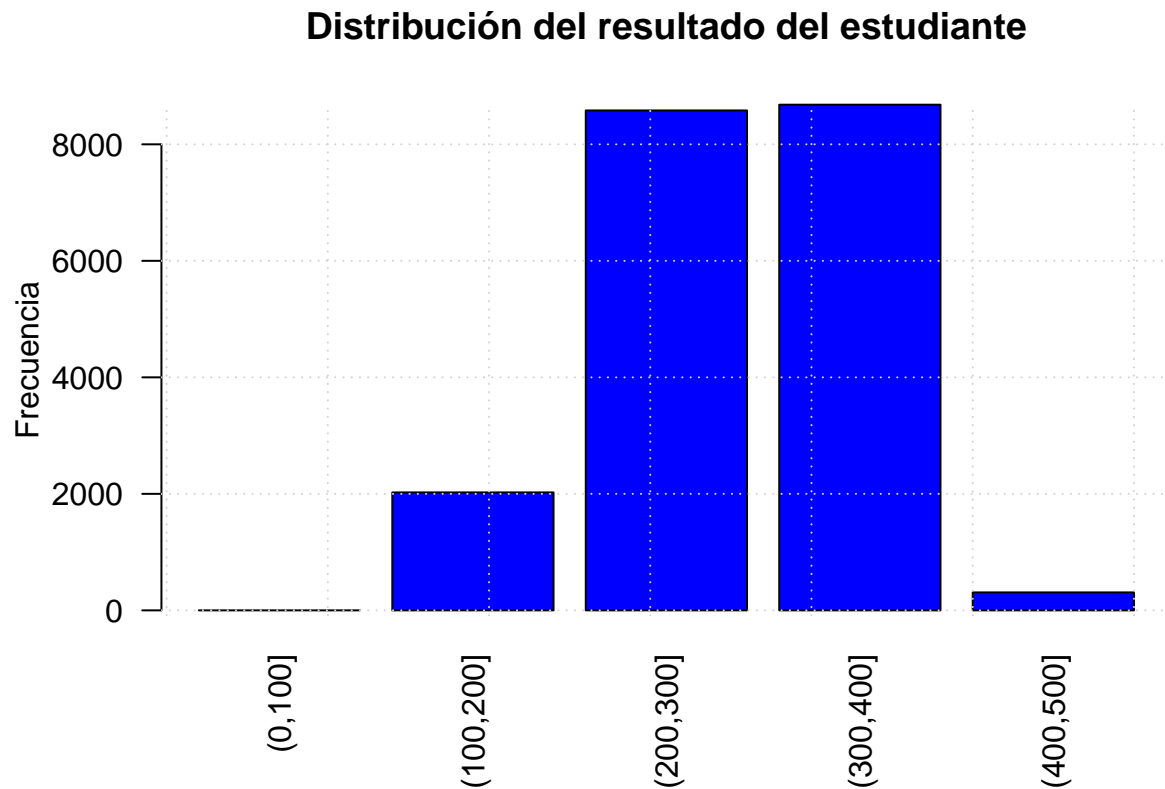
Donde, en términos generales, se puede ver un mejor desempeño en los colegios que son de carácter no oficial, con respecto a su riqueza, aunque es de notar que para algunos departamentos los desempeños se ven similares. Sin embargo, se decide observar la distribución de riqueza de los colegios mediante un gráfico de barras:

Gráfico de Distribución de Riqueza por colegio



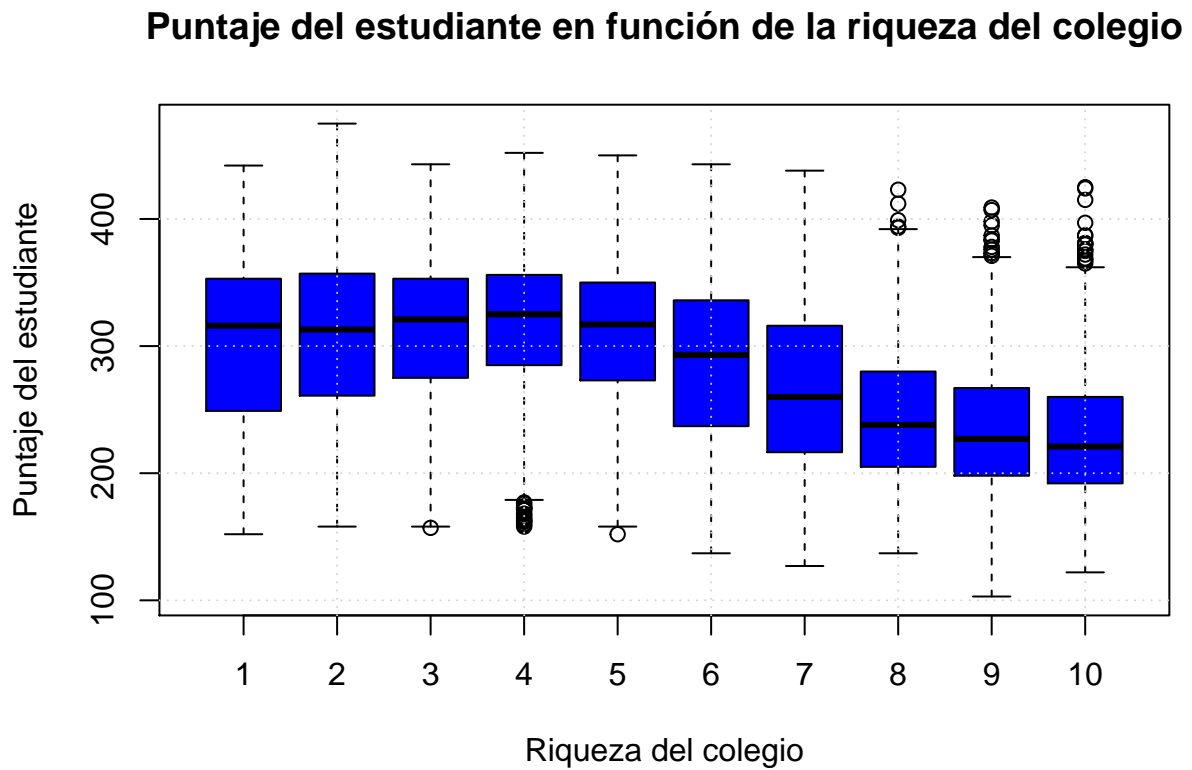
Se puede observar que la riqueza de los colegios en su mayoría se encuentran entre el nivel 4 y 7 del indicador de riqueza creado.

Gráfico para la variable del resultado por colegio.



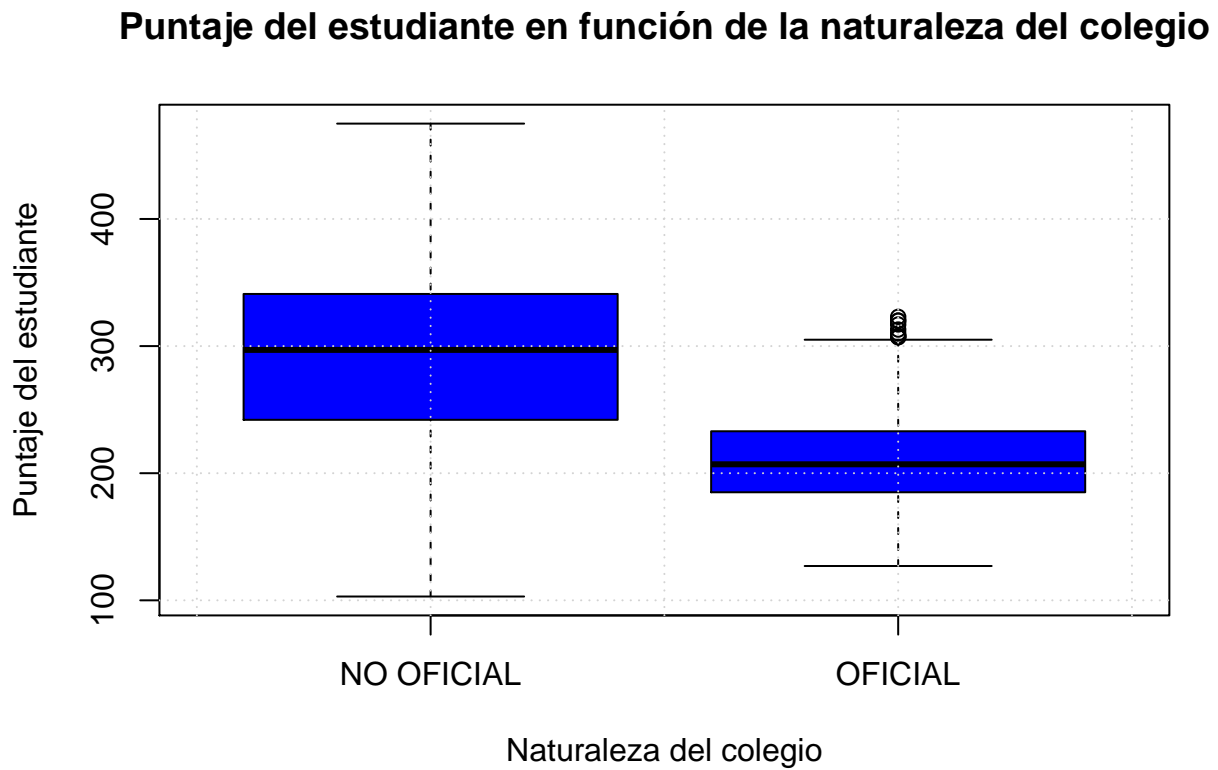
El anterior gráfico demuestra que es más natural observar que los estudiantes posean puntuaciones de la prueba Saber 11 entre $(200, 300]$ y $(300, 400]$, que en los otros rangos. Así que hay pocos estudiantes que obtuvieron malos resultados, en términos generales, y también pocos con resultados sobresalientes.

Boxplot para el Puntaje del estudiante en función de la riqueza del colegio



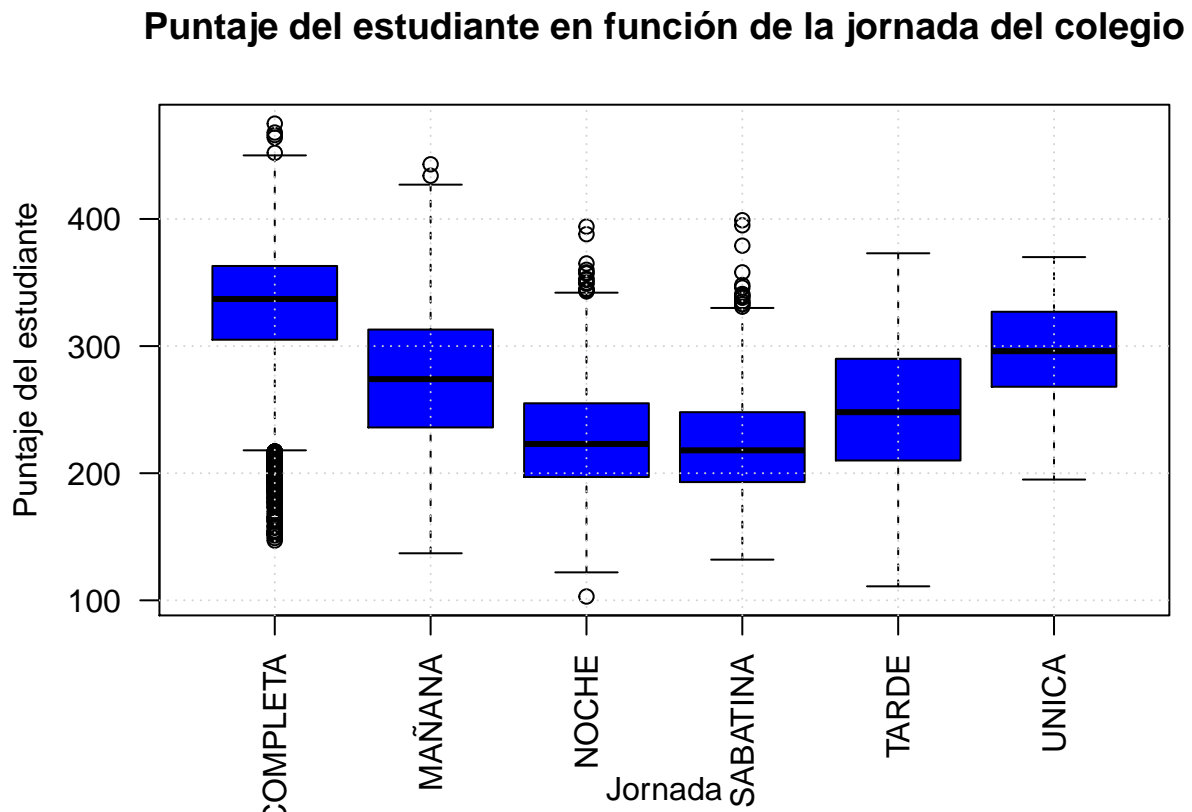
Del gráfico de cajas anterior, se puede pensar que la riqueza del colegio sí influye en el desempeño de los estudiantes en la prueba de Estado.

Boxplot para el Puntaje del estudiante en función de la naturaleza del colegio



Donde se observa que la media del puntaje para los colegios no oficiales es mayor que la de los colegios de carácter oficial, sin embargo se puede observar igualmente que el BoxPlot de los colegios de carácter no oficial, tiene mayor variabilidad.

Boxplot para el Puntaje del estudiante en función de jornada del colegio



Se observa que la media cambia con respecto a los diversos tipos de jornadas, siendo los colegios con jornada completa aquellos cuyo puntaje en promedio es el mejor en presencia de las demás, pero también cabe recalcar que posee datos atípicos en ambos extremos.

Finalmente, se puede pensar que la jornada del colegio también afecta el rendimiento de los estudiantes en la prueba.

Tablas de Contingencia del Nivel de Riqueza vs. Tipo de jornada con respecto al puntaje en las Pruebas Saber 11

Ahora, dados los resultados anteriores se decide trabajar con los niveles de riqueza y con las jornadas de los colegios, así con éstas se crean tablas de contingencia para comparar el resultado de los colegios no oficiales y oficiales, con el mismo nivel de riqueza y jornada.

NOTA: Como son muchas posibles combinaciones, lo cual implicaría la creación, de a lo sumo, 60 tablas de contingencia, se deciden mostrar algunas de ellas:

Nivel de riqueza 5 vs Jornada de la Mañana:

PUNTAJE	NO OFICIAL	OFICIAL
(0,100]	0.0000000	0.0000000
(100,200]	0.0253672	0.0652174
(200,300]	0.6154873	0.8478261
(300,400]	0.3591455	0.0869565
(400,500]	0.0000000	0.0000000

Donde se observa que para los colegios de carácter No Oficial los estudiantes con nivel de riqueza 5 de la jornada de la mañana tienen mejores puntajes que los estudiantes de colegio de carácter Oficial; se puede visualizar que el porcentaje de estudiantes que tienen una puntuación entre (300,400] es de un 35.9 % para los de colegio No Oficial, mientras que para los estudiantes de carácter Oficial solamente se encuentra en ese rango un 8.7 % del estudiantado.

Sin embargo, el 84.8 % de estudiantes de colegio Oficial se encuentran en el rango (200,300], es decir, la mayoría de dichos estudiantes, bajo el nivel de riqueza y jornada mencionados, tienen un rendimiento aceptable.

Nivel de riqueza 8 vs Jornada de la Mañana:

PUNTAJE	NO OFICIAL	OFICIAL
(0,100]	0.0000000	0.0000000
(100,200]	0.1393643	0.6666667
(200,300]	0.7139364	0.3333333
(300,400]	0.1466993	0.0000000
(400,500]	0.0000000	0.0000000

De la tabla anterior, se observa que para los colegios de carácter No Oficial de la jornada de la mañana, los estudiantes con nivel de riqueza 8 tienen mejores puntajes, en promedio, que los estudiantes de colegio de carácter Oficial; se puede visualizar que el porcentaje de estudiantes que tienen una puntuación entre (200,300] es de un 71.4 % para los de colegio No Oficial, mientras que para los estudiantes de carácter Oficial solamente se encuentra en ese rango un 33.3 % del estudiantado. Incluso en el rango (300,400] tiene un porcentaje de 14.7 % los estudiantes de colegio No Oficial, y los estudiantes del otro tipo de colegio ni siquiera se encuentran en dicho rango.

También, es notorio que los estudiantes de colegio Oficial su mayor porcentaje se encuentra en el rango de puntuación (100,200] con un 66.7 % de la población, mientras que para los estudiantes de colegio No Oficial, se encuentran con un porcentaje de 71.4 % en el rango (200,300].

Nivel de riqueza 5 vs Jornada de la Noche:

PUNTAJE	NO OFICIAL	OFICIAL
(0,100]	0.0000000	0.0
(100,200]	0.3939394	0.7
(200,300]	0.5757576	0.3
(300,400]	0.0303030	0.0
(400,500]	0.0000000	0.0

De la tabla anterior, se observa claramente que para los colegios de carácter No Oficial de la jornada de la noche, los estudiantes con nivel de riqueza 5 tienen mejores puntajes nuevamente, en comparación de los estudiantes que son de colegio de carácter Oficial; se puede observar que el porcentaje de estudiantes que tienen una puntuación entre (200,300] es de un 57.6 % para los de colegio No Oficial, mientras que para los estudiantes de carácter Oficial solamente se encuentra en ese rango un 30 % del estudiantado.

También, en el rango (100,200] tiene un porcentaje de 39.4 % los estudiantes de colegio No Oficial, y los estudiantes del colegio Oficial su mayor porcentaje se encuentra en dicho rango de puntuación del examen con un 70 % de la población. Y finalmente, el 3 % aproximadamente, de los estudiantes de colegio No Oficial, se encuentran en el rango (300,400].

Nivel de riqueza 10 vs Jornada de la Noche:

PUNTAJE	NO OFICIAL	OFICIAL
(0,100]	0.0000000	0.0000000
(100,200]	0.3261803	0.4297521
(200,300]	0.6309013	0.5619835
(300,400]	0.0429185	0.0082645
(400,500]	0.0000000	0.0000000

De la tabla anterior, se observa que tanto para los colegios de carácter No Oficial de la jornada de la noche, los estudiantes con nivel de riqueza 10 como para los estudiantes que son de colegio de carácter Oficial tienen un rendimiento semejante con respecto a los rangos de puntuación en donde se encuentran. Entonces, da indicios que para este tipo de nivel de riqueza y jornada las puntuaciones del examen no se ven afectadas por el tipo de dichos colegios (No Oficial y Oficial), así que se puede pensar en realizar una prueba chi-cuadrado de independencia.

Por tanto, se plantea la siguiente prueba de Hipótesis:

H_0 : Los dos factores son independientes vs. H_1 : Los dos factores no son independientes.

Pearson's Chi-squared test		
X-squared	df	p-value
0.041635	2	0.9794

Como el valor-p es mayor a un α de 0.05 hay suficiente evidencia estadística para no rechazar la hipótesis nula, así que los puntajes para los colegios con un nivel de riqueza de 10 puntos y de jornada de la noche, son factores independientes del tipo de colegio.

Nivel de riqueza 5 vs Jornada de la Tarde:

PUNTAJE	NO OFICIAL	OFICIAL
(0,100]	0.0000000	0.0000000
(100,200]	0.3076923	0.3333333
(200,300]	0.6923077	0.5000000
(300,400]	0.0000000	0.1666667
(400,500]	0.0000000	0.0000000

De la tabla anterior, igualmente se observa que tanto para los colegios de carácter No Oficial de la jornada de la tarde, los estudiantes con nivel de riqueza 5 como para los estudiantes que son de colegio de carácter Oficial tienen un rendimiento semejante con respecto a los rangos

de puntuación en donde se encuentran. Entonces, da indicios que para este tipo de nivel de riqueza y jornada las puntuaciones del examen no se ven afectadas por el tipo de dichos colegios (No Oficial y Oficial), así que se puede pensar en realizar una prueba chi-cuadrado de independencia.

Por tanto, se plantea la siguiente prueba de Hipótesis:

H_0 : Los dos factores son independientes vs. H_1 : Los dos factores no son independientes.

Pearson's Chi-squared test		
X-squared	df	p-value
0.19871	2	0.9054

Como el valor-p es mayor a un α de 0.05 hay suficiente evidencia estadística para no rechazar la hipótesis nula, así que los puntajes para los colegios con un nivel de riqueza de 5 puntos y de jornada de la tarde, son factores independientes del tipo de colegio.

Nivel de riqueza 1 vs Jornada Sabatina:

PUNTAJE	NO OFICIAL	OFICIAL
(0,100]	0.0000000	0.0000000
(100,200]	0.0833333	0.3529412
(200,300]	0.8333333	0.6176471
(300,400]	0.0833333	0.0294118
(400,500]	0.0000000	0.0000000

De la tabla anterior, se observa que para los colegios de carácter No Oficial de la jornada Sabatina, los estudiantes con nivel de riqueza 1 tienen mejores puntajes, en comparación de los estudiantes que son de colegio de carácter Oficial; se puede observar que el porcentaje de estudiantes que tienen una puntuación entre (200,300] es de un 83.3 % para los de colegio No Oficial, mientras que para los estudiantes de carácter Oficial solamente se encuentra en ese rango un 61.8 % del estudiantado.

También, se observa que en el rango (100,200] tiene un porcentaje de 8.3 % los estudiantes de colegio No Oficial, y los estudiantes del colegio Oficial un gran porcentaje se encuentra allí, un 35.3 % de la población con respecto a ese nivel de riqueza.

Nivel de riqueza 4 vs Jornada Sabatina:

PUNTAJE	NO OFICIAL	OFICIAL
(0,100]	0.0000000	0.0000000
(100,200]	0.3333333	0.6206897
(200,300]	0.6666667	0.3793103
(300,400]	0.0000000	0.0000000
(400,500]	0.0000000	0.0000000

Se observa que para los colegios de carácter No Oficial de la jornada sabatina, los estudiantes con nivel de riqueza 4 tienen mejores puntajes, en promedio, que los estudiantes de colegio de carácter Oficial; se puede visualizar que el porcentaje de estudiantes que tienen una puntuación entre (200,300] es de un 66.7 % para los de colegio No Oficial, mientras que para los estudiantes de carácter Oficial solamente se encuentra en ese rango un 37.9 % del estudiantado.

Luego, para los estudiantes de colegio Oficial su mayor porcentaje se encuentra en el rango de puntuación (100,200] con un 62.1 % de la población, mientras que para los estudiantes de colegio No Oficial, se encuentran con un porcentaje de 33.3 % en dicho rango.

Nivel de riqueza 10 vs Jornada Sabatina:

PUNTAJE	NO OFICIAL	OFICIAL
(0,100]	0.0000000	0.0000000
(100,200]	0.4000000	0.5544554
(200,300]	0.5673469	0.4356436
(300,400]	0.0326531	0.0099010
(400,500]	0.0000000	0.0000000

De la tabla anterior, se observa claramente que tanto para los colegios de carácter No Oficial de la jornada Sabatina, los estudiantes con nivel de Riqueza 10 como para los estudiantes que son de colegio de carácter Oficial tienen un rendimiento semejante con respecto a los rangos de puntuación en donde se encuentran, sobretodo en los rangos de (100,200] y (200,300], lo cual da indicios que, para este tipo de nivel de riqueza y jornada, las puntuaciones del examen no se ven afectadas por el tipo de colegios (No Oficial y Oficial), así que se puede pensar en realizar una prueba chi-cuadrado de independencia.

Por tanto, se plantea la siguiente prueba de Hipótesis:

H_0 : Los dos factores son independientes vs. H_1 : Los dos factores no son independientes.

Pearson's Chi-squared test		
X-squared	df	p-value
0.054454	2	0.9731

Como el valor-p es mayor a un α de 0.05 hay suficiente evidencia estadística para no rechazar la hipótesis nula, así que los puntajes para los colegios con un nivel de riqueza de 10 puntos y de jornada sabatina, son factores independientes del tipo de colegio.

Conclusiones Generales

- Se considera que de manera eficiente se realizó un estudio estadístico por medio de la metodología de Análisis de Correspondencias Múltiples (MCA).
- Mediante tablas de contingencia se resumió de manera eficiente información relevante para dar solución al problema planteado inicialmente.
- Se concluye que los colegios de carácter no oficial de Colombia presentan mejores resultados en las pruebas Saber 11°, con respecto a su nivel de riqueza y tipo de jornada.
- Se observa que para algunos casos, el rendimiento de los estudiantes independientemente del tipo de colegio es similar, esto indica que si bien en la mayoría de los casos, las instituciones de carácter no oficial de Colombia presentan mejores resultados, también existen tipos de jornadas y niveles de riqueza en donde el rendimiento medio denota unos resultados similares con respecto a los puntajes obtenidos para dichos estudiantes.

Comentarios

- Dichos análisis presentados en el transcurso del trabajo aplican para los resultados de la Prueba Saber 11° 2018-I, por tanto no es aconsejable que se generalice dicho estudio para otros años de presentación de la prueba, ya que posiblemente pueden haber variaciones, sobretodo en el marco de resultados de pruebas en medio de la situación de Pandemia mundial.
- Es recomendable que se analicen otros años en este caso de estudio, esto con el fin de saber si el comportamiento de las puntuaciones de las pruebas Saber 11, con respecto a la riqueza tiene alguna mejoría o cambios drásticos.

- Las herramientas digitales jugaron, sin duda alguna, un papel fundamental (R, R-Studio) para la realización de dicho trabajo, es por eso que se debe resaltar la importancia que tiene el saber implementar la estadística en dichos softwares.

Anexos

- En el transcurso de la semana, en dicho enlace, se adjuntará un vídeo explicativo del trabajo realizado:

<https://drive.google.com/drive/folders/1ctkTZW9Z1roizOGtKfI9voCpWVCP4FaD?usp=sharing>

- En este enlace se podrá visualizar el trabajo de una manera más didáctica:

https://rpubs.com/Kleider_V/Icfes_Research

Bibliografía

- C, Juan. (2022). *Notas de clase Datos Categóricos*. Semestre 2022. Universidad Nacional de Colombia. Sede Medellín.
- Equipo IBM. (2021). *Análisis de correspondencias múltiple*. IBM. Obtenido de <https://www.ibm.com/docs/es/spss-statistics/SaaS?topic=categories-multiple-correspondence-analysis>
- Equipo ICFES DATOS ABIERTOS. (17 de octubre de 2019, última actualización). *Resultados Saber 11 2018-1 Refinado*. Obtenido de <https://www.datos.gov.co/Educacion/Resultados-Saber-11-2018-1-Refinado/ptck-fi3s>
- Equipo R CODER. (2022). *Tabla de contingencia en R*. R CODER. Obtenido de <https://r-coder.com/tabla-contingencia-r/>
- Equipo Wikipedia. (s.f.). *ICFES*. Wikipedia. Obtenido de <https://es.wikipedia.org/wiki/ICFES>