

upna

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

fec>>

facultad de ciencias
económicas y empresariales
ekonomia eta enpresa
zientzien fakultatea
school of economics
and business administration

Facultad de Ciencias Económicas y Empresariales

TRABAJO FIN DE GRADO/MASTER EN
Administración y Dirección de Empresas

DESIGUALDAD Y EDUCACIÓN: UNA APROXIMACIÓN DESDE LA
ESTADÍSTICA MULTIVARIANTE.

Sandra Otazu Bara

Pamplona-Iruña día 20 de diciembre de 2023

Módulo: métodos cuantitativos

Dirigido por:

María Carmen García Olaverri

RESUMEN: El presente trabajo tiene como objetivo principal investigar la relación entre la desigualdad social y económica en diversos países y su repercusión en el nivel educativo. A través de herramientas estadísticas avanzadas como el análisis cluster, se lleva a cabo la clasificación de países afines con base en variables cuidadosamente elegidas relacionadas con la desigualdad. Una vez conformados los grupos de países, se analiza si la pertenencia a un clúster específico incide en los resultados y en la importancia atribuida a la educación en ese conjunto de naciones.

En resumen, la investigación busca comprender la dinámica que conecta la desigualdad en un país con su nivel educativo, explorando patrones y tendencias mediante análisis estadísticos avanzados. La metodología utilizada no solo tiene como objetivo identificar relaciones significativas, sino también comprender cómo distintos niveles de desigualdad pueden influir en las prioridades y el énfasis otorgados a la educación en contextos nacionales particulares. Este enfoque multidimensional proporciona una perspectiva integral para abordar la compleja relación entre desigualdad y educación a escala global, contribuyendo así a una comprensión más profunda de los factores que afectan la formación académica en diferentes entornos sociopolíticos.

Palabras clave: Cluster, desigualdad, ANOVA, países, educación.

ABSTRACT: this bachelor's thesis aims to investigate the relationship between social and economic inequality in diverse countries and its impact on educational levels. Employing advanced statistical tools such as analysis of variance and cluster analysis, the classification of similar countries is conducted based on carefully selected variables related to inequality. Once country groups are formed, the study examines whether belonging to a specific cluster influences outcomes and the significance attributed to education within that set of nations.

In essence, the research seeks to comprehend the dynamics linking a country's inequality with its educational level, exploring patterns and trends through advanced statistical analyses. The methodology employed not only seeks to identify significant relationships but also to understand how varying levels of inequality may influence priorities and emphasis placed on education in specific national contexts. This multidimensional approach provides a comprehensive perspective for addressing the intricate relationship between inequality and education on a global scale, contributing to a deeper

understanding of the factors influencing academic formation in different sociopolitical environments.

Key words: Cluster, inequality, ANOVA, countries, education.

Índice

1.	PRESENTACIÓN Y OBJETIVOS.....	4
2.	DESCRIPCIÓN DE LAS VARIABLES Y LA BASE DE DATOS.	6
2.1	Bases de datos utilizadas.....	6
2.2	Descripción de las variables empleadas en los distintos análisis.....	9
2.2.1	Medidas de desigualdad económica.	9
2.2.2	Indicadores de desigualdad social.	11
2.2.3	Indicadores relativos a la importancia que se otorga a la educación..	14
2.2.4	Indicadores relativos a los resultados de la educación.	16
2.3.3	Análisis existencia de modelo europeo y latinoamericano.....	17
3.	ANÁLISIS CLUSTER BASADO EN LAS MEDIDAS DE DESIGUALDAD ECONÓMICA Y SOCIAL.	19
3.1	Metodología del análisis cluster.....	20
3.2	Resultados del análisis.	21
3.3	La posición de España.....	24
3.4	La movilidad social en los clusters descritos.	25
4.	RELACIÓN ENTRE DESIGUALDAD Y EDUCACIÓN.	28
4.1	Metodología del ANOVA de un factor.	28
4.2	Prueba de homogeneidad de las varianzas.	30
4.3	Resultados de los análisis.....	31
4.3.1.	Variables de importancia de la educación.	31
4.3.2.	Variables de resultados de la educación.	35
4.3.3	Conclusiones generales.....	38
5.	BIBLIOGRAFÍA.	41

1. PRESENTACIÓN Y OBJETIVOS.

En el presente trabajo vamos a tratar de establecer la relación que pueda existir entre diferentes indicadores de desigualdad social y económica en relación con la importancia y resultados de la educación en diferentes países.

Para entender esto mejor, en primer lugar, debemos definir conceptos clave como “desigualdad social” o “desigualdad económica”.

La desigualdad económica se refiere a las disparidades en la distribución de los recursos económicos, como ingresos y riqueza, entre diferentes individuos, grupos o regiones dentro de una sociedad. En este caso, analizaremos, para un conjunto de países, la disparidad entre los individuos de cada país. Esta disparidad puede manifestarse de diversas maneras, incluyendo la brecha entre los ingresos más altos y bajos, la concentración de la riqueza en manos de unos pocos.

Para ello, hemos tomado como referencia variables relativas al reparto de ingresos, como puede ser la brecha entre los ingresos más altos y bajos, la concentración de la riqueza en manos de unos poco o el índice de Gini.

Analizaremos estas variables más detenidamente más adelante, pero a modo introductorio se podría decir que este último indicador, el índice de Gini es la medida estadística utilizada para evaluar la desigualdad en la distribución de ingresos o riqueza por excelencia, que cuantifica la distribución de ingresos dentro de una población. Cuando existe una desigualdad económica significativa, puede tener impactos negativos en la cohesión social, la movilidad económica, el acceso a servicios básicos y la estabilidad económica en general. La reducción de la desigualdad económica a menudo es un objetivo clave en las políticas públicas para promover una sociedad más justa y sostenible.

En segundo lugar, en lo que respecta a la desigualdad social, podemos decir que esta se refiere a las disparidades y diferencias sistemáticas en el acceso, control y distribución de recursos, oportunidades, derechos y beneficios dentro de una sociedad. Estas disparidades pueden manifestarse en diversos aspectos, como falta de acceso equitativo a oportunidades educativas y laborales, ingresos económicos, educación, salud, acceso a empleo, vivienda, poder político y otros recursos que influyen en la calidad de vida de los individuos.

La desigualdad social abarca cuestiones relacionadas con el género, la raza, la etnia, la orientación sexual, la religión y otras características que pueden influir en la posición social de una persona.

En definitiva, hemos tenido en cuenta en un primer momento variables que establecen la relación entre tasa de participación mujeres y hombres en la fuerza laboral, tasas de trabajadores en empleos vulnerables, diversidad que encontramos en las escuelas de los diferentes países...

Al igual que ocurre con la desigualdad económica, la lucha contra la desigualdad social es también un objetivo clave en muchas sociedades, ya que se busca promover la equidad, la justicia y la inclusión, creando condiciones para que todos los individuos tengan acceso a oportunidades y derechos básicos independientemente de sus circunstancias.

El objetivo de este trabajo es analizar la relación entre los indicadores de desigualdad descritos previamente y ver si existe una asociación con diferentes variables relativas a la educación. Para ello vamos a tratar de analizar en profundidad un conjunto de variables relacionadas con la desigualdad social, económica y la educación de una serie de países seleccionados. Estos países se dividen en dos áreas geográficas: Europa y Latinoamérica. De un total de 48 países, 37 son países europeos y 11 Latinoamericanos.

Entre las cuestiones que vamos a tratar de responder se encuentran:

¿Existe una relación de causalidad entre ser un país desigual e invertir poco en educación?

¿Puede ser un país desigual en términos económicos, pero ser igualitario en términos sociales?

¿Es la desigualdad del país lo que explica los bajos valores en las variables relativas a la educación o es el bajo nivel educativo el que convierte al país en un país con mucha desigualdad?

¿Existe un modelo diferenciado para Latinoamérica y otro para Europa en lo que a desigualdad se refiere?

¿Qué posición ocupa España en este análisis respecto de Europa? ¿Y en el conjunto global de los países?

¿Qué agrupación llevan a cabo los diferentes países en función de la semejanza en materia de desigualdad social y económica?

El trabajo se organiza como sigue:

En el apartado segundo se lleva a cabo una descripción de la base de datos, así como de las variables que se van a analizar en el estudio. Las variables se dividen en tres grupos. En primer lugar, se encuentran las variables relativas a la desigualdad económica,

por otro lado, tenemos las variables que se engloban en la desigualdad social y por último tenemos los indicadores relativos a la educación.

En el tercer apartado se lleva a cabo un Análisis cluster mediante el cual agruparemos a los países en función de las variables relativas a la desigualdad, tanto económica como social. Analizaremos en este apartado también en lugar que ocupa España en relación con el resto de los países y a los clusters creados.

En este mismo apartado analizaremos la movilidad social y la relación que esta guarda tanto con la desigualdad como con el PIB per cápita.

El siguiente apartado está dedicado a explorar las relaciones entre la pertenencia a uno de los clusters anteriormente descritos y los indicadores de educación, analizaremos también la existencia de un único modelo para Europa y otro para Latinoamérica. Todos estos análisis incluirán en este mismo apartado las conclusiones pertinentes relativas a cada cuestión.

El trabajo incluye la bibliografía consultada en la elaboración de este.

2. DESCRIPCIÓN DE LAS VARIABLES Y LA BASE DE DATOS.

Este trabajo es fundamentalmente empírico por lo que resulta esencial describir las fuentes de información utilizadas y así como la metodología estadística empleada.

2.1 Bases de datos utilizadas.

Para llevar a cabo el análisis hemos utilizado los datos del informe llamado “The Global Social Mobility Report 2020 Equality, Opportunity and a New Economic Imperative “. Informe que se centra principalmente en el concepto de movilidad social a nivel mundial. Este informe, publicado por el Foro Económico Mundial, analiza la movilidad social, que se refiere a la capacidad de las personas para mejorar su posición socioeconómica en comparación con la generación de sus padres.

Por otro lado, aborda temas referidos a la desigualdad, agrupada en 10 pilares, estos son los siguientes:

Pilar 1: Salud Este pilar evalúa la capacidad de los países para ofrecer atención médica de alta calidad a sus ciudadanos, reconociendo que el acceso a servicios de salud impacta de por vida la movilidad social. Se destaca cómo las desigualdades en la salud, como la brecha en la esperanza de vida entre diferentes estratos socioeconómicos, pueden

influir directamente en el acceso al empleo. Además, se señala que factores relacionados con la salud mental y adicciones también contribuyen a la brecha entre salud y riqueza.

Pilares 2-4: Educación (Acceso, Calidad y Equidad, y Aprendizaje de por Vida) Estos pilares evalúan la capacidad de los países para proporcionar acceso, calidad y equidad en la educación, así como para fomentar el aprendizaje continuo a lo largo de la vida. Se destaca la importancia de estrategias de desarrollo de capital humano desde la infancia hasta la vida adulta, con un énfasis especial en los primeros 1,000 días de la vida del niño. Además, se enfatiza la necesidad de abordar barreras estructurales que limitan el acceso de niños desfavorecidos a la educación de calidad.

Pilar 5: Acceso a la Tecnología Este pilar mide el nivel de acceso y adopción de la tecnología entre la población. Se destaca cómo el acceso a la tecnología puede actuar como un igualador al proporcionar información de manera equitativa, independientemente del trasfondo socioeconómico. Se reconoce el papel crucial de la tecnología, especialmente la educación en línea, para mejorar el aprendizaje a lo largo de la vida y aumentar las oportunidades de empleo en una economía en constante cambio.

Pilares 6-8: Oportunidades Laborales Justas: Estos tres pilares concretamente son: Oportunidades Laborales, Salarios Justos y Condiciones Laborales. Estos pilares evalúan la capacidad de las economías para proporcionar oportunidades laborales, salarios justos y condiciones laborales equitativas. Destacan la necesidad de convertir habilidades en ingresos para impulsar la movilidad económica. También se abordan los desafíos del cambio tecnológico y la transición verde en el mercado laboral, así como la importancia de transiciones efectivas de la escuela al trabajo. La calidad de los contratos laborales y la equidad en el acceso a oportunidades laborales son elementos críticos.

Pilares 9-10: Protección Social e Instituciones Inclusivas Estos pilares miden la capacidad de las economías para ofrecer protección social, instituciones inclusivas y servicios públicos eficientes. Destacan la importancia de las redes de seguridad social y las instituciones justas para mitigar el impacto de las transiciones laborales y reducir las disparidades en los estándares de vida. Se señala la limitada cobertura de protección social en muchas economías, excluyendo a trabajadores vulnerables. La inclusividad de las

instituciones, la eficiencia del gobierno y la lucha contra la corrupción son aspectos cruciales para promover la equidad y la participación equitativa en la sociedad.

En nuestro trabajo hemos tomado variables de los pilares 2 (acceso a la educación), 3 (educación de calidad y equitativa), 9 (protección social), 6 (oportunidades laborales) y 7 (salarios justos).

En cuanto a los países elegidos, aunque en un primer momento se pensó en incluir únicamente territorios europeos se comprobó que había muy poca variabilidad entre ellos por lo que se decidió incluir un grupo de países de América Latina. Inicialmente la base de datos estaba compuesta por 50 países, pero se descartaron algunos de ellos (Malta y Luxemburgo) por su escasa relevancia estadística y en general con información estadística poco fiable. Perú es un país que se ha tenido en cuenta en todos los análisis descriptivos de las variables, pero se ha excluido en la parte final del trabajo, como se explicará más adelante.

El grupo de países para los cuales se lleva a cabo el estudio multivariante es:

Países Europeos: Albania, Alemania, Armenia, Austria, Bélgica, Bulgaria, Chipre, Croacia, Dinamarca, Eslovaquia, Eslovenia, España, Estonia, Finlandia, Francia, Georgia, Grecia, Hungría, Irlanda, Islandia, Italia, Letonia, Lituania, Moldavia, Noruega, Países Bajos, Polonia, Portugal, Reino Unido, República Checa, Rumania, Rusia, Serbia, Suecia, Suiza, Turquía, Ucrania.

Países Latinoamericanos: Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, México, Panamá, Paraguay, Perú, Uruguay

La siguiente ilustración, tomada del “The global social mobility report 2020 equality, opportunity and a new economic imperative” sintetiza la estructura de la base de datos empleada



Algunas de las variables empleadas, como por ejemplo el índice de niños sin escolarizar, no estaban completas para todos los países. Estos datos los hemos completado con los extraídos de la página “knoema” (<https://knoema.es/atlas?view=snowflake>)

Por otro lado, hemos buscado en fuentes externas el Índice de Gini al ser uno de los principales indicadores de desigualdad en la distribución de ingresos. Estos valores los hemos extraído de la fuente “banco mundial de datos” (<https://datos.bancomundial.org/indicador/SI.POV.GINI>)

2.2 Descripción de las variables empleadas en los distintos análisis.

Clasificamos las variables en dos grupos: Las relativas a desigualdad y las relativas a educación.

Dentro de los indicadores de desigualdad distinguimos entre los de Desigualdad económica (4 variables) y los de Desigualdad Social (4 variables). En cuanto a las medidas de educación distinguimos entre indicadores de Importancia otorgada a la Educación (3 variables) y Resultados de la Educación (3 variables)

Describimos a continuación todos ellos.

2.2.1 Medidas de desigualdad económica.

Índice de Gini

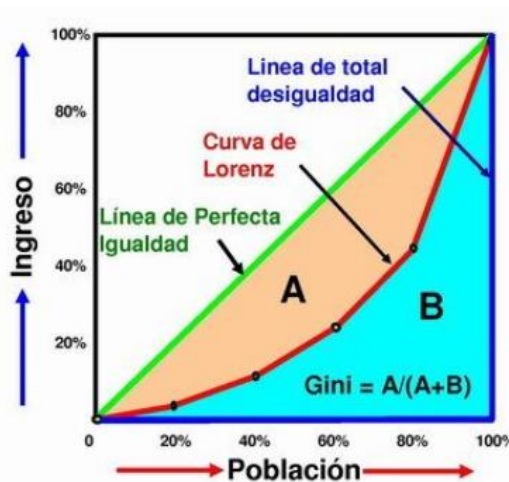
El índice de Gini es una medida estadística utilizada para evaluar la desigualdad en la distribución de ingresos o riqueza en una población o región, se calcula como un valor numérico entre 0 y 1, donde 0 representa la igualdad perfecta, lo que significa que todos los individuos tienen los mismos ingresos o riqueza, y 1 representa la desigualdad perfecta, lo que significa que una sola persona posee todos los ingresos o riqueza y todos los demás no tienen nada.

Gráficamente el índice de Gini (G) es:

$$G=A/(A+B)$$

A es el área entre la línea de Lorenz (la curva que representa la distribución real de ingresos) y la línea de igualdad perfecta (donde todos tienen ingresos iguales), y B es el área bajo la línea de igualdad perfecta.

En la práctica, el índice de Gini se expresa como un porcentaje multiplicando el resultado por 100.



$$G = \left| 1 - \sum_{k=1}^{n-1} (X_{k+1} - X_k)(Y_{k+1} + Y_k) \right|$$

X: proporción acumulada de población.

Y: proporción acumulada de ingresos.

Concretamente en nuestros datos destaca Eslovenia con la menor desigualdad (0,24) y Brasil siendo el país con mayor desigualdad (0,529). Siendo el promedio de todos los países de 0,347.

Relación entre el 40% inferior y el 10% superior de la participación en los ingresos laborales:

Estadístico calculado a partir del siguiente cociente:

media del 40% de las rentas más bajas/ media del 10% de las rentas más altas

Es un valor que podría superar el 100 ya que estamos analizando la suma de 4 rentas bajas en relación con 1 alta, por lo tanto, la suma de las 4 rentas bajas podría superar a 1 renta alta.

Sin embargo, este caso se da únicamente en Eslovaquia obteniendo un valor de 102.3, mientras que la media de todos los países es de 59.17 lo cual quiere decir que, de media, la suma de 4 rentas medias pertenecientes al 40% inferior no llegan a representar el 60% de una renta media perteneciente al 10% superior.

España se encuentra por encima de la media con un valor de 62.4%.

Conforme menor sea el valor del estadístico mayor será la desigualdad de ese país.

Relación entre el 50% inferior y el 50% superior de participación laboral en los ingresos

Este índice representa el valor de dividir la media del 50% de las rentas del trabajo más bajas entre la media del 50% restante, es decir, de las rentas más altas de ese país.

Dicho valor oscilará siempre entre 0 y 1. Cuanto más cercano a 1 sea el valor obtenido menor desigualdad existirá en dicho país ya que las medias de rentas laborales bajas y las altas serían más iguales.

Sin embargo, la realidad es que el mayor valor se corresponde al país de Eslovaquia con un valor del 46.7%, lo cual quiere decir que ni siquiera juntando dos rentas medias del 50% de rentas más bajas se lograría llegar a la renta media del 50% con rentas más altas.

Siendo Perú el País con el índice más bajo (12.2) donde se necesitarían más de 8 rentas medias pertenecientes al 50% de rentas más bajas para llegar a contabilizar la misma renta de una persona perteneciente al 50% de rentas más altas. Es decir, la diferencia entre rentas bajas y altas es mucho mayor.

Renta media del 40% inferior (% de la renta media nacional)

Este índice representa la renta media del 40% más pobre de cada país en función de la renta media nacional. Es decir, cuanto más cercano a 1 sea ese valor menos pobreza existirá en ese país ya que la renta del 40% más pobre se correspondería a la renta media del país.

(renta 40% más pobre/renta media del país)

Nos llama la atención los 4 países con más igualdad respecto de este índice que son: Eslovenia (60.2), Moldavia (60.2), República Checa (60.8) y Ucrania (61.2).

Así como también cabe destacar los países de Brasil y Panamá con valores de 28.8 y 28.9 respectivamente lo cual quiere decir que la renta media del 40% más pobre no representa ni el 30% de la renta media de dicho país. La diferencia entre pobres y ricos debe ser muy grande, lo que eleva la renta media. De ahí ese valor tan reducido.

2.2.2 Indicadores de desigualdad social.

Relación entre tasa de participación mujeres y hombres en la fuerza laboral

Relación entre la tasa de participación de mujeres y hombres en la fuerza laboral, se calcula dividiendo la participación femenina en la fuerza laboral por la tasa de

participación masculina en la fuerza laboral, y multiplicando por 100. Representa la cantidad de mujeres que trabajan por cada 100 hombres trabajando.

Es decir, un 100 representaría igualdad en la participación de hombres y de mujeres, por cada 100 hombres que trabajan, trabajarían 100 mujeres.

Destaca Turquía siendo el país con el índice más bajo de toda la tabla (46,2) donde cada 2 hombres no llegan a trabajar ni una mujer. Mientras que la media de todos los países seleccionados es de 77,29 mujeres trabajando por cada 100 hombres.

En el caso de España, observamos un valor notablemente alto de 81,8 mujeres trabajando por cada 100 hombres. Este indicador sugiere una mayor inclusión de mujeres en el ámbito laboral en comparación con la media global.

Trabajadores en empleos vulnerables (%)

Representa la tasa de trabajadores familiares auxiliares o trabajadores por cuenta propia como porcentaje del empleo total

Cabe destacar el amplio rango que oscila entre el país con menos trabajadores en empleos vulnerables (Noruega con 4,7) y el país que abarca más trabajadores de esta categoría (Albania 54,3). En este contexto, España presenta un valor de 11,3, lo que sugiere que la proporción de trabajadores en empleos vulnerables en el país se encuentra en una posición intermedia en comparación con los extremos mencionados.

El valor de 11,3 en España indica una situación relativamente favorable en términos de empleos vulnerables en comparación con la media global que cuenta con un valor de 18,45.

Cobertura de la protección social (% de la población)

Proporción de la población efectivamente cubierta por un sistema social de protección, incluidos pisos de protección social, infancia, maternidad, prestaciones y apoyo a personas desempleadas, personas con discapacidades, víctimas de accidentes laborales y personas mayores.

Destaca Georgia donde únicamente el 28.6% de los ciudadanos cuentan con esta cobertura de protección social. Sin embargo, la media total de todos los países es de 82.09%. Esto se debe a que la gran mayoría de los países cuentan con una alta cobertura llegando a ser del 100% en países como Francia, Eslovenia, Finlandia, Bélgica o Suecia.

Sin embargo, a la hora de realizar el análisis cluster no nos era útil debido a que en el informe utilizado para la extracción de datos no se encontraban los valores relativos a todos los países para esta variable. Por lo que esta fue la única variable descartada.

Diversidad social en escuelas

Puntuación en el índice PISA de inclusión social. El índice PISA de inclusión social se calcula como $100*(1-\rho)$, donde ρ representa la correlación intraclase del estatus socioeconómico. La correlación intraclase, a su vez, es la variación en el estatus socioeconómico de los estudiantes entre escuelas, dividido por la suma de la variación en el estatus socioeconómico de los estudiantes entre escuelas y la variación en el nivel socioeconómico de los estudiantes dentro de la escuela.

$\rho = \text{variación nivel socioeconómico entre escuelas} / (\text{variación nivel socioeconómico entre escuelas} + \text{variación nivel socioeconómico estudiantes})$

Cuanto menor sea ρ implica menor variación socioeconómico entre las escuelas y más variación de nivel socioeconómico entre estudiantes. Es decir, el indicador $100*\rho$ será mayor, implicará más diversidad social en las escuelas y una tasa más cercana a 100.

Como era de esperar son los países de América del sur donde más diferencia de nivel socioeconómico hay entre escuelas (existen escuelas para ricos y escuelas para pobres) con un valor en torno a 60. Y, por otro lado, países del norte de Europa como Noruega, Finlandia, Dinamarca, Islandia donde menos variación hay entre escuelas y más entre los propios alumnos de las escuelas. Es decir, no hay escuelas de ricos ni pobres, en una misma escuela conviven alumnos de diferentes rangos sociales y económicos.

Destacan países como Hungría, Eslovaquia o Bulgaria donde el nivel de diversidad social es tan bajo como en los países de América del sur.

Es importante que tengamos en cuenta que estas medidas de desigualdad tienen distinto sentido de variación. Así, un país muy igualitario tiene un índice de Gini y un índice de trabajadores en empleos vulnerables por debajo de la media.

En cambio, ese país igualitario ideal tendrá una relación entre el 40% inferior y el 10% superior de la participación en los ingresos laborales, relación entre el 50% inferior y el 50% superior de participación laboral en los ingresos, renta media del 40% inferior (% de la renta media nacional), relación entre tasa de participación mujeres y hombres en la fuerza laboral y diversidad social en escuelas por encima de la media.

El siguiente cuadro muestra el sentido de la variación de los indicadores donde un signo negativo implica estar por debajo de la media y positivo implica estar por encima. Representamos los modelos idílicos de un país igualitario y un país desigual.

Tabla 1: Sentido de la variación de las variables de desigualdad.

	PAIS IGUALITARIO	PAÍS DESIGUAL
Gini	-	+
Trabajadores vulnerables	-	+
40 inferior /10 superior	+	-
50 inferior /50 superior	+	-
40 inferior /media	+	-
Mujeres/hombres	+	-
Diversidad social escuelas	+	-

2.2.3 Indicadores relativos a la importancia que se otorga a la educación.

Gasto público en educación per cápita (PIB per cápita en educación)

Como su propio nombre indica, representa el gasto público por persona que se invierte en educación. Es un indicador de la importancia que cada estado otorga a la educación.

Para obtener este indicador se ha utilizado el PIB per cápita de cada país y el porcentaje del PIB dedicado a educación.

Se observan grandes diferencias entre el país donde menos se invierte en educación por persona siendo este Paraguay con 203,05 euros persona y el país donde más se invierte, Islandia, donde se invierte 5613,53 euros por persona. Es decir, se invierte más de 27 veces la cantidad invertida en Paraguay.

España invierte en educación un 4,6% del total de los bienes y servicios producidos en España, además cuenta con un PIB per cápita de 29350,2 lo que se corresponde con 1350,11 euros invertidos en educación por persona.

Niños sin escolarizar

Representa el porcentaje de niños no escolarizados como porcentaje en relación del total de niños en edad de escuela primaria. (niños no escolarizados/total de niños en edad de escolarización)

El país con mayor tasa de niños sin escolarizar es Paraguay donde por cada 100 niños en edad de escolarizar no se escolarizan a 10,8.

La mediana de dicha variable es 1,5 lo cual indica que la mitad de los países analizados poseen 1,5 niños sin escolarizar por cada 100 niños en edad de escolarización como máximo. Valor que coincide con la tasa en España de niños sin escolarizar (1,5)

Porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo

Proporción de estudiantes en escuelas cuyo director reportó falta de material educativo entre estudiantes desfavorecidos.

Destaca la amplitud del rango entre el país con menos porcentaje de alumnos desfavorecidos con falta de material. Dicho rango oscila entre el 7.1% en Países Bajos y 85.2% en Colombia, obteniéndose así una amplitud total de 78,1% de diferencia.

Dichos valores notifican que mientras que en Países Bajos de todos los estudiantes desfavorecidos únicamente el 7.1% carecen de material educativo, en Colombia es el 85.2% de esos estudiantes desfavorecidos los que están desprovistos de dichos materiales.

Como en el caso de los indicadores de desigualdad en la tabla 2 se sintetiza el tipo de puntuación que se espera para las tres variables.

Tabla 2: Sentido de variación de las variables relativas a la importancia de la educación.

	MUCHA IMPORTANCIA	POCA IMPORTANCIA
Niños sin escolarizar (%)	Menos puntuación	Más puntuación
Estudiantes desfavorecidos en escuelas que reportan falta de material educativo (%)	Menos puntuación	Más puntuación
Gasto público en educación	Más puntuación	Menos puntuación

2.2.4 Indicadores relativos a los resultados de la educación.

Índice de educación ajustado por desigualdad (0-1 el mejor)

Promedio ajustado por desigualdad entre los años medios de escolaridad y años esperados de escolaridad.

Los años medios y esperados de escolaridad se basan en datos de encuestas de hogares armonizados en bases de datos internacionales (Estudio de Ingresos de Luxemburgo; Eurostat, Encuesta de la Unión Europea sobre ingresos y condiciones de vida; Datos de distribución internacional del ingreso del Banco Mundial, Encuestas Macro demográficas y de Salud del ICF; Encuesta por grupos de indicadores múltiples del Fondo de las Naciones Unidas para la Infancia; el Centro de Estudios Distributivos, Laborales y Sociales; la base de datos socioeconómica del Banco Mundial para América Latina y el Caribe; y la Base de datos sobre desigualdad de ingresos en el mundo de la Universidad de las Naciones Unidas)

Cabe destacar el valor más bajo de este indicativo que es de 0,5 en países como Brasil, Colombia y Paraguay. Este valor indica que en estos países se llega por lo menos a la mitad de los años esperados de escolaridad.

España tiene un valor de 0,7 (menor que la moda que es de 0,8) lo cual indica que un niño escolarizado medio en España realiza un 70% de los años totales esperados de escolaridad.

Niños por debajo del nivel mínimo de competencia en lectura (%)

Proporción de niños al final de la escuela primaria que leen por debajo del nivel mínimo de competencia, según lo definido por la Alianza Global para Monitorear Aprendizaje (GAML), en el contexto de la sostenibilidad Seguimiento del Objetivo de Desarrollo 4.1.1.

Nos llama la atención que los tres países con menos niños por debajo del nivel mínimo son Letonia, Rusia y Croacia con unos valores de 0.8, 0.9 y 1 respectivamente.

Así como también nos llama la atención el alto valor adquirido en el país de Uruguay donde el 71,3% de los niños que finalizan la escuela primaria se encuentran por debajo del nivel mínimo de competencia.

Resultados de aprendizajes armonizados

Puntuación compuesta que representa el logro en 7 pruebas de rendimiento internacionales y regionales

Los resultados de aprendizaje armonizados se producen utilizando un factor de conversión para comparar internacional y regionalmente pruebas de rendimiento estandarizadas. Estas pruebas incluyen PISA, TIMSS, PIRLS, SACMEQ, LLECE y PASEC.

Todas las puntuaciones medias fueron calculadas en una escala con un punto central de 500. Es decir, los países que obtengan una puntuación por debajo de 500 se encuentran por debajo de la media y a la inversa los países con una puntuación por encima de 500.

La media de los países europeos es de 510.19 y la media de los países sudamericanos estudiados es de 419.75, prácticamente 80 puntos por debajo de la media y 90 por debajo de la media de los países europeos.

La tabla 3 sintetiza el tipo de puntuación que se espera para las tres variables de resultados de la educación.

Tabla 3: Sentido de variación de las variables relativas a los resultados en educación

	MEJOR RESULTADO	PEOR RESULTADO
Índice de educación ajustado por desigualdad	Cercano a 1	Cercano a 0
Niños por debajo del nivel mínimo (%)	Menor porcentaje	Mayor porcentaje
Resultados de aprendizajes armonizados	Mayor porcentaje	Menor porcentaje

2.3.3 Análisis existencia de modelo europeo y latinoamericano.

En este apartado llevamos a cabo un análisis para evaluar el impacto que tiene sobre la educación el hecho de ser un país europeo o latinoamericano. Nos preguntamos si se da la misma importancia y resultados en ambas regiones geográficas.

Pruebas de homogeneidad de varianzas

		Estadístico de Levene	gl1	gl2	Sig.
Niños sin escolarizar	Se basa en la media	9,697	1	45	,003
Niños por debajo del nivel mínimo (%)	Se basa en la media	12,545	1	38	,001
Resultados de aprendizajes armonizados	Se basa en la media	,742	1	44	,394
Porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo	Se basa en la media	,094	1	41	,760
gasto_educ_porhabitante	Se basa en la media	12,814	1	45	<,001
Índice de educación ajustado por desigualdad (0-100 el mejor)	Se basa en la media	,067	1	45	,796

En relación con la anterior prueba de homocedasticidad realizada (para los 5 clusters) podemos decir que siguiendo esta agrupación las varianzas de las variables cualitativas son bastante homogéneas ya que mantenemos la hipótesis para 3 de las 6 variables. Esta prueba nos demuestra que siguiendo esta distribución las varianzas de las diferentes variables para el cluster países europeos son bastante similares a las varianzas para el cluster países Latinoamericanos. Igualmente deberemos estudiar el resto de los análisis para llegar a conclusiones más precisas.

Descriptivo:

		N	Media	Dev. estándar
Niños sin escolarizar	América	10	4,210	3,9168
	Europa	37	2,265	2,3621
	Total	47	2,679	2,8312
Índice de educación ajustado por desigualdad (0-100 el mejor)	América	10	,610	,0994
	Europa	37	,797	,0897
	Total	47	,757	,1193
Niños por debajo del nivel mínimo (%)	América	10	48,860	13,7022
	Europa	30	5,743	6,3429
	Total	40	16,523	20,7546
Resultados de aprendizajes armonizados	América	10	421,260	23,5905
	Europa	36	511,239	29,0259
	Total	46	491,678	46,6328
Porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo	América	8	56,750	19,3095
	Europa	35	39,689	18,8531
	Total	43	42,863	19,8748
gasto_educ_porhabitante	América	10	583,8447	272,99115
	Europa	37	1828,2850	1507,42435
	Total	47	1563,5105	1434,55710

Contraste de diferencia de medias:

		F	Sig.
Niños sin escolarizar	Entre grupos	3,955	,053
Índice de educación ajustado por desigualdad (0-100 el mejor)	Entre grupos	32,813	<,001
Niños por debajo del nivel mínimo (%)	Entre grupos	185,482	<,001
Resultados de aprendizajes armonizados	Entre grupos	80,818	<,001
Porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo	Entre grupos	5,289	,027
gasto_educ_porhabitante	Entre grupos	6,652	,013

Atendiendo a los resultados del contraste de diferencia de medias podemos afirmar que pertenecer a un país Europeo o Latinoamericano influye en los resultados obtenidos en el ámbito educativo y en la importancia que se le da a la educación en dichos países. Esta conclusión se cumple para todas las variables analizadas, a excepción de los niños sin escolarizar que con un p-valor de 0,053 se encuentra en el límite. De manera general podemos afirmar que este resultado nos aporta fiabilidad al análisis.

Observamos como para las variables en las que una media inferior denota unos mejores resultados/ más importancia a nivel educativo (niños sin escolarizar, porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo y niños por debajo del nivel mínimo) son los países europeos los que obtienen esa media más baja. Ocurre lo mismo en el caso opuesto. Para las variables gasto en educación por estudiante, índice de educación ajustado por desigualdad y resultado de aprendizajes armonizados es el cluster “Europa” el que obtiene mayores medias, lo cual es sinónimo de mejores resultados/ más importancia.

Como conclusión general podemos afirmar que existe un modelo claramente diferenciado para los países europeos y otro para los países latinoamericanos. Además, estos modelos son coherentes con las expectativas previas a la realización del análisis.

3. ANÁLISIS CLUSTER BASADO EN LAS MEDIDAS DE DESIGUALDAD ECONÓMICA Y SOCIAL.

Para poder responder a las preguntas planteadas para el presente trabajo es necesario realizar una agrupación de los países en función de las semejanzas o diferencias

entre los países en cuanto a las variables que representan la desigualdad económica y social. Para ello, vamos a llevar a cabo un Análisis cluster.

3.1 Metodología del análisis cluster.

Se trata de que los países semejantes entre sí se hallen en un mismo grupo/cluster (homogeneidad interna) y los grupos difieran unos de otros (heterogeneidad externa). Existen distintos métodos de agrupación que se clasifican en: técnicas jerárquicas y técnicas de optimización. En nuestro caso concreto vamos a realizar la agrupación en base a las técnicas jerárquicas.

En este método los grupos se van formando en sucesivas etapas por unión (caso ascendente) de grupos ya existentes (los diferentes países), de modo que si las observaciones se han unido permanecerán en ese mismo estado hasta el final del proceso. Estas uniones se realizarán en función a la distancia que haya entre los diferentes países, uniéndose así los más parecidos (con distancias menores) y formando un nuevo grupo al que se le pueden seguir uniendo otros países en etapas posteriores.

Existen diferentes aspectos teórico-prácticos que debemos tener en cuenta a la hora de realizar nuestro análisis cluster:

La influencia de las unidades de medida de las variables en los resultados. Es decir, no representa lo mismo una unidad de la variable relativa a tasa de niños no escolarizados o de la variable renta media del país. Para equiparar estas unidades se recomienda tipificar las variables.

La estabilidad de la solución. Existen variantes en el proceso del análisis del cluster como puede ser el método de agrupación en clusters (enlace entre grupos, enlace dentro de grupos...) o la medida para establecer la distancia que hay entre los grupos (distancia euclídea, distancia euclídea al cuadrado...). Los grupos son más consistentes y por lo tanto la solución es estable cuando, aunque cambien las diferentes variantes en el procedimiento de agrupación, el resultado final es similar. En nuestro caso, se agrupan los mismos países, aunque se modifiquen las condiciones iniciales. Es decir, la solución presentada es robusta.

No existe una regla fija a la hora de establecer el número de grupos en la solución final, dependerá del analista y de los objetivos perseguidos.

Tan importante es identificar los grupos como caracterizarlos, es decir observar y definir como son y qué es lo que les diferencia de los otros grupos.

Concretamente en nuestro análisis vamos a utilizar como base el método de agrupación mediante enlaces entre grupos y para el intervalo la distancia euclídea al cuadrado. Al analizar la estabilidad de la solución cambiaremos estos parámetros.

3.2 Resultados del análisis.

Realizamos un primer análisis con las 7 variables relativas a la desigualdad obteniendo así diferentes agrupaciones de las cuales la agrupación en 6 clusters nos pareció la más interesante dado el número de países perteneciente a cada cluster, decidiendo que no tendríamos en cuenta el 6º cluster ya que está compuesto únicamente por un país (Perú).

Para poder describir gráficamente la clasificación obtenida debemos recordar que un país con poca desigualdad tendrá un índice de Gini por debajo de la media, la tasa que representa el 40% inferior y el 10% superior de participación estará por encima de la media al igual que la tasa relativa a la relación entre el 50% inferior y el 50% superior de participación laboral en los ingresos, la renta media del 40% inferior (% de la renta media nacional) y la relación entre tasa de participación mujeres y hombres en la fuerza laboral.

Debemos recordar que a lo largo del análisis cluster vamos a hablar de variables con tasas positivas o negativas, esto nos indica si está por encima o por debajo de la media ya que hemos realizado una tipificación de las variables como mencionamos anteriormente.

Por último, queremos mencionar que vamos a omitir el cluster 6 por estar formado por un único país (Perú), cluster relativamente parecidas al 5 con la diferencia de que la tasa de trabajadores en empleos vulnerables es superior y una tasa en la relación entre la participación de mujeres y hombres en la fuerza laboral positiva (por encima de la media), mientras que en el cluster 5 este valor es negativo.

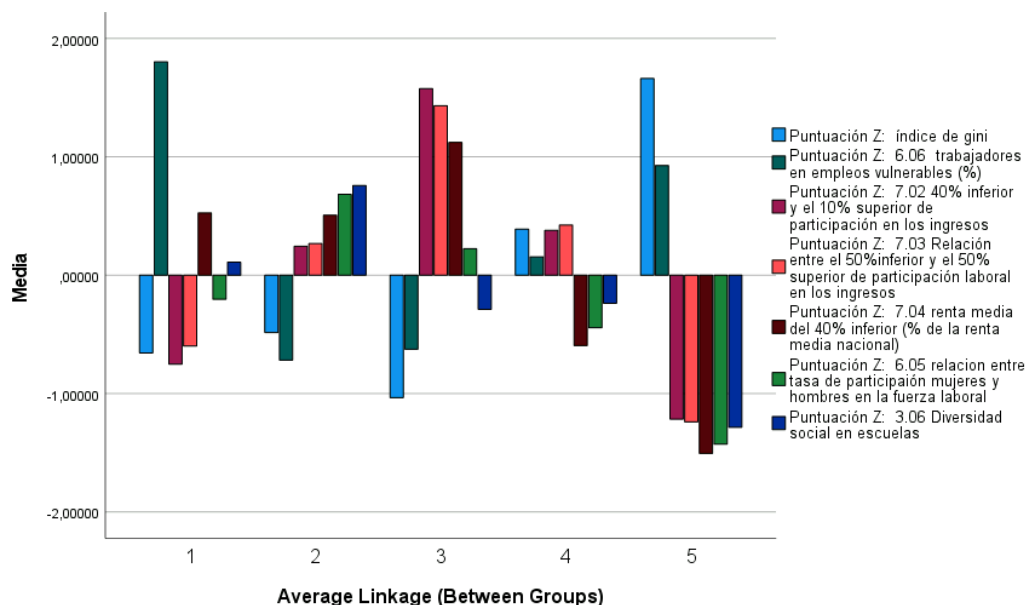
LOS 5 CLUSTERS FORMADOS SERÍAN:

Tabla 4: composición de cada cluster.

Cluster 1 (4)	Cluster 2 (22)	Cluster 3 (5)	Cluster 4 (7)	Cluster 5 (9)
Albania Armenia Georgia Moldavia	Alemania, Austria, Chipre, Croacia, Dinamarca, España, Estonia, Finlandia, Francia, Irlanda, Islandia, Letonia, Lituania, Noruega, Países Bajos, Polonia, Portugal, Reino Unido, Rusia, Ucrania, Suecia, Suiza.	Bélgica Eslovaquia Eslovenia Hungría Republica Checa	Bulgaria Grecia Italia Rumania Serbia Argentina Uruguay	Turquía Brasil Chile Colombia Costa Rica Ecuador México Panamá Paraguay

A continuación, pasamos a describir los 5 clústers principales teniendo en cuenta las características de los indicadores:

Gráfico 1: Representación de los clusters



Hemos ordenado las variables de tal forma: las dos primeras (índice de Gini y trabajadores en empleos vulnerables) son las únicas variables en las que un valor positivo refleja más desigualdad. El resto, unos valores positivos representan países más equitativos. De esta manera el gráfico es más visual.

CLUSTER 1: ATÍPICO.

Destaca por ser el cluster con más trabajadores en empleos vulnerables. Únicamente cuenta con valores que le acercan a ser un país igualitario en la variable “renta media del 40% inferior” y “diversidad social en las escuelas”. Este cluster está compuesto por 4 países europeos. En general son países desiguales, pero con un perfil relativamente atípico.

CLUSTER 2: IGUALITARIO – EQUITATIVO ECONÓMICO Y SOCIALMENTE.

Observamos valores negativos para las dos primeras variables lo cual representa un país igualitario. A su vez, las restantes variables se sitúan por encima de la media siendo la diversidad social la que mayor valor adquiere, están en coherencia con las dos primeras variables ya que ahora un valor por encima de la media implica estar formado por Países igualitarios.

También debemos tener en cuenta que el cluster 2 está formado por 22 del total de 48 países, lo cual explica que la distancia de las variables respecto de la media sea inferior a las de los otros clusters.

CLUSTER 3: REPARTO DE RENTAS, NO DE OPORTUNIDADES.

El presente cluster puede parecernos en un primer momento un modelo similar al cluster 2. Sin embargo, el presente cluster es más igualitario en cuanto al reparto de riquezas que el cluster 2 pero menos igualitario en lo social. Las 4 medidas de desigualdad económica toman valores más extremos que en el cluster 2. El reparto de rentas se acerca más a la equidad, además es el cluster que toma un valor más negativo en el índice de Gini.

Sin embargo, en cuanto a las medidas de desigualdad social los valores son más modestos. Participación mujeres en el mercado laboral y los trabajadores con empleos vulnerables muestran valores menos igualitarios que en el cluster 2 y la desigualdad en las escuelas se sitúa en valores positivos. En definitiva, muestra un perfil de países en los que la riqueza económica está mejor repartida, pero las oportunidades no lo están.

No obstante, debemos tener en cuenta lo mencionado anteriormente y es que la media se encuentra definida en gran medida por los países que forman el cluster 2 por estar formado por prácticamente la mitad de los Países de la muestra.

CLUSTER 4: PAÍSES EN TRANSICIÓN.

A modo general las variables toman valores relativos a un país desigual. Las dos primeras variables por encima de la media reflejando mayor desigualdad y las 3 últimas por debajo manteniendo lo reflejado por las 2 primeras. Las únicas variables que reflejan valores relativos a un país igualitario son “Relación entre el 40% inferior y el 10% superior de la participación en los ingresos laborales” y “Relación entre el 50% inferior y el 50% superior de participación laboral en los ingresos” tomando valores por encima de la media.

Sin embargo, debemos tener en cuenta también que ninguna de las variables toma valores excesivamente extremos. Es más, todas sus variables tienen valores muy cercanos a la media.

Podemos decir que este cluster está formado por los países más desiguales dentro de los países europeos y los más igualitarios de los países latinoamericanos.

CLUSTER 5: DESIGUAL EN TODO.

Es el cluster que más destaca por la desigualdad que reflejan sus valores en todas las variables. Este cluster refleja el ideal de un país desigual. Las dos primeras variables toman valores positivos y el resto de variables negativos. Podríamos decir que este cluster toma los valores inversos del cluster 2.

A modo general destaca la variable diversidad social en las escuelas por ser la variable que menos sigue una coherencia respecto de los diferentes clusters. Es la que se comporta de un modo más independiente respecto al resto de variables.

A su vez, podemos decir que los países que muestran desigualdad económica suelen mantener esa desigualdad en las variables relativas a la desigualdad social. Ocurre lo mismo con los clusters compuestos por Países más igualitarios, excepto en el cluster 3 que presenta una adecuada repartición de la riqueza económica pero no tanto de las oportunidades.

3.3 La posición de España.

Como se ha comentado anteriormente España se clasifica en el cluster 2 compuesto por 22 países europeos con perfil predominantemente igualitario tanto en los aspectos económicos como en los sociales.

La tabla adjunta muestra las puntuaciones de España en los 7 indicadores de desigualdad y las puntuaciones medias dentro de su cluster, lo que nos permite describir su posición a nivel global y dentro de su cluster.

Tabla 5: comparación de España con su cluster de pertenencia

	Puntuación España	Puntuación media Cluster 2
Gini	34,9	30,64
Trabajadores vulnerables	11,3	9,1
40 inferior /10 superior	62,4	64,5
50 inferior /50 superior	32,3	34,18
40 inferior /media	43,5	51,72
Mujeres/hombres	81,8	83,48
Diversidad social escuelas	75,8	80,34

Como puede observarse España presenta más indicios de desigualdad en todos los indicadores en comparación con la media de su cluster. En otras palabras, aunque España pertenece al cluster de los países más igualitarios no está entre los mejores, tanto si nos referimos a la desigualdad económica como a la social.

3.4 La movilidad social en los clusters descritos.

La desigualdad económica y social no solo es un indicador de la diferencia de oportunidades que tienen los ciudadanos. Es también un obstáculo para la movilidad social. En un país desigual la pertenencia a las clases sociales más bajas se perpetúa durante generaciones. Existen indicadores de movilidad social que miden la facilidad para cambiar de clase social.

En general el concepto de movilidad social se refiere al cambio en el estatus social, económico o profesional de un individuo o grupo a lo largo de su vida, en comparación con la posición que ocupaban originalmente, generalmente en relación con sus padres o la generación anterior. Este cambio puede manifestarse en ascensos o descensos en la escala social, y la movilidad social puede ser ascendente, descendente o lateral, dependiendo de la dirección del cambio en la posición social. La movilidad social es un indicador importante para comprender la dinámica de igualdad de oportunidades y la distribución de recursos en una sociedad.

Las economías con mayor movilidad social proporcionan oportunidades más equitativas, es decir, un terreno igual y meritocrático independientemente del origen socioeconómico, ubicación geográfica, género u origen. En este apartado vamos a tratar de demostrar la existencia de una asociación directa entre la desigualdad económica y social de un país y su puntuación de movilidad social en el índice. Una baja movilidad social afianza las desigualdades históricas, y mayores desigualdades de ingresos alimentan una menor movilidad social. Mejorar la movilidad social puede convertir este círculo vicioso en uno virtuoso y tener beneficios positivos en el crecimiento económico general.

El puntaje general del Índice Global de Movilidad Social es el promedio de los puntuaciones de los 10 pilares descritos en los apartados 2.2.1 y 2.2.2. Cada pilar, por lo tanto, representa el 10% del puntaje general.

Para los indicadores individuales, antes de la agregación, los valores en bruto se transforman en una puntuación de progreso que varía de 0 a 100, siendo 100 el estado ideal

Por lo tanto, un índice cercano a 100 representa una mayor movilidad social y un índice cercano a 0 representa un país con nula movilidad social.

En el caso concreto de España contamos con un índice de movilidad social de 70 puntos, algo superior aunque cercano al valor que toma la media de todos los países analizados que es de 67,38.

En el presente apartado vamos a tratar de contrastar la siguiente hipótesis de igualdad de medias .

H0: la media del Índice Global de Movilidad Social es igual para los 5 clusters).

En el caso de mantenerse la hipótesis implicaría que el pertenecer a un país más desigual o más igualitario no afecta a la movilidad social del propio país.

Descriptivos:

	N	Media	Desv. estándar
1	4	56,1750	2,41988
2	22	75,1909	7,09009
3	5	73,1000	5,85021
4	7	63,1857	3,65031
5	9	53,3667	4,73022
Total	47	67,3830	10,80218

Contrastes de igualdad de medias:

CLU6_1	N	Subconjunto para alfa = 0.05		
		1	2	3
5	9	53,3667		
1	4	56,1750	56,1750	
4	7		63,1857	
3	5			73,1000
2	22			75,1909
Sig.		,759	,156	,804

De los resultados del análisis se concluye que pertenecer a un cluster u otro se asocia con distintos niveles de movilidad social.

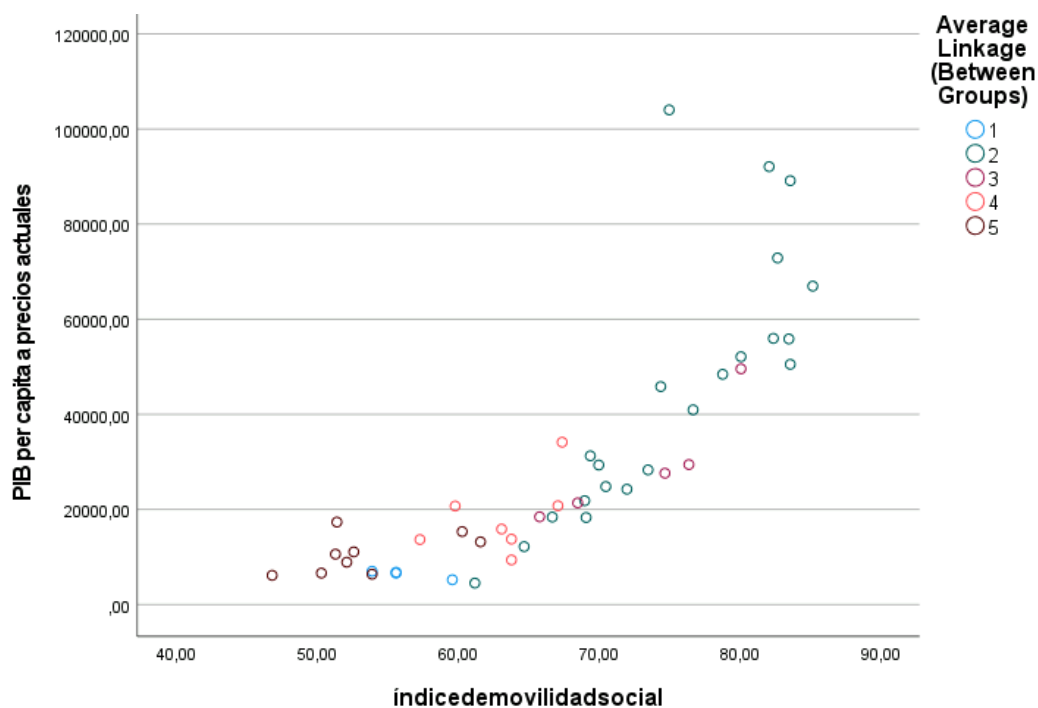
El cluster 5, formado por los países más desiguales, es el que muestra una peor movilidad social con una media de 53,37 seguido del cluster 1 con una puntuación media de 56,17. En el lado opuesto se encuentran los clusters 2 y 3 con las mejores movilidades sociales con una puntuación de 75,19 y 73,1 respectivamente. Por último, el cluster 4 se encuentra en el punto medio con una media de 63,19.

En el siguiente gráfico observamos de manera más visual la relación existente entre el PIB per cápita de cada país, agrupados estos a su vez en los diferentes clusters organizados por colores, y el índice de movilidad de estos.

Observamos como se forma una posible línea de puntos con pendiente positiva, lo cual nos indica, como es lógico, que existe una relación directa entre el aumento del PIB per cápita y el aumento de la movilidad social.

A su vez, podemos ver como los colores, es decir, los diferentes países pertenecientes a un mismo cluster, se organizan de manera bastante homogénea. Siendo el cluster 1 y 5 los que peor movilidad social y un PIB per cápita más bajo representan. El siguiente sería el cluster 4 que sus países se colocan entre los relativos al cluster 1 y los relativos al cluster 3. Por último y liderando la gráfica se sitúan los países relativos al cluster 2.

Gráfico 2: Relación entre PIB per cápita e índice de movilidad social



4. RELACIÓN ENTRE DESIGUALDAD Y EDUCACIÓN.

En este apartado abordamos la posible relación entre la pertenencia a cierto cluster y la importancia y resultados en educación. En otras palabras, vamos a tratar de comprobar si el pertenecer a un cluster u otro explica los resultados y la importancia que se le da en esos clusters a la educación.

Hemos organizado los clusters en función de unos criterios de desigualdad tanto social como económica. Por lo tanto, el fondo de la cuestión es averiguar si es cierto que la educación (tantos resultados como importancia) viene influida por la desigualdad (tanto económica como social) de cada país.

4.1 Metodología del ANOVA de un factor.

El Análisis de la varianza es una técnica de dependencia cuyo objetivo es detectar el posible efecto de ciertas variables de tipo cualitativo (que denominamos factores) sobre otras variables cuantitativas (que denominamos variables dependientes o variables respuesta)

En el presente trabajo nos encontramos con una serie de variables cuantitativas, 6 concretamente (niños sin escolarizar, índice de educación ajustado por desigualdad, niños por debajo del nivel mínimo, resultado aprendizajes armonizados, porcentaje de estudiantes desfavorecidos en escuelas con falta de material educativo, gasto público en educación).

Estas variables cuantitativas son nuestras variables “dependientes” y nos preguntamos si la variable cualitativa de perteneces a alguno de los 5 clusters seleccionados las explica.

La hipótesis nula en cualquier análisis de la varianza es que la media de la variable dependiente es la misma en todos los niveles del factor. Para resolver esta cuestión tenemos que plantear el mismo número de hipótesis que de variables cuantitativas que tengamos, en nuestro caso, 6. Por lo tanto, las hipótesis a contrastar en todo análisis de la varianza son las siguientes:

H0: la media de la variable dependiente es la misma en todos los niveles del factor: (media de niños sin escolarizar) cluster 1 = (media de niños sin escolarizar) cluster 2 = ... = (media de niños sin escolarizar) cluster 5.

H1: La media de la variable dependiente es diferente en alguno de los niveles del factor

Nos interesará rechazar las hipótesis ya que esto se interpreta como que el factor cualitativo tiene influencia sobre la variable dependiente cuantitativa y que alguna media es diferente. En cambio, si no rechazamos la hipótesis significaría que el pertenecer a un cluster o a otro no afecta a la tasa de niños sin escolarizar (la media es la misma en todos los niveles del factor).

El estadístico de contraste (para N observaciones y k niveles del factor) compara la variabilidad creada por los niveles del factor (SCE) con la variabilidad interna de los datos (SCR).

$$F = \frac{SCE/(k - 1)}{SCR/(N - k)}$$

En definitiva, comprobados los supuestos y conocida la distribución de probabilidad del estadístico (F), se realiza el contraste de hipótesis ANOVA y en función del p-valor obtenido se rechaza o se mantiene la hipótesis nula.

El resultado permite conocer si todas las medias son iguales o no, pero sin conocer cuál o cuáles de los niveles del factor tienen medias diferentes. Con la finalidad de

averiguarlo, una vez determinada que alguna media es diferente, se puede realizar una prueba post-hoc. Este tipo de prueba muestra subconjuntos homogéneos cuyas medias no difieren entre sí.

Por otro lado, para poder aplicar esta técnica, existe una propiedad que es deseable en los datos. Esta propiedad consiste en la homogeneidad de la varianza (homocedasticidad). Esto se debe a que no se puede realizar un ANOVA fiable si la varianza de una de las variables cualitativas es muy grande y la de otra de las variables cualitativas es muy pequeña. Para comprobar esta propiedad realizaremos un contraste de hipótesis del “test de Levene”, donde la hipótesis que se contrasta es la siguiente:

$$H_0: (\text{Var}X) \text{ cluster1} = (\text{Var}X) \text{ cluster 2} = \dots = (\text{Var}X) \text{ Cluster 5}$$

En este contraste es preferible no rechazar la hipótesis, ya que esto indicaría que no hay una diferencia significativa entre las diferentes varianzas y por lo tanto demostraría la existencia de homogeneidad en la varianza. Como hemos dicho anteriormente, este test sirve para poder comprobar la idoneidad de los datos.

4.2 Prueba de homogeneidad de las varianzas.

Observando los resultados relativos a la homogeneidad de las varianzas podemos decir que de manera general no hay demasiada homogeneidad en los datos. Mantenemos la hipótesis en 2 de las 6 variables analizadas. Esta distribución es algo menos homogénea que la realizada anteriormente atendiendo a criterios continentales.

Los datos en conjunto son los mismos, sin embargo, el cambio producido en la prueba de homocedasticidad se explica por la alteración en la distribución de estos datos. Hemos pasado de agrupar los datos en 2 clusters a hacerlos en 5. Tendremos que ver los resultados generales del análisis para llegar a una conclusión más precisa.

Pruebas de homogeneidad de varianzas

		Estadístico de Levene	gl1	gl2	Sig.
Niños sin escolarizar	Se basa en la media	10,382	4	42	<,001
Niños por debajo del nivel mínimo (%)	Se basa en la media	11,614	4	35	<,001
Resultados de aprendizajes armonizados	Se basa en la media	4,712	4	41	,003
Porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo	Se basa en la media	1,053	4	38	,393
gasto_educ_porhabitante	Se basa en la media	6,674	4	42	<,001
Índice de educación ajustado por desigualdad (0-100 el mejor)	Se basa en la media	,935	4	42	,453

4.3 Resultados de los análisis.

A continuación, mostramos los principales resultados del análisis realizado sobre el ANOVA de un factor, con su análisis post-hoc respectivo y su gráfico.

4.3.1. Variables de importancia de la educación.

Niños sin escolarizar

Descriptivos:

		N	Media	Desv. estándar
Niños sin escolarizar	1	4	4,750	4,1893
	2	22	1,391	1,4149
	3	5	1,740	1,1632
	4	7	3,086	2,9317
	5	9	5,111	3,6157
	Total	47	2,679	2,8312

ANOVA:

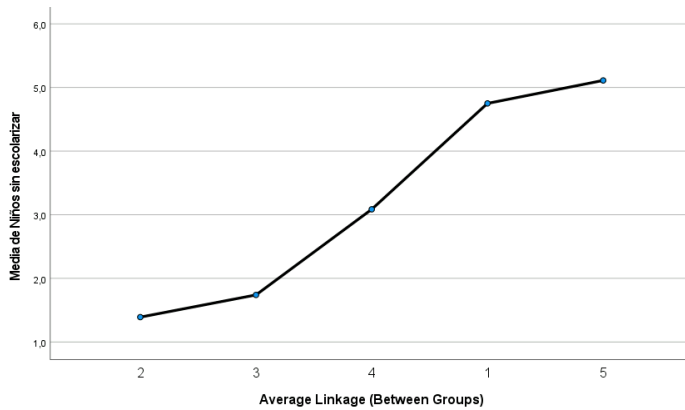
		ANOVA				
		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Niños sin escolarizar	Entre grupos	112,461	4	28,115	4,608	,004
	Dentro de grupos	256,258	42	6,101		
	Total	368,719	46			

Análisis post-hoc:

Ryan-Einot-Gabriel-Welsch F

Average Linkage (Between Groups)	N	Subconjunto para alfa = 0.05	
		1	2
2	22	1,391	
3	5	1,740	1,740
4	7	3,086	3,086
1	4	4,750	4,750
5	9		5,111
Sig.		,069	,082

Gráfico:



En primer lugar, observamos como la media total es de 2,7 niños sin escolarizar por cada 100 niños en edad de escolarización. Donde el valor mínimo de entre todos los datos es de 0 y el valor máximo de 10,8.

En lo que al ANOVA respecta rechazamos la hipótesis de que la media sea igual para todos los clusters. Es en el análisis post-hoc y en los gráficos donde se observan como se distribuyen las diferencias. Debemos tener en cuenta que en la presente variable un porcentaje bajo representa mayor importancia para la educación en los respectivos clusters.

Mientras que el cluster 2 es el que representa un menor porcentaje de niños sin escolarizar el cluster 5 es el que representa el mayor porcentaje lo cual se traduce en una menor importancia en la educación.

Porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo

Descriptivo:

Porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo	1	3	61,267	11,9709
	2	21	33,052	18,6974
	3	5	46,280	11,3590
	4	7	49,543	14,5403
	5	7	55,286	22,5057
Total	43	42,863	19,8748	

ANOVA:

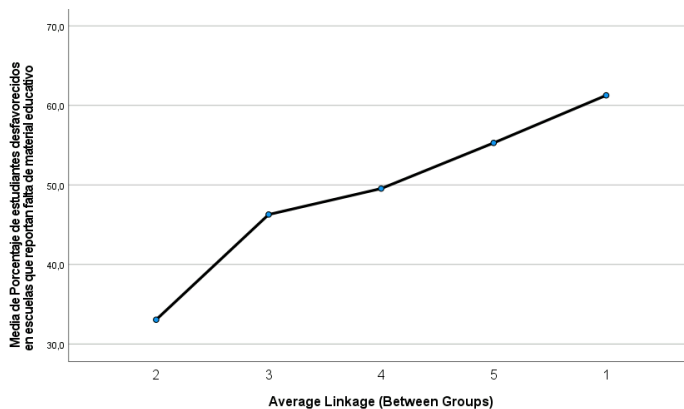
Porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo	Entre grupos	4488,288	4	1122,072	3,523	,015
	Dentro de grupos	12102,113	38	318,477		
	Total	16590,400	42			

Análisis post-hoc:

Ryan-Einot-Gabriel-Welsch F

Average Linkage (Between Groups)	N	Subconjunto para alfa = 0.05	
		1	2
2	21	33,052	
3	5	46,280	46,280
4	7	49,543	49,543
5	7		55,286
1	3		61,267
Sig.		,120	,643

Gráfico:



En relación con el ANOVA volvemos a rechazar la hipótesis de que todas las medias de los diferentes clusters sean iguales.

En la actual variable destacamos la gran diferencia que se observa entre las medias de los cluster 1 y 2 donde el cluster 1 tiene una media de casi el doble a la del cluster 2. Esto se observa tanto en el análisis post-hoc donde son los clusters que claramente están diferenciados (el cluster 5 se asemeja mucho al 1 y también está diferenciado del resto en este análisis) como en el gráfico.

En este caso al igual que con la variable anterior, un valor bajo refleja una mayor importancia en la educación de los países de los clusters. Concretamente, el cluster 2 es el que más importancia en educación representa y el cluster 1 y 5 los que menos.

Gasto público en educación por habitante

Descriptivo:

gasto_educ_porhabitante				
1	4	245,1538	62,24968	
2	22	2467,5566	1542,53699	
3	5	1632,1308	1012,93066	
4	7	785,5648	368,91687	
5	9	506,5027	254,34948	
Total	47	1563,5105	1434,55710	

ANOVA:

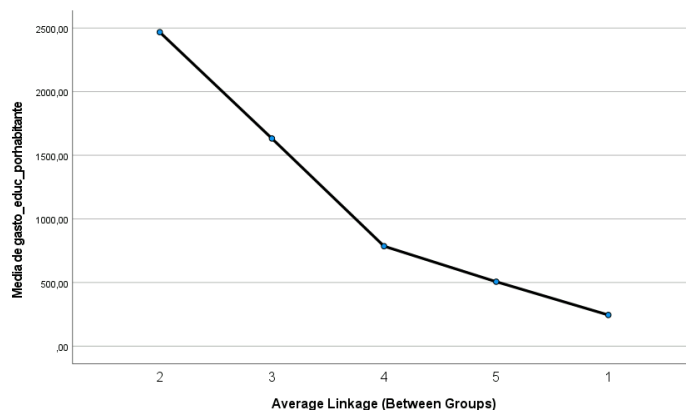
gasto_educ_porhabitante	Entre grupos					
	Dentro de grupos	39248172,995	4	9812043,249	7,436	<,001
	Total	55417714,230	42	1319469,386		
		94665887,224	46			

Análisis post-hoc:

Ryan-Einot-Gabriel-Welsch F

Average Linkage (Between Groups)	N	Subconjunto para alfa = 0.05	
		1	2
1	4	245,1538	
5	9	506,5027	
4	7	785,5648	
3	5	1632,1308	1632,1308
2	22		2467,5566
Sig.		,263	,333

Gráfico:



Se vuelve a rechazar la hipótesis de igualdad de medias. En esta variable a diferencia con las dos anteriores un alto valor representa mayor importancia en la educación. Destacan el cluster 2 con el mayor valor mientras que el 1 y 5 destacan por ser los clusters con valores más inferiores. Los cluster 3 y 4 toma valores intermedios

Volvemos a observar en los descriptivos al igual que comentamos en la definición de la variable que existe una gran diferencia entre las inversiones que se realizan en los diferentes países y por lo tanto también la diferencia existente entre los diferentes clusters, siendo la media del cluster 1 245,15 euros por habitante y la media del cluster 2 2467,56. Es decir en los países del cluster 2 se invierte en educación de media 10 veces más que la cantidad invertida en los países del cluster 1.

A modo general en las variables relativas a la importancia en la educación podemos decir que los clusters 2 y 3 son los que más importancia dan a la educación mientras que el cluster 1 es el que menos seguido del cluster 4 y 5.

4.3.2. Variables de resultados de la educación.

Índice de educación ajustado por desigualdad

Descriptivos:

Índice de educación ajustado por desigualdad (0-100 el mejor)	1	4	,725	,0500
	2	22	,827	,0827
	3	5	,840	,0548
	4	7	,729	,0488
	5	9	,578	,0667
	Total	47	,757	,1193

ANOVA:

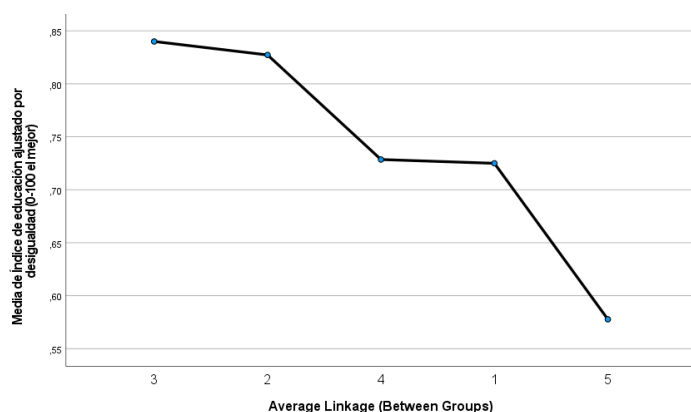
Índice de educación ajustado por desigualdad (0-100 el mejor)	Entre grupos	,442	4	,110	21,787	<,001
	Dentro de grupos	,213	42	,005		
	Total	,655	46			

Análisis post-hoc:

Ryan-Einot-Gabriel-Welsch F

Average Linkage (Between Groups)	N	Subconjunto para alfa = 0.05		
		1	2	3
5	9	,578		
1	4		,725	
4	7		,729	
2	22			,827
3	5			,840
Sig.		1,000	,999	,959

Gráfico:



De nuevo rechazamos la hipótesis de igualdad de medias. En este caso un alto valor representa mejor resultado. Siguiendo el análisis post-hoc las medias se dividen en 3 agrupaciones donde el cluster 5 es el que toma un valor menor, seguido del 1 y 4 que se colocan en la 2ª agrupación diferenciados del cluster 2 y 3 por ser los clusters con mayores valores. Concretamente el cluster 3 destaca por ser el cluster con el mejor valor (0,84).

Niños por debajo del nivel mínimo (%)

Descriptivo:

Niños por debajo del nivel mínimo (%)	1	2	21,750	11,6673
	2	18	3,394	3,1735
	3	5	4,260	1,5758
	4	6	20,633	21,5292
	5	9	45,689	17,6718
Total	40	16,523	20,7546	

ANOVA:

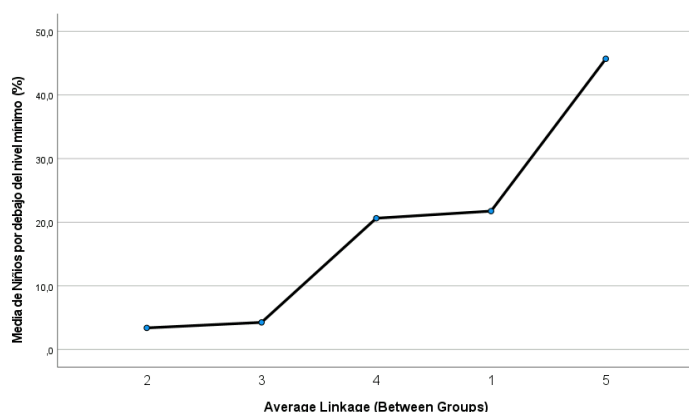
Niños por debajo del nivel mínimo (%)	Entre grupos	11666,221	4	2916,555	19,886	<,001
	Dentro de grupos	5133,129	35	146,661		
Total		16799,350	39			

Análisis post-hoc:

Ryan-Einot-Gabriel-Welsch F

Average Linkage (Between Groups)	N	Subconjunto para alfa = 0.05		
		1	2	3
2	18	3,394		
3	5	4,260	4,260	
4	6		20,633	
1	2		21,750	
5	9			45,689
Sig.		,996	,110	1,000

Gráfico:



Se vuelve a rechazar la hipótesis de igualdad de medias. En la presente variable un valor más bajo representa un mejor resultado. Destacan de nuevo los clusters 2 y 3 por ser los clusters con valores más bajos en ese orden mientras que el cluster 5 es el cluster que representa un mayor porcentaje de niños por debajo del nivel mínimo (45,69%). Además, esta vez el cluster 5 se encuentra claramente diferenciado del cluster 4 y 1 que venían destacando también por ser los clusters con peores resultados.

Resultados de aprendizajes armonizados

Descriptivo:

Resultados de aprendizajes armonizados	1	3	446,133	1,4012
	2	22	523,895	16,7089
	3	5	519,820	12,6338
	4	7	477,629	38,2979
	5	9	423,400	27,8284
	Total	46	491,678	46,6328

ANOVA:

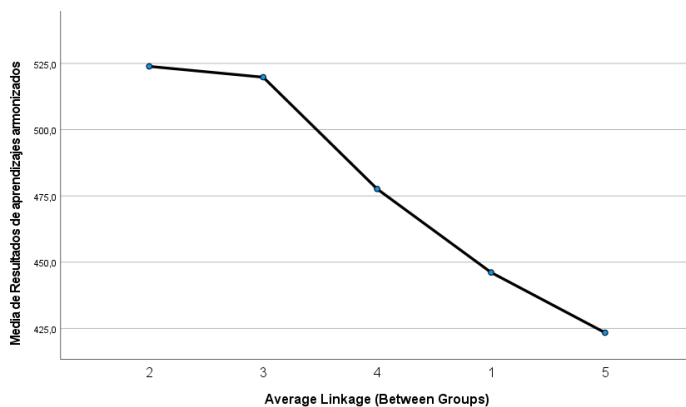
Resultados de aprendizajes armonizados	Entre grupos	76356,700	4	19089,175	36,401	<,001
	Dentro de grupos	21501,058	41	524,416		
	Total	97857,758	45			

Análisis post-hoc:

Ryan-Einot-Gabriel-Welsch F

Average Linkage (Between Groups)	N	Subconjunto para alfa = 0.05		
		1	2	3
5	9	423,400		
1	3	446,133	446,133	
4	7		477,629	
3	5			519,820
2	22			523,895
Sig.		,322	,127	,959

Gráfico:



En esta ocasión un mayor valor representa mejores resultados. Observando los descriptivos destaca de nuevo el cluster 5 por ser el que peores resultados obtiene (esto se observa también tanto en el análisis post-hoc como en el gráfico) y los cluster 2 y 3 por ser los que mejores resultados obtienen en ese orden. El cluster 1 sigue de cerca al cluster 5 con resultados también bajos al igual que el cluster 4, con la diferencia de que este se aleja algo más como se observa en el análisis post-hoc.

4.3.3 Conclusiones generales.

Como conclusión al análisis podemos decir que se rechaza que la media sea igual en los 5 clusters para las diferentes variables relativas a la educación

Ho: (media de niños sin escolarizar) cluster 1 = (media de niños sin escolarizar) cluster 2 = ... = (media de niños sin escolarizar) cluster 5.

Es decir, pertenecer a un cluster o a otro influye en la importancia que se le otorga a la educación y a los diferentes resultados que se obtienen en esta.

Sin embargo, con el ANOVA únicamente somos capaces de rechazar la hipótesis y de decir que la media para las diferentes variables no es igual para todos los clusters. Para saber que clusters son los que se diferencian, cuales se parecen entre sí y cómo son esas diferencias hemos acudido al análisis post-hoc.

Si bien es cierto que el cluster 5 destaca de manera general por ser el que peores resultados/menos importancia en educación tiene, en las variables relativas a la importancia en educación es el cluster 1 el que destaca en 2 de las 3 variables analizadas por ser el que menos gasto público invierte en educación y el que representa un mayor porcentaje de estudiantes desfavorecidos en escuelas que reportan falta de material educativo. No obstante, el cluster 5 le sigue de cerca siendo uno de los clusters que representa menos importancia en la educación.

En las variables relativas a los resultados en educación destaca el cluster 5 por ser el que peores resultados obtiene, además la diferencia con el cluster 1 se amplifica creando una mayor semejanza entre el cluster 1 y 4 que se encuentran más distanciados del cluster 5.

Por el contrario, los clusters 2 y 3 destacan por ser los clusters mejores parados en el análisis. Concretamente en las variables relativas a la importancia en educación el cluster 2 destaca en 2 de las 3 variables por ser el que más importancia obtiene al igual que en las variables relativas a los resultados obtenidos en educación donde el cluster 2 vuelve a destacar en 2 de las 3 variables. Destacando así en un cómputo global en 4 de las 6 variables por lo que podemos concluir estableciendo al cluster 2 como el cluster con mejores resultados y más importancia en educación. No obstante, debemos tener en cuenta que la diferencia con el cluster 3 es mínima en todas las variables; además en todos los análisis post-hoc los clusters 2 y 3 aparecen siempre juntos.

A modo conclusión podríamos establecer un ranking de los 5 clusters de peor posición a mejor en materia educativa que seguiría el siguiente orden:

Cluster 5 – Cluster 1 – Cluster 4 – Cluster 3 – Cluster 2.

Cabe destacar el lugar que ocupa España en este análisis. Si bien es cierto que como hemos comentado anteriormente todos sus valores de las variables se encuentran por debajo de la media del cluster al que pertenece debemos tener en cuenta que pertenece al cluster 2 siendo este el mejor cluster en materia de educación.

Por último, no debemos perder el hilo de las preguntas que mediante este análisis tratamos de resolver. Entre ellas, ¿Existe realmente una relación entre la desigualdad y la

educación? Pues bien, si acudimos a la descripción de los clusters en función de su desigualdad, vemos como el cluster 5 ha sido el cluster que hemos catalogado como el más desigual correspondiéndose a su vez con el cluster con peores resultados y menos importancia en educación. Los siguientes peores en el análisis relativo a la educación son el cluster 1 y 4. A su vez, en el análisis de los clusters son los siguientes más desiguales después del cluster 5. Por último, los clusters 2 y 3 son los clusters con países más igualitarios y con mejores resultados y más importancia en educación.

Otra de las cuestiones que nos planteábamos al comienzo del trabajo era la posible relación que podía existir entre la desigualdad, la educación y la movilidad social de un país. Acudiendo al apartado relacionado con la movilidad social vemos como, de nuevo, se repite el ranking establecido para la desigualdad de los países y para la educación. Por lo tanto, concluimos estableciendo una estrecha relación entre la desigualdad, la educación y la movilidad social de un país. Se ha respetado el mismo orden para los 3 análisis. A más desigualdad del cluster peores son los resultados, menos la importancia que se le da a la educación y peor es la movilidad social en los respectivos países del cluster.

5. BIBLIOGRAFÍA.

Aiyar, S., & Ebeke, C. (2019). *Inequality of Opportunity, Inequality of Income and Economic Growth*. IMF Paper WP/19/34. International Monetary Fund.

Anand, P., Behrman, J., Dang, H. A., & Jones, S. (2018). *Inequality of opportunity in education: Accounting for the contributions of Sibs, schools and sorting across East Africa*. World Bank Policy Research Working Paper, (8622).

Brunori, P., Ferreira, F. H. G., & Peragine, V. (2013). *Inequality of Opportunity, Income Inequality and Economic Mobility: Some International Comparisons*. IZA Discussion Paper No. 7155. Retrieved from <http://ftp.iza.org/dp7155.pdf>.

Global Social Mobility Index 2020. (2023, 9 noviembre). World Economic Forum. <https://www.weforum.org/publications/global-social-mobility-index-2020-why-economies-benefit-from-fixing-inequality/>

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2012). *Análisis multivariante*.

Peña, D. (2013). *Análisis de datos multivariantes*. Cambridge: McGraw-Hill España.

Snowflake data discovery. Knoema. <https://knoema.es/atlas?view=snowflake>

World Bank Open Data. (s. f.). World Bank Open Data. <https://datos.bancomundial.org/indicador/SI.POV.GINI>

Trabajo fin de grado/máster en estructura salarial: un análisis estadístico sobre la brecha de género, Elisa Irujo Izcue.

World Inequality Database. (2015-2016). Retrieved from <http://wid.world/data/>