

MÁSTERES de la UAM

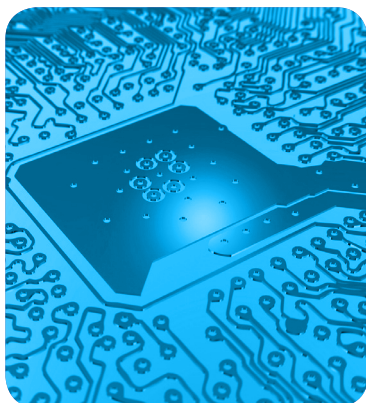
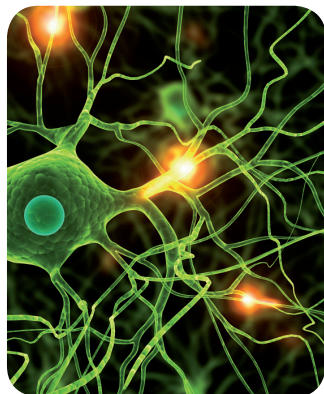
Facultad
de Ciencias Económicas
y Empresariales / 14-15

Desarrollo
Económico



**Rendimiento
académico del nivel
de primaria en
el Perú: un análisis
regional**

Rosana Zevallos Vargas





Rendimiento académico del nivel de primaria en el Perú: un análisis regional

Master en Desarrollo Económico y Políticas Públicas
Facultad de Ciencias Económicas y Empresariales

Trabajo de Fin de Máster

Alumna: Rosana Zevallos Vargas
Tutora: Profesora Dra. Marta Rahona López

Madrid, julio de 2015

Índice

Justificación.....	4
Marco Teórico	5
Metodología.....	9
Limitaciones.....	9
Justificación de las variables a emplear	10
Variable dependiente (VD).....	11
Variables independientes (VI's)	11
Hipótesis de trabajo	14
Ámbito del Análisis.....	14
Análisis estadístico de los años 2000 y 2013.....	15
Análisis descriptivo.....	15
Regresión Lineal Múltiple.....	16
Idoneidad del modelo.....	16
Test de Normalidad	18
Test de Linealidad.....	19
Homogeneidad o Igualdad de Varianzas (Homocedasticidad).....	20
Independencia de los errores.....	20
Análisis de la Correlación	21
Análisis de Clústers	23
Análisis de la Varianza de los Clústers	24
Análisis de Correlaciones a nivel de Clústers	25
Comprobación de Hipótesis	27
Conclusiones	28
Bibliografía	30

Índice de Tablas

Tabla 1. Indicadores de Pobreza Multidimensional (Scorecard)	4
Tabla 2. Matriz de Correlaciones, 2000 y 2013	10
Tabla 3. Comparativo de indicadores - Media, Perú 2000 y 2013.....	12
Tabla 4. Estadísticos descriptivos básicos de las Regiones, 2000 vs 2013.....	13
Tabla 5. Idoneidad de Modelo, 2000	15
Tabla 6. Idoneidad de Modelo, 2013	15
Tabla 7. Tests de Normalidad, 2000 y 2013	17
Tabla 8. Test de Shapiro-Wilk de normalidad, 2000 y 2013	17
Tabla 9. Nuevo test Shapiro-Wilk, 2000 y 2013.....	18
Tabla 10. Matriz de Correlación, 2000 y 2013	20
Tabla 11. Pertenencia a los clúster por regiones, 2000 y 2013	23
Tabla 12. Varianzas de los Clústers, 2000 y 2013	23
Tabla 13. Correlación de Clústers, 2000 y 2013.....	24

Rendimiento académico del nivel de primaria en el Perú: un análisis regional

Resumen

La educación es considerada como el factor determinante que le permite a la gente escapar de la pobreza. En países en vías de desarrollo con ingresos medios como lo es el Perú desde hace algunos años se vienen realizando importantes esfuerzos presupuestarios por mejorar la calidad educativa básica que se ofrece a todos los niños y jóvenes, luego de ampliar su cobertura y acceso como paso previo; sin embargo existen aún algunas barreras que no permiten que los resultados deseados se correspondan con la inversión y expectativas del gobierno. El presente trabajo demuestra que las políticas públicas en el Perú, a pesar del cambio de enfoque de pobreza usado al pasar de la monetaria a la multidimensional, siguen presentando carencias importantes que se manifiestan en los pobres rendimientos escolares. Usaremos para el análisis el enfoque multidimensional de variables de educación y condiciones de vida de la población y aplicando métodos estadísticos usaremos correlaciones y análisis de varianzas para evaluar 3 conglomerados regionales que se conforman en los años 2000 y 2013 en el Perú, así como los comportamientos de las variables para cada clúster en el tiempo.

Abstract

Education is considered to be a key factor that allows people to escape poverty. In recent years, middle-income developing countries like Peru have been making important budgetary efforts to improve the quality of the basic education offered to all children and young people, after expanding access and education coverage to all as a first step; however, there are some barriers still preventing the desired results from matching government investments and expectations. The present paper shows that public policies in Peru, despite the change of approach from monetary to multidimensional poverty, keep showing important gaps that translate into poor school performance. For the analysis we will use education and living conditions variables multidimensional approach, and through the application of statistical methods we will use correlations and variance analysis to evaluate 3 regional clusters that are formed for the years 2000 and 2013 in Peru, as well as the behaviour of each cluster's variables for those periods.

Justificación

Desde la década de los 90's el Perú ha realizado importantes esfuerzos para cumplir con el mandato constitucional de ofrecer acceso universal a la educación¹. La tasa de escolarización

¹ En la Constitución Política del Perú de 1993, el tema educativo de enseñanzas obligatorias (inicial, primaria y secundaria) está contemplado en los artículos 13 al 17. El artículo 16 dice que "Tanto el sistema como el régimen educativo son descentralizados", aunque es el estado central el que coordina las políticas a nivel nacional y se le asigna un rol prioritario en el presupuesto nacional, mientras que el artículo 15 señala que "La ley establece los requisitos para desempeñarse como director o profesor de un centro educativo, así como sus derechos y

inicial, primaria y secundaria, así como de alfabetización entre 1997 y el 2013 ha pasado de ser 45.9%/73.4%/56.4% respectivamente a 71.9%/89.5%/78.5% (GRADE, 2015) y de alfabetización de jóvenes entre 15 y 24 años de 98.7% y de población adulta de 93.84% en 2013 (INEI, 2015).

Uno de los objetivos de cualquier estado al estudiar la evolución de sus indicadores de progreso social es el de diseñar las políticas públicas más adecuadas que permitan mejorar las condiciones de vida de su población, además de verse “como fuente potencial de ventaja competitiva” (OECD, 2011). Para el caso peruano, recientemente, esos análisis se vienen haciendo desde un enfoque multidimensional y ya no sólo desde el enfoque de la pobreza monetaria. Uno de esos indicadores que más atención y preocupación recibe es el del rendimiento escolar.

A pesar de que el Perú no participa de manera recurrente en las evaluaciones internacionales PISA, el Ministerio de Educación, a través de su unidad de Medición de la Calidad Educativa ESCALE, viene realizando desde el año 2006 evaluaciones a los alumnos de segundo grado y sexto grado de educación primaria pública (Guadalupe, León, & Cueto, 2013).

Lo que se pretende analizar en este trabajo es la posible incidencia de una serie de variables de salud, educación y condiciones de infraestructura escolar sobre el rendimiento de los alumnos expresado en el indicador *tasa de aprobados*, al considerar que la tasa de alfabetización no es el objetivo final del estado sino un paso previo a lograr mejores resultados académicos.

Marco Teórico

Las formas en que se manifiesta los efectos de la pobreza son múltiples y muy complejos, sin embargo el enfoque que se le ha dado para analizar sus causas, determinar su impacto y diseñar políticas que ayuden a aliviarla y superarla ha sido mayoritariamente monetarios, ya que se consideraba que el bienestar económico era el punto de partida hacia el bienestar social. Esta visión cambió radicalmente hace pocos años cuando el premio Nobel Amartya Sen, desde el *Oxford Centre Oxford University's Poverty and Human Development Initiative*,

obligaciones. El Estado y la sociedad procuran su evaluación, capacitación, profesionalización y promoción permanentes” (Tribunal Constitucional del Perú, 1993).

y el Programa de Desarrollo de las Naciones Unidas, desarrollaron el enfoque de pobreza multidimensional y a partir de él un índice que facilita la identificación de áreas prioritarias para la asignación de recursos públicos y por lo tanto de políticas públicas (UNDP, 2014).

La metodología del MPI (en español: indicador de pobreza multidimensional), consiste en agrupar una serie de carencias en 3 grupos o dimensiones, a los que se le asigna un peso específico (óp. Cit.). El marco teórico utilizado define un hogar en pobreza multidimensional como aquel en que la carencia o privación sopesada supera 1/3 o 33% del total de los indicadores medidos (Kovacevic & Calderon, 2014).

Tabla 1. Indicadores de Pobreza Multidimensional (Scorecard)

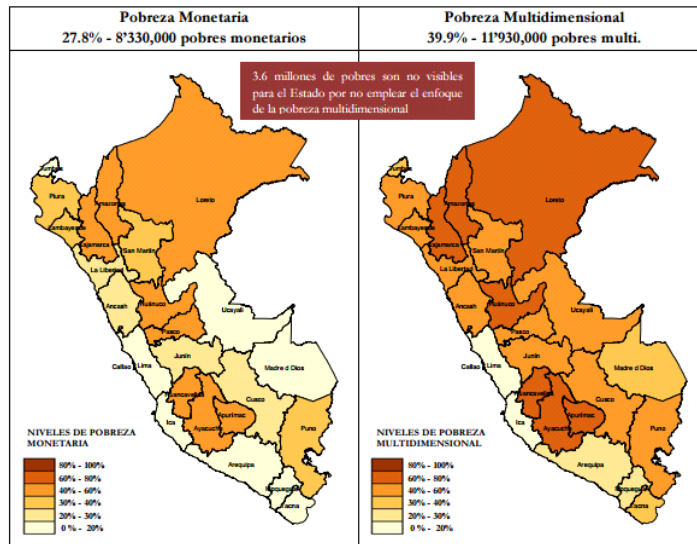
Dimensión	Variable	Peso
Educación	Asistencia a la escuela de niños en edad escolar (hogares con y sin niños en edad escolar)	1/6
	Nivel escolar alcanzado de los miembros del hogar	1/6
Salud	Mortalidad infantil en hogares con mujeres entre 15-49 años, con niños menores de 5 años.	1/6
	Déficit calórico y nutrición de hombres, mujeres y niños (peso, retraso de crecimiento, malnutrición)	1/6
Condiciones de vida	Acceso a electricidad	1/15
	Acceso a agua potable (menos de 30 minutos de camino a la fuente)	1/15
	Acceso a instalaciones sanitarias (desagüe)	1/15
	Uso de combustibles sólidos para cocina y calefacción	1/15
	Pisos de la vivienda terminados	1/15
	Acceso a medios de comunicación, transporte, artefactos electrodomésticos, tierras arables, otros.	1/15

Fuente: UNDP (2014) y Enrique Vásquez, El Perú de los pobres no visibles para el Estado, (2011).

En el caso del Perú, nos señala Vásquez que “el enfoque de pobreza multidimensional presenta una realidad distinta a la mostrada por el enfoque de pobreza monetaria. En concreto, para el 2012, la tasa de pobreza multidimensional es de 36.6%, mientras que, según la pobreza monetaria, la tasa es de 25.8%” (Vásquez, 2012; 2014).

Considerando este enfoque, el presente trabajo busca analizar los resultados de promoción de alumnos de primaria de las escuelas del sector público y la posible relación de algunas variables incluidas en la metodología de multidimensionalidad de la pobreza antes descrita.

Mapa 1. Pobreza, Monetaria y Multidimensional, según regiones: Perú, 2011



Fuente: Centro de Estudios de la Universidad del Pacífico – Perú (Vásquez, 2011)

La educación para la población peruana es considerada como fundamental para el desarrollo económico y social del conjunto, siendo la enseñanza primaria la “etapa del desarrollo es posible actuar positivamente y de manera eficaz en la formación de las personas”, lo que explica que todas las declaraciones, acuerdos y programas mundiales sobre este tema pidan su universalización, entendida además como “acceso y cobertura” (CEPAL & OEI, 2009).

Sabemos que el rendimiento de los estudiantes depende de una infinidad de causas, muchas de ellas difíciles de determinar y en muchos casos difíciles de medir. En un estudio sobre los predictores del rendimiento académico se señalaba que la multidimensionalidad de esos predictores seguía una clasificación: factores aptitudinales (como la inteligencia y coeficiente intelectual), factores de entorno (ambiente familiar, estimulación de los profesores, del entorno de clases, etc.), motivación, auto-concepto² y ambiente educativo (aceptación por parte de los compañeros, instalaciones, educación del padre, ocupación de los padres, etc.) (Sánchez-Cabezudo Sancho, J., 1986).

De las variables a considerar en el presente trabajo se utilizará como base la información estadística recogida por la Unidad de Estadística Educativa del Ministerio de Educación del Perú (MINEDU) y serán complementadas con datos proporcionados por el Instituto Nacional

² Autoimagen o autoestima., según se desprende del contexto de la investigación.

de Estadística e Informática (INEI), advirtiendo que considerando la clasificación de predictores de rendimiento antes mencionada, la encuesta y estadísticas extraídas NO contemplan los factores de entorno, motivación o ambiente. La idea es analizar primero el grado de correlación entre las variables que sí están incluidas en la encuesta con respecto a la variable dependiente que plantea el problema a investigar: tasa de alumnos aprobados de nivel primario en el Perú. Al tener como unidades de observación a 26 regiones y sólo 2 años de comparación (2000 y 2013), por motivos de manejo estadístico sólo se podrán 5 variables como máximo.

Además de esto, se empleará un análisis de clústers que tengan en cuenta el nivel de desarrollo de cada región. Esta metodología se fundamenta en varios estudios: el Observatorio Chileno de Política Educativas cita a otros estudios realizados desde los años 60's sobre la diferencia del peso de los factores que inciden en una educación efectiva en países en vías de desarrollo en comparación con los ya desarrollados, siendo éstos “el entorno familiar, la calidad de la comunidad donde residen y la efectividad de la escuela”.

En este análisis se muestra que es la varianza de esos factores la que cambia de ponderación conforme un país mejora sus indicadores socioeconómicos y por ende podría extrapolarse la metodología a nivel regional, especialmente cuando existen marcadas brechas entre regiones.

La OCDE también reconoce que existe “preocupación... por la existencia de niveles sustanciales de heterogeneidad en el desempeño escolar junto con diferencias significativas en resultados educativos para grupos reconocibles en la población [lo que haría que se] generen tensiones sociales y creen ineficiencias económicas” (OECD, 2011). Finalmente, el Grupo de Análisis para el Desarrollo del Perú, GRADE marca el mismo camino cuando señala que la distribución de la inversión y los recursos en la escuela deberían compensar las diferencias entre grupos de estudiantes; sin embargo, son los niños más ricos, urbanos y castellanohablantes quienes reciben más o mejores recursos” (GRADE, 2015).

Siendo el gasto en educación un factor clave para el desarrollo regional, no es necesariamente la cantidad de recursos empleados lo que explicará el desempeño escolar sino el uso apropiado de esos recursos y conocimiento de las necesidades específicas de cada zona.

Metodología

Se realizará un análisis estadístico de fuentes secundarias oficiales para extraer información sobre las variables de interés en forma de modelos estáticos para los años 2000 y 2013, e intentaremos comprobar los supuestos necesarios para efectuar regresiones multivariantes que nos indique la relevancia de cada variable en el rendimiento académico de los niños de primaria de las escuelas públicas del Perú. De no ser posible, que es lo que se asume, emplearemos el análisis de clústers o conglomerados y sus varianzas.

A partir de las tres dimensiones empleadas para medir la pobreza (educación como dimensión básica, salud y condiciones de la vivienda como dimensiones subyacentes), así como del total de indicadores con las respectivas ponderaciones considerados en el PMI, para este trabajo se han seleccionado variables que, debido a su importancia y efecto en los niveles de ingreso de la población y su movilidad social, podrían ser equiparables al modelo PMI y por lo tanto empleadas en un futuro – y con más datos para completar series temporales – para diseñar modelos que ayuden a proyectar los impactos de esas variables en los logros educativos de los niños y jóvenes estudiantes en el Perú.

Limitaciones

- Pocos años e información para las unidades observadas. Esto no obliga emplear un número limitado máximo de 5 variables.
- La mayoría de estudios realizados por el MINEDU y terceras organizaciones para el MINEDU no clasifican a la población por regiones políticas sino por regiones sociales (costa urbana, costa rural, sierra urbana... selva rural, Lima), por lo que muchas variables del INEI no se pueden utilizar.
- El gasto presupuestario por regiones no se encuentra desglosado por partidas, lo que no permite extraer información precisa.
- Inconsistencia en la clasificación utilizada al agrupar la región Lima. En este caso, algunos censos y encuestas la dividen en Lima Metropolitana/Lima Provincias/Lima y Callao, mientras algunos otros diferencian sólo Lima de Callao, sin niveles metropolitanos y provinciales. Para superar este escollo se ha procedido a calcular el promedio diferencial

entre las variables allí donde sí se discriminaban las regiones y aplicar el porcentaje diferencial para aquellas donde no se hacía esa separación.

- Series de datos incompletas. En el caso de las encuestas nacionales de educación, éstas empezaron siendo evaluaciones realizadas cada 4 años. Para cubrir los años 2000 y 2013 allí donde no existía información, se ha empleado la tendencia de la media, práctica comúnmente aceptada para el caso de encuestas periódicas no anuales anuales (Honaker & King, 2010).
- Encuestas de calidad de educación y estadísticas educativas no contemplan mediciones de variables tales como motivación del alumno, del maestro, nivel de participación de los padres o tutores, etc.

Justificación de las variables a emplear

Debido a las limitaciones de la información recolectada por las agencias evaluadoras de la calidad educativa, no es posible emplear la variable “Rendimiento Satisfactorio nivel primario” ni en el área de matemáticas ni en el de comprensión lectora como variables dependientes, por lo que estamos asumiendo que el superar el nivel primario (tasa de alumnos aprobados) equivale a adquirir los conocimientos y habilidades necesarios en cada área académica.

Como ya hemos mencionado, utilizaremos un enfoque multidimensional, ya que se incluyen elementos que también influyen en el desempeño y desarrollo escolar. Al respecto, la OECD indica que “con este enfoque en los resultados de aprendizaje, las medidas confiables y válidas de desempeño escolar pueden servir como base para políticas y programas educativos para la rendición de cuentas, los procesos de mejora escolar y para que familias y la sociedad en su conjunto cuenten con mayor información sobre la efectividad de escuelas” (OECD, 2011).

Con respecto al análisis estadístico, utilizaremos un modelo estático-comparativo y no de serie temporal, al no disponer de datos para todas las variables. El primer tipo análisis a realizar será el de correlaciones, para el cual se empleará el test de *Spearman Tau* al esperarse que muchas de las variables no sigan un comportamiento normal y existan *outliers* o

anormalidades considerables. Este test reduce la sensibilidad de la normalidad esos *outliers* (Shong Chok, 2010).

Como se puede observar en la tabla 2 a continuación, existen muchas variables de las consideradas en la encuesta de la Unidad de Estadística del MINEDU “ESCALE” que no están fuertemente relacionadas con la variable dependiente (VD) de tasa de aprobación de nivel primario. Se descartan desde ya las variables de salud 14 y 16 (tuberculosis y tasa de enfermedad en la población), y nos quedamos con las de nivel de vida y educación que se detallan a continuación.

Variable dependiente (VD)

Porcentaje de alumnos aprobados de educación primaria, dando por descontado que la tasa de aprobación de los alumnos es equiparable al concepto de “rendimiento satisfactorio alcanzado” para el último grado de educación primaria.

Variables independientes (VI's)

Según nuestra matriz de correlación, y descontando el efecto doble que existe entre algunas de las variables (acceso a electricidad, agua y desagüe, por ejemplo), analizaremos el comportamiento de las siguientes:

Var2 gasto de educación regional sobre el PIB regional. A pesar de no tener desglose del gasto por partidas a nivel regional, se asume comúnmente que existe un impacto significativo de esta variable en la VD. La Var2 mide el gasto corriente y de capital en instituciones educativas y administración de la educación realizado por el gobierno central y los gobiernos regionales y locales, expresado como porcentaje del producto bruto interno regional.

Var3 tasa de asistencia a nivel inicial o pre-escolar. Consideramos que el asistir o no a los grados de inicial (3 a 5 años) influye directamente en el rendimiento de los alumnos en nivel primario.

Var6 tasa de atraso escolar, que mide el porcentaje de alumnos que para su edad cronológica se encuentran cursando grados inferiores debido a repeticiones. Los datos no proporcionan las razones por las cuales los alumnos sufren ese atraso escolar.

Tabla 2. Matriz de Correlaciones, 2000 y 2013

Kendall's tau_b		Correlations			
		DPTE_ aprobados primaria			
		2000	2013	Var %	Fuerza Asociación
DPTE_ aprobados primaria	Corr. Coefficient	1.000	1.000	-	-
	Sig. (2-tailed)				
var1_gas toxalum no	Corr. Coefficient	.145	.138	4.26%	Débil
	Sig. (2-tailed)	.300	.321		
var2_educpib	Corr. Coefficient	-.464**	-.507**	9.27%	Fuerte
	Sig. (2-tailed)	.001	.000		
var3_tasasistenciainicial	Corr. Coefficient	.361**	.495**	37.12%	Moderada
	Sig. (2-tailed)	.010	.000		
var4_tasasistenciaprimaria	Corr. Coefficient	.009	.003	-66.87%	Muy Débil
	Sig. (2-tailed)	.947	.982		
var6_atrasoescolar	Corr. Coefficient	-.797**	-.723**	-9.28%	Fuerte
	Sig. (2-tailed)	.000	.000		
var9_accesointernetprimaria	Corr. Coefficient	.625**	.711**	13.76%	Fuerte
	Sig. (2-tailed)	.000	.000		
var10_profes titulados primaria	Corr. Coefficient	.083	-.175	-311.11%	Débil
	Sig. (2-tailed)	.552	.209		
var11_tasa analfechos 15 a 24	Corr. Coefficient	-.418**	-.305*	-27.03%	Moderada
	Sig. (2-tailed)	.003	.029		
var12_sueldopromedio	Corr. Coefficient	.567**	.305*	-46.21%	Moderada
	Sig. (2-tailed)	.000	.029		
var14_tasas TBC	Corr. Coefficient	.351*	.163	-74.96%	Débil
	Sig. (2-tailed)	.012	.243		
var15_desempleo	Corr. Coefficient	.446**	.360**	-19.28%	Moderada
	Sig. (2-tailed)	.001	.010		
var16_pobenfemedad	Corr. Coefficient	-.028	-.017	39.71%	Muy Débil
	Sig. (2-tailed)	.843	.907		
var19_prof titulados inicial	Corr. Coefficient	.046	.126	173.30%	Muy Débil
	Sig. (2-tailed)	.741	.366		
var20_sindesague	Corr. Coefficient	-.532**	-.446**	-16.17%	Moderada
	Sig. (2-tailed)	.000	.001		
Var21_electricidad	Corr. Coefficient	.657**	.662**	0.76%	Fuerte
	Sig. (2-tailed)	.000	.000		
Var23_tasas nec insatisfechos	Corr. Coefficient	-.477**	-.526**	10.27%	Fuerte
	Sig. (2-tailed)	.001	.000		
Var24_rendimiento TEXTO	Corr. Coefficient	.581**	.606**	4.30%	Fuerte
	Sig. (2-tailed)	.000	.000		
Var25_rendimiento MATE	Corr. Coefficient	.344*	.495**	43.90%	Moderada
	Sig. (2-tailed)	.014	.000		

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Elaboración: propia

Var9 Porcentaje de escuelas que cuentan con acceso a Internet, primaria (% del total) – Según algunos estudios, los factores que impactan el rendimiento académico cuando existe acceso a las TIC’s son aquellos que se desarrollan alrededor del acceso al equipo informático y a la internet³, sin embargo, con la inclusión de esta variable buscamos también entender el

³ Los estudios considerados demuestran que el efecto de tener acceso a un ordenador en el rendimiento escolar es mayor que el de tener acceso a internet – caso de Brasil (Wainer, Vieira, & Melguizo, 2014); que en realidad es el ambiente de estudios provisto en casa y el nivel de involucramiento de los padres en los estudios de sus hijos lo que afecta positivamente el resultado de los alumnos con acceso a tecnologías de informática y comunicaciones – caso de los Países Bajos (Aesaert, y otros, 2015), o que son las capacidades del profesor con

impacto indirecto de la inversión en banda ancha y en equipamiento informático en este aspecto específico sobre la tasa de alumnos aprobados.

Var23 Tasa de necesidades insatisfechas (al menos 1 necesidad insatisfecha). Puesto que el nivel de ingresos varía mucho entre regiones lo mismo el coste de vida, y estamos tratando de discriminar pobreza monetaria de multidimensional, usaremos este indicador que mide los siguientes factores: viviendas inadecuadas por sus materiales, hacinamiento crítico por habitación, falta de servicios para la eliminación de excretas, inasistencia a escuelas primarias de los menores (que no es igual a atraso escolar), y la capacidad económica del hogar, que asocia el nivel educativo del Jefe del hogar con la tasa de dependencia económica.

Con respecto a las variables que dejamos fuera, 3 que llaman especialmente la atención por el efecto a priori esperado y no cumplido, son:

Var10 y Var19, tasa de profesores de educación primaria y de inicial con títulos de docentes – En el año 2011 el Congreso aprobó mediante LEY N° 29944 - Ley de Reforma Magisterial y de mejora de la calidad de la educación (MINEDU, 2012), en la cual se incorporaban pruebas de acceso a la carrera docente para el personal que venía laborando como tal. Con esta variable intentábamos medir el efecto de esa medida sobre los resultados escolares antes y después de esa medida. Profesores titulados, primaria - inicial (% del total).

Var14 Tasa de todas las formas de Tuberculosis (por 100 mil habitantes) - tasa de incidencia de todos los tipos de tuberculosis (TBC) entre niños hasta 36 meses años – La TBC representó hasta hace pocos años un problema crónico en la población de menores recursos. Al incluir esta variable quisimos medir el impacto sobre la educación de una variable de salud aun tan presente en los sectores más necesitados, sin embargo parece ser al ser un indicador bajo control, no incide en el rendimiento escolar. Los datos de malnutrición que podrían servir para reemplazar este indicador sólo llevan medidos a nivel regional desde el año 2007.

respecto a las TIC las que contribuyen a mejorar el rendimiento y no el simple hecho del acceso a ellas – caso de Chile (Salinas & Sánchez, 2009).

departamentos y una provincia constitucional (Callao), todas ellas llamadas regiones. Estas regiones serán nuestras unidades de análisis.

A efectos metodológicos censales, el INEI ha considerado oportuno dividir el análisis del departamento o región Lima en dos: Lima metropolitana (capital) y el resto de provincias limeñas, ya que estos espacios muestran comportamientos disímiles a pesar de la cercanía y pertenencia a la misma región.

Análisis estadístico de los años 2000 y 2013

En los estadísticos básicos comparativos que se muestran a continuación podemos observar principalmente dos cosas: que las medias de las variables que se espera mejoren con el tiempo lo hacen y las que se esperan se reduzcan (Var6 y Var15) también, y que los coeficientes de variación tienen un comportamiento más heterogéneo.

Análisis descriptivo

Tabla 4. Estadísticos descriptivos básicos de las Regiones, 2000 vs 2013

Descriptive Statistics 2000													
	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Coef.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	of variation	Statistic	Std. Error	Statistic	Std. Error
DPTE_aprobadosprimaria	26	20.61	73.04	93.65	83.9137	1.22241	6.23307	38.851	0.074	.045	.456	-1.188	.887
var2_educpib	26	8.50	1.54	10.04	4.2626	.45118	2.30059	5.293	0.540	.913	.456	.273	.887
var3_tasasistenciainicial	26	47.10	28.50	75.60	54.3058	2.53932	12.94805	167.652	0.238	-.144	.456	-.663	.887
var6_atrasoescolar	26	33.59	7.93	41.51	21.2731	1.82415	9.30139	86.516	0.437	.327	.456	-.790	.887
var9_accesointernetprimaria	26	7.05	.09	7.15	1.6063	.35536	1.81199	3.283	1.128	2.227	.456	5.271	.887
Var23_tasasnecinsatisfecha	26	65.25	21.52	86.77	48.0057	3.12879	15.95378	254.523	0.332	.264	.456	-.189	.887
Valid N (listwise)	26												

Descriptive Statistics 2013													
	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Coef.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	of variation	Statistic	Std. Error	Statistic	Std. Error
DPTE_aprobadosprimaria	26	12.58	83.32	95.90	91.8682	.59674	3.04281	9.259	0.033	-.938	.456	1.047	.887
var2_educpib	26	14.82	1.77	16.59	5.8644	.75542	3.85191	14.837	0.657	1.133	.456	.990	.887
var3_tasasistenciainicial	26	27.38	66.00	93.37	78.8824	1.39242	7.09997	50.410	0.090	.391	.456	-.281	.887
var6_atrasoescolar	26	18.61	3.01	21.62	9.5406	1.03004	5.25221	27.586	0.551	.621	.456	-.354	.887
var9_accesointernetprimaria	26	64.44	4.61	69.05	26.0827	3.92018	19.98910	399.564	0.766	1.003	.456	-.180	.887
Var23_tasasnecinsatisfecha	26	49.34	8.05	57.39	25.1590	2.66273	13.57731	184.343	0.540	.865	.456	.118	.887
Valid N (listwise)	26												

Aquellas variables en las que se reduce el coeficiente de variación (que mide el cambio de la diferencia entre los datos y las medias por período) son la VD, la Var3 de tasa de asistencia a nivel inicial y la Var9 de acceso a internet en las escuelas, perteneciendo las 2 últimas al ámbito de políticas educativas y económicas del gobierno central. En el caso de la Var6 de

atraso escolar del nivel primaria, a pesar de una reducción en la media, se amplía la brecha entre regiones.

Regresión Lineal Múltiple

La regresión múltiple tiene como objetivo central el diseñar un modelo que consiga “*explicar el comportamiento de una variable (endógena, explicada o dependiente)... utilizando la información proporcionada por un conjunto de variables explicativas (exógenas o independientes)...*” (Pérez, 2012). Por lo tanto, a fin de medir la significancia de las variables escogidas para el trabajo, se aplicarán regresiones lineales multivariantes a los sets de datos comparativos 2000 y 2013 a nivel regional para ver si se puede modelizar los resultados. El objetivo primordiales realizar un análisis de comparación de medias y de las varianzas, pero de existir posibilidad de modelizar, se procederá a hacerlo.

Para ello, deberemos comprobar si las variables cumplen los supuestos necesarios para poder validar la regresión y se parte por el ratio variables y observaciones, que según lo recomendado no debe bajar de 1 a 5, idealmente 20 a 5, con lo que al tener 5 VI’s para 26 observaciones cumplimos con el requisito mínimo exigido.

Idoneidad del modelo

La idoneidad y robustez del modelo a plantear se mide inicialmente a través del valor del R^2 .

Tabla. Idoneidad de Modelo, 2000

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.902 ^a	.814	.767	3.00815	2.529

a. Predictors: (Constant), Var23_tasasnecinsatisfecha, var3_tasasistenciainicial, var2_educpib, var9_accesointernetprimaria, var6_atrasoescolar

b. Dependent Variable: DPTE_aprobadosprimaria

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	790.300	5	158.060	17.467	.000 ^b
	Residual	180.980	20	9.049		
	Total	971.279	25			

a. Dependent Variable: DPTE_aprobadosprimaria

b. Predictors: (Constant), Var23_tasasnecinsatisfecha, var3_tasasistenciainicial, var2_educpib, var9_accesointernetprimaria, var6_atrasoescolar

Para el año 2000, con un R^2 de 0.814 y un p-valor significativo para todo nivel de confianza, aún se debe revisar los supuestos que nos indiquen que se puede modelizar el comportamiento

de la VD a partir de las VI's escogidas, ya que al ser el estadístico Durbin-Watson de 2.529, y al no estar claramente cerca de 2, nos indica que deberemos revisar los residuos estandarizados estimados.

Tabla 6. Idoneidad de Modelo, 2013

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.904 ^a	.817	.772	1.45447	1.239

a. Predictors: (Constant), Var23_tasasnecinsatisfecha, var2_educpib, var3_tasasisistenciainicial, var6_atrasoescolar, var9_accesointernetprimaria

b. Dependent Variable: DPTE_aprobadosprimaria

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	189.158	5	37.832	17.883	.000 ^b
	Residual	42.310	20	2.115		
	Total	231.468	25			

a. Dependent Variable: DPTE_aprobadosprimaria

b. Predictors: (Constant), Var23_tasasnecinsatisfecha, var2_educpib, var3_tasasisistenciainicial, var6_atrasoescolar, var9_accesointernetprimaria

Para el año 2013 el R^2 mejora mínimamente a 0.817, el p-valor es también significativo y el valor Durbin-Watson es de 1.239, lo que refuerza la idea de revisar los residuos.

Supuestos

SPSS nos ayudará a comprobar los siguientes supuestos básicos para poder determinar si se modeliza o no el comportamiento de la VD con respecto a las VI's:

- Normalidad
- Linealidad
- Homogeneidad o Igualdad de Varianzas (Homocedasticidad)
- Independencia de las variables

Además podremos comprobar otros supuestos no tan obvios como:

- No presencia de multicolinealidad
- Distribución normal de los errores

Test de Normalidad

Este contraste se realiza para comprobar si se verifica la hipótesis de normalidad necesaria para que el resultado de los análisis sea fiable, es decir, que los datos observados se comporten dentro como una distribución normal o campaniforme. En el cuadro siguiente observaremos las pruebas Kolmogorov-Smirnov, que se aplica para contrastar la hipótesis de normalidad de la población y cuyo estadístico de prueba es la máxima diferencia, así como el de Shapiro Wilk, que se ajusta mejor a muestras de tamaño 50 o menos. El SPSS nos arroja ambas pruebas por defecto.

Hipótesis Nula H_0 : Las distribuciones del conjunto de datos de las variables siguen un comportamiento normal

Hipótesis Alternativa H_1 : El comportamiento de las distribuciones del conjunto de datos de las variables no siguen la normalidad

Tabla 7. Tests de Normalidad, 2000 y 2013

Variables	Tests of Normality											
	2000						2013					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk			Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.
var2_educpib	.160	26	.085	.914	26	.033	.144	26	.178	.891	26	.010
var3_tasasistenciainicial	.096	26	.200 [*]	.961	26	.418	.098	26	.200 [*]	.968	26	.561
var6_atrasoescolar	.104	26	.200 [*]	.957	26	.327	.131	26	.200 [*]	.932	26	.089
var9_accesointernetprimaria	.245	26	.000	.713	26	.000	.203	26	.007	.855	26	.002
Var23_tasasnecinsatisfecha	.095	26	.200 [*]	.964	26	.474	.162	26	.078	.905	26	.021
DPTE_aprobadosprimaria	.109	26	.200 [*]	.947	26	.200	.112	26	.200 [*]	.931	26	.081

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

A un nivel de significación de 0.05, para el año 2000 las variables dependiente, Var2 y Var9 no presentan una distribución normal, mientras que para el 2013 no presenta normalidad además la Var23, por lo tanto se debería rechazar la hipótesis nula de normalidad. Ahondaremos un poco más en el supuesto de normalidad para comprobar la curtosis y sesgo.

Tabla 8. Test de Shapiro-Wilk de normalidad, 2000 y 2013

Variables		Descriptives					
		2000			2013		
		Statistic	Std. Error	Z-value	Statistic	Std. Error	Z-value
var2_educpib	Skewness	.913	.456	2.005	1.133	.456	2.486
	Kurtosis	.273	.887	0.308	.990	.887	1.117
var3_tasasistenciainicial	Skewness	-.144	.456	-0.315	.391	.456	0.857
	Kurtosis	-.663	.887	-0.748	-.281	.887	-0.317
var6_atrasoescolar	Skewness	.327	.456	0.717	.621	.456	1.363
	Kurtosis	-.790	.887	-0.891	-.354	.887	-0.400
var9_accesointernetprimaria	Skewness	2.227	.456	4.889	1.003	.456	2.201
	Kurtosis	5.271	.887	5.946	-.180	.887	-0.203
var23_tasasnecinsatisfecha	Skewness	.264	.456	0.580	.865	.456	1.899
	Kurtosis	-.189	.887	-0.213	.118	.887	0.133
DPTE_aprobadosprimaria	Skewness	.045	.456	0.099	-.938	.456	-2.059
	Kurtosis	-1.188	.887	-1.341	1.047	.887	1.181

Calculando el valor z que deberá estar comprendido entre +/-1.96 (valor del estadístico sobre el error estándar), podemos observar algunos resultados que no cumplen el requisito necesario. Procederemos a realizar una transformación de variables para conseguir la necesaria normalidad en ellas.

Tabla 9. Nuevo test Shapiro-Wilk, 2000 y 2013

Tests of Normality 2												
Variables	2000						2013					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk			Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.
LOGvar2	.099	26	.200 [*]	.962	26	.426	.103	26	.200 [*]	.958	26	.346
var3_tasasistenciainicial	.096	26	.200 [*]	.961	26	.418	.098	26	.200 [*]	.968	26	.561
var6_atrasoescolar	.104	26	.200 [*]	.957	26	.327	.131	26	.200 [*]	.932	26	.089
LOGvar9	.138	26	.200 [*]	.938	26	.118	.098	26	.200 [*]	.958	26	.352
LOGvar23	.148	26	.149	.954	26	.293	.155	26	.110	.953	26	.276
DPTE_aprobadosprimaria	.109	26	.200 [*]	.947	26	.200	.112	26	.200 [*]	.931	26	.081

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Luego de la transformación de las Var2, 9 y 23, conseguimos la normalidad necesaria para seguir evaluando el resto de supuestos, al lograr que no se rechace la hipótesis nula de normalidad al darse Sig. > 0.05 para todas las variables.

Test de Linealidad

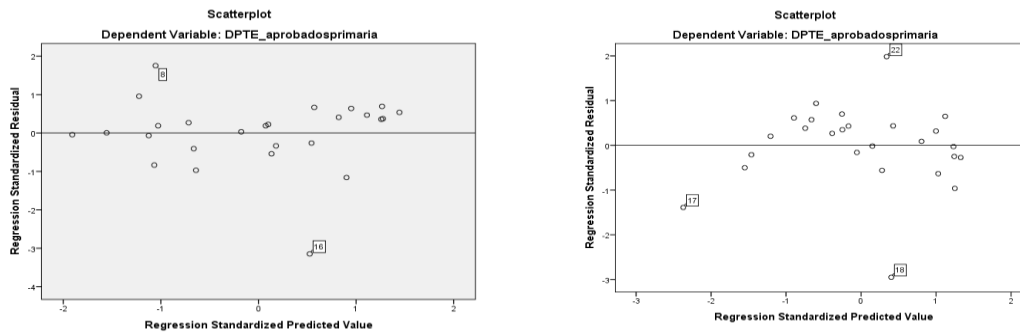
Lo que se buscará con este test es comprobar que el modelo será lineal o no.

Hipótesis Nula H_0 : El conjunto de datos de las variables muestran una tendencia lineal

Hipótesis Alternativa H_1 : El conjunto de datos de las variables NO muestran una tendencia lineal

Para este test, necesitamos graficar los residuos estandarizados de la VD versus su valor predictivo. La regla a superar es que los residuos no deben superar el valor de +/- 3 en ambos ejes, si trazáramos un línea paralela al eje X quedarán los valores repartidos 50% hacia arriba y 50% hacia debajo de la misma y que los valores residuales no muestren un patrón reconocible sino que se encuentren aleatoriamente dispersos.

Ilustración. Linealidad. Prueba visual



No parece estar claro en ninguno de los casos a linealidad, al no estar los residuos distribuidos de forma claramente aleatoria dentro de los rectángulos. Antes de rechazar la hipótesis nula, comprobaremos el resto de supuestos y en todo caso, procederemos a retirar las observaciones anómalas (etiquetadas).

Homogeneidad o Igualdad de Varianzas (Homocedasticidad)

La variación de los valores de los predictores deberá ser constante, independientemente de si los valores de los predictores son grandes o pequeños.

Hipótesis Nula H_0 : La variación de los valores de los predictores es constante.

Hipótesis Alternativa H_1 : La variación del término de error NO es constante.

Ya que usando los mismos gráficos de los supuestos de linealidad y de independencia de errores, no es suficiente para determinar la hetero/homocedasticidad, seguiremos comprobando supuestos.

Independencia de los errores

Para cualquier par de observaciones, los términos de error no deberán estar correlacionados.

Hipótesis Nula H_0 : No existe correlación entre los términos de error; hay independencia

Hipótesis Alternativa H_1 : Existe correlación entre los términos de error; no hay independencia

Usando los gráficos anteriores, la dispersión de los residuos parecen seguir en un patrón, los errores tampoco están dispersos de forma aleatoria al existir agrupación de los datos, lo que indicaría que NO se cumple este supuesto. Ya podríamos determinar la no linealidad del

modelo, por lo que no intentaremos probar más supuestos y comprobar el valor de los residuos.

Cabe mencionar que al intentar ajustar las variables con métodos curvilíneos y no lineales, tampoco se consigue explicar el comportamiento de la VD a través de un modelo que pueda servir para predecir, por lo que pasaremos a analizar la correlación entre variables y sus varianzas, tal como se contemplaba en un principio.

Análisis de la Correlación

Puesto que las variables analizadas no siguen comportamiento lineal, aunque sí normal, el método de correlación de *Spearman* nos permite medir mejor el grado de asociación entre dos o más variables tanto en dirección como en fuerza.

Tabla 10. Matriz de Correlación, 2000 y 2013

			Correlations					
			dpte	logvar2	var3	var6	logvar9	logvar23
Kendall's tau_b	dpte	Correlation Coefficient	1.000	-.464**	.361**	-.797**	.625**	-.477**
		Sig. (2-tailed)	.	.001	.010	.000	.000	.001
		N	26	26	26	26	26	26
logvar2	logvar2	Correlation Coefficient	-.464**	1.000	-.217	.556**	-.383**	.470**
		Sig. (2-tailed)	.001	.	.122	.000	.006	.001
		N	26	26	26	26	26	26
var3	var3	Correlation Coefficient	.361**	-.217	1.000	-.454**	.392**	-.380**
		Sig. (2-tailed)	.010	.122	.	.001	.005	.007
		N	26	26	26	26	26	26
var6	var6	Correlation Coefficient	-.797**	.556**	-.454**	1.000	-.631**	.594**
		Sig. (2-tailed)	.000	.000	.001	.	.000	.000
		N	26	26	26	26	26	26
logvar9	logvar9	Correlation Coefficient	.625**	-.383**	.392**	-.631**	1.000	-.557**
		Sig. (2-tailed)	.000	.006	.005	.000	.	.000
		N	26	26	26	26	26	26
logvar23	logvar23	Correlation Coefficient	-.477**	.470**	-.380**	.594**	-.557**	1.000
		Sig. (2-tailed)	.001	.001	.007	.000	.000	.
		N	26	26	26	26	26	26

** . Correlation is significant at the 0.01 level (2-tailed).

			Correlations					
			DPTE_aptosprimaria	Logvar2	var3_tasasistencial	var6_atrasoescolar	Logvar9	Logvar23
Kendall's tau_b	DPTE_aptosprimaria	Correlation Coefficient	1.000	-.507**	.495**	-.723**	.711**	-.526**
		Sig. (2-tailed)	.	.000	.000	.000	.000	.000
		N	26	26	26	26	26	26
Logvar2	Logvar2	Correlation Coefficient	-.507**	1.000	-.284*	.532**	-.587**	.519**
		Sig. (2-tailed)	.000	.	.042	.000	.000	.000
		N	26	26	26	26	26	26
var3_tasasistencial	var3_tasasistencial	Correlation Coefficient	.495**	-.284*	1.000	-.440**	.440**	-.440**
		Sig. (2-tailed)	.000	.042	.	.002	.002	.002
		N	26	26	26	26	26	26
var6_atrasoescolar	var6_atrasoescolar	Correlation Coefficient	-.723**	.532**	-.440**	1.000	-.717**	.545**
		Sig. (2-tailed)	.000	.000	.002	.	.000	.000
		N	26	26	26	26	26	26
Logvar9	Logvar9	Correlation Coefficient	.711**	-.587**	.440**	-.717**	1.000	-.557**
		Sig. (2-tailed)	.000	.000	.002	.000	.	.000
		N	26	26	26	26	26	26
Logvar23	Logvar23	Correlation Coefficient	-.526**	.519**	-.440**	.545**	-.557**	1.000
		Sig. (2-tailed)	.000	.000	.002	.000	.000	.
		N	26	26	26	26	26	26

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Lo primero a destacar en la correlación media entre las VI's y la VD es que todas las VI's están correlacionadas con la VD de manera significativa a un nivel de confianza del 99% y por lo tanto también al 95%.

Con respecto al grado de fuerza de la asociación entre las VI's y la VD, notamos que:

La variable LogVar2, gasto en educación con respecto al PIB regional no sólo tiene una relación negativa con el desempeño escolar sino que se hace un poco (9.48%) más negativa con el tiempo, al pasar de $-.464$ a $-.507$. Esto se puede deber a que la inversión pública en sí es un dato agregado que está compuesto por la suma gastos corrientes y de capital sin proporcionar un desglose más detallado y que no explica mucho en este caso excepto los montos invertidos en total. Al no estar sub-clasificada por partidas presupuestarias, hace que mayores montos invertidos no incidan en la tasa de aprobados de primaria, como parece ser el caso.

La Var3 tasa de asistencia en nivel inicial sí se incrementa 13 puntos, lo que indicaría que con el tiempo, a mayor cobertura de escuelas de nivel pre-escolar, más posibilidades de aprobar tienen los niños. Este comportamiento también parece seguir la lógica de las políticas educativas sugeridas por Naciones Unidas a través de Meta 2021, en la que se solicita ampliar la cobertura y acceso a estudios pre-escolares de los niños.

La Var6 de atraso escolar sigue estando muy correlacionada con la VD, por lo que se deberían concentrar esfuerzos en este sentido. Con respecto a esta observación, debemos recordar que ya GRADE había indicado la necesidad de enfocar esfuerzos en reducir esta tasa de manera más eficaz y evaluar las principales causas del problema, que podrían ser comunes a las causas de la tasa de aprobación y tasa de rendimiento escolar.

Con la Var9 de acceso de los colegios a internet, vemos que es la variable logarítmica que tiene una relación más fuerte con la VD. Se observa que el coeficiente mejora entre el 2000 y el 2013 en un 13.76%, al pasar de $.625$ a $.711$. Tal como los estudios referenciados indicaban, el acceso a TIC's en las escuelas parece tener un impacto mayor en las regiones en vías de desarrollo, cosa que se comprueba con este coeficiente.

Finalmente, la Var23, de tasa de necesidades insatisfechas (al menos 1 necesidad insatisfecha), sigue teniendo un impacto de moderado a alto en la tasa de alumnos aprobados de primaria, cosa que demuestra que existe una correlación entre la pobreza no monetaria (o multidimensional) y los resultados educativos.

A nivel de relación entre las VI's, la Var9_accesointernet y Var6 de atraso escolar están más correlacionadas con el tiempo superando el .70, con lo que se refuerza la hipótesis de la importancia de las TIC's en el desarrollo educativo de los alumnos y además están también ambas muy correlacionadas en ambos períodos con la Var23 de necesidades insatisfechas, que nuevamente confirma la importancia de la reducción de los niveles de pobreza multidimensional.

Análisis de Clústers

Una de las formas de clasificación de casos más utilizadas de acuerdo a las variables es el método de clústers. Este método o análisis es un “conjunto de técnicas multivariantes utilizadas para clasificar a un conjunto de individuos en grupos homogéneos” (Universidad de Valencia, s.f.).

En los métodos jerárquicos los individuos no se dividen en clústers una sola vez, sino que se van haciendo divisiones sucesivas o iteraciones. El método que se empleará será el método de Ward, que es uno de los más utilizados porque “posee casi todas las ventajas del método de la media y suele ser más discriminativo en la determinación de los niveles de agrupación” y está comprobado que “este método [es] capaz de acertar mejor con la clasificación óptima que otros métodos (mínimo, máximo, media y centroide)” (Op. Cít.). También se utilizará la distancia Euclídea normalizada, con lo que “solventamos el inconveniente de los efectos de unidades de medida distintas de las variables y obtenemos una distancia que no dependerá de las unidades de medida” (Op. Cít.).

Como se verá en la tabla siguiente, de entre los tres conglomerados extraídos en 2000 y 2013 se puede ver que el grupo 1, compuesto por las regiones con mayor incidencia de pobreza, se reduce en favor del grupo 2, con resultados medios. Los que presentan mejores resultados también se reducen. En general todas las regiones mejoran y por lo tanto se reorganizan las

sub-clasificaciones, manteniéndose en ambos períodos Amazonas, Loreto, Pasco, San Martín y Ucayali como las que menor progreso han tenido de todas las regiones y Lima / Callao, Ica, Arequipa, Tacna y Tumbes como las que mejor desempeño muestran.

Tabla 11. Pertenencia a los clúster por regiones, 2000 y 2013

Cluster Membership 2000				Cluster Membership 2013			
Case Number	region	Cluster	Distance	Case Number	region	Cluster	Distance
1	Amazonas	1	13.248	1	Amazonas	1	11.479
5	Ayacucho	1	17.618	17	Loreto	1	12.086
8	Cusco	1	3.623	20	Pasco	1	6.117
9	Huancavelica	1	23.219	23	San Martín	1	7.627
10	Huánuco	1	18.187	26	Ucayali	1	7.262
17	Loreto	1	5.384	2	Ancash	2	7.596
20	Pasco	1	11.054	3	Apurímac	2	14.181
23	San Martín	1	10.183	5	Ayacucho	2	10.177
26	Ucayali	1	8.041	6	Cajamarca	2	13.387
2	Ancash	2	8.103	8	Cusco	2	1.849
3	Apurímac	2	18.756	9	Huancavelica	2	14.646
6	Cajamarca	2	10.675	10	Huánuco	2	17.070
12	Junín	2	22.170	12	Junín	2	8.807
13	La Libertad	2	13.605	13	La Libertad	2	14.324
14	Lambayeque	2	9.601	14	Lambayeque	2	18.842
18	Madre de Dios	2	12.291	16	Lima Provincias 2/	2	19.747
21	Piura	2	12.722	18	Madre de Dios	2	10.330
22	Puno	2	4.533	19	Moquegua	2	23.092
4	Arequipa	3	18.883	21	Piura	2	9.442
7	Callao	3	6.094	22	Puno	2	10.365
11	Ica	3	4.847	4	Arequipa	3	11.739
15	Lima Metropolitana 1/	3	8.017	7	Callao	3	12.289
16	Lima Provincias 2/	3	5.988	11	Ica	3	4.185
19	Moquegua	3	5.574	15	Lima Metropolitana 1/	3	12.508
24	Tacna	3	13.137	24	Tacna	3	15.267
25	Tumbes	3	21.922	25	Tumbes	3	17.508

Análisis de la Varianza de los Clústers

Si bien es cierto no se puede contrastar los p-valores sobre la igualdad de las varianzas de los clústers ya que los niveles observados de significancia no están corregidos, sí sirve para hacernos una idea de la maximización de las diferencias entre los casos.

La hipótesis nula con la que trabajamos es que las varianzas, es decir, la dispersión de los datos que componen cada variable por cada región, no son muy diferentes entre ellas, y por lo tanto el comportamiento de cada clúster –dentro de sus diferencias- es similar.

Tabla 12. Varianzas de los Clústers, 2000 y 2013

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
var2	21.109	2	3.917	23	5.388	.012
var3	1097.511	2	86.795	23	12.645	.000
var6	765.069	2	27.511	23	27.809	.000
var9	17.894	2	2.013	23	8.890	.001
Var23	2302.374	2	76.449	23	30.116	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
var2	28.931	2	13.612	23	2.125	.142
var3	282.198	2	30.254	23	9.328	.001
var6	178.884	2	14.429	23	12.397	.000
var9	4104.651	2	77.383	23	53.044	.000
var23	1692.712	2	53.181	23	31.829	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Para el año 2000, se rechazaría la observación de que las varianzas de las variables independientes son iguales, mientras que se rechazaría en todos los casos en el año 2013 excepto para la var2, de gasto público sobre PIB por regiones. Esto nos indicaría que existen diferencias importantes entre los conglomerados por cada variable empleada, con lo cual deberemos realizar una nueva evaluación y comparación, esta vez identificando cada caso o unidad observada, a fin de establecer tendencias en el tiempo y contrastar las hipótesis de trabajo.

Análisis de Correlaciones a nivel de Clústers

Realizaremos un análisis de las correlaciones a nivel de Clústers que nos ayudará a ver el efecto de cada variable por cada conglomerado y así determinar qué tipo de prioridad debería dársele a cada una.

Tabla 13. Correlación de Clústers, 2000 y 2013

Correlation 2000					
Kendall's tau_b		Var Dependiente			Fuerza de Relación
		Cluster1	Cluster2	Cluster3	
Var Dependiente Aprobados Prim	Correlation Coefficient	1.000	1.000	1.000	-
	Sig. (2-tailed)				
var2 Gasto Educación / PIB	Correlation Coefficient	-.333	-.222	.038	Moderada
	Sig. (2-tailed)	.211	.404	.899	a baja
var3 Tasa asistencia nivel inicial	Correlation Coefficient	.278	-.222	.071	Moderada
	Sig. (2-tailed)	.297	.404	.805	a baja
var6 Atraso escolar	Correlation Coefficient	-.611*	-.833**	-.714*	Muy
	Sig. (2-tailed)	.022	.002	.013	Fuerte
var9 Acceso a internet de colegios	Correlation Coefficient	.167	.556*	.643*	Moderada
	Sig. (2-tailed)	.532	.037	.026	a fuerte
var23 Tasa necesidades insatisf.	Correlation Coefficient	.167	-.056	-.214	Débil
	Sig. (2-tailed)	.532	.835	.458	a moderada

*. Correlation is significant at the 0.05 level (2-tailed).
 **. Correlation is significant at the 0.01 level (2-tailed).

Correlation 2000					
Kendall's tau_b		Var Dependiente			Fuerza de Relación
		Cluster1	Cluster2	Cluster3	
Var Dependiente Aprobados Prim	Correlation Coefficient	1.000	1.000	1.000	-
	Sig. (2-tailed)				
var2 Gasto Educación / PIB	Correlation Coefficient	-.200	-.543**	.000	Moderada
	Sig. (2-tailed)	.624	.005	1.000	a nula
var3 Tasa asistencia nivel inicial	Correlation Coefficient	.200	.219	.200	Baja
	Sig. (2-tailed)	.624	.255	.573	
var6 Atraso escolar	Correlation Coefficient	-1.000**	-.600**	-0.46667	Muy fuerte
	Sig. (2-tailed)		.002	.188	a fuerte
var9 Acceso a internet de colegios	Correlation Coefficient	.800	.562**	-0.06667	Muy fuerte
	Sig. (2-tailed)	.050	.004	.851	a débil
var23 Tasa necesidades insatisf.	Correlation Coefficient	-.400	-.467*	.333	Moderada
	Sig. (2-tailed)	.327	.015	.348	

*. Correlation is significant at the 0.05 level (2-tailed).
 **. Correlation is significant at the 0.01 level (2-tailed).

Como se puede ver, para ambas tablas existe sólo correlación significativa en dos variables comunes: var6 de atraso escolar y var9 de acceso a internet en los colegios, aunque no para todos los clústers, y la var23 de tasa de necesidades insatisfechas y var2 de gasto de educación sobre el PIB regional para el clúster 2 del año 2013. Tal como lo habíamos mencionado, la reorganización de clústers en el 2013 se debía a mejoras en los indicadores de la mayoría de las regiones.

Al ser el gasto educativo sobre el PIB un dato agregado, se deberá analizar a profundidad qué elementos específicos inciden en esa mejora, o eliminar del todo la variable y reemplazarla con algún otro elemento que no está siendo censado hasta el momento o de lo que no se

cuentan datos comparables o que se están ya considerando o duplicando dentro de la variable 23, que es a su vez otro agregado.

Comprobación de Hipótesis

Con respecto a la **Hipótesis H_1** , la tasa de alumnos aprobados de primaria (VD) se incrementa en el tiempo, esta se acepta. Considerando que son 14 años entre períodos analizados, era de esperarse que este indicador mejore, especialmente cuando el punto de partida dejaba tanto rango para mejoras.

Con respecto a la **Hipótesis H_2** , existe correlación significativa entre la VD y cada una de las variables independientes incluidas en ambos períodos de análisis, ésta se rechaza. Existe correlación significativa en sólo 2 y 3 de las variables, dependiendo del período analizado y se rechaza también a nivel de clústers.

Tal como había extraído de los estudios referenciados, es muy difícil encontrar factores explicativos “modelizables” que sirvan para entender el rendimiento escolar, especialmente en períodos tan amplios de tiempo y con tantos avances sociales, económicos, tecnológicos y metodológicos experimentados, por lo que por ahora podemos realizar solamente seguimientos a las correlaciones y comportamiento de las varianzas.

Con respecto a la **Hipótesis H_3** , siguiendo los resultados de los estudios referenciados con respecto a la influencia de los maestros - la tecnología e infraestructura en los países menos desarrollados versus otras consideraciones en las más desarrollados-, se demuestra que a nivel regional se sigue la misma tendencia, por lo que podríamos aceptarla parcialmente ya que la var9 de acceso a internet de los colegios sí afecta la tasa de aprobados de nivel primario y además sí existe una diferencia entre las regiones menos desarrolladas (en las que no hay efecto por necesitar de elementos más básicos) y las más desarrolladas.

Considerando el período 2013, esta variable pierde su efecto para el caso del clúster 3 que aglomera las regiones más desarrolladas, con lo que podría intuirse que son otros los factores que impactan la variable dependiente. De acuerdo a los propios estudios base referenciados en este trabajo, estas variables están más relacionadas con las condiciones familiares y ambiente

de estudio de los alumnos, elementos que por el momento el censo no mide ni se contempla medir.

Conclusiones

El presente trabajo ha buscado verificar la posible incidencia de una serie de variables de salud, educación y condiciones de infraestructura escolar sobre el rendimiento de los alumnos expresado en el indicador *tasa de aprobados*. En la primera fase del trabajo se han descartado varias de ellas por existir una correlación muy baja entre cada una de ellas con respecto a la variable dependiente, que es precisamente esa tasa de aprobados.

De las variables con mayor fuerza relacional resultantes del análisis de la matriz de correlaciones inicial, nos quedamos con el gasto educativo en cada región sobre el PIB regional (Var2), la tasa de asistencia de educación inicial o pre-escolar (Var3), el atraso escolar que mide la repetición escolar (Var6), el acceso de las escuelas a internet (Var9) y la tasa de necesidades insatisfechas (al menos 1 necesidad insatisfecha), que se relaciona con la pobreza multidimensional (Var23).

Luego de aplicar los análisis estadísticos correspondientes, vemos que se refuerza la idea de la necesidad de enfocar el problema del rendimiento escolar desde una perspectiva multidimensional y el de no incrementar el gasto en educación sin saber exactamente qué factores inciden directa e indirectamente en el rendimiento escolar. Tal como se ha mencionado anteriormente, el gasto en educación por ser un agregado y su incremento presupuestario no es indicador suficiente para predecir mejores resultados en el tiempo.

Existe por lo tanto mucho espacio de mejora en el rendimiento académico de los alumnos de nivel primario a partir de muchas variables que también deben mejorar. Dentro de todas las que se han descartado por su bajo nivel de correlación con la VD, llama la atención la poca incidencia de las políticas magisteriales que permitieron ampliar la tasa de profesores titulados o certificados de nivel de primaria, que pasan de tener una correlación baja en el año 2000 a una correlación baja pero negativa en el 2013. Curiosamente, la variable de los profesores titulados de educación inicial pasa de una correlación muy baja (casi nula) en el 2000 – considerando la poca cobertura de la época- a una correlación un poco menos baja en 2013, aunque siempre positiva. Eso nos indicaría posibles problemas en la forma en que la Ley

29944 ha implementado la política de certificación de profesores de nivel de primaria. Al ser el nivel de preparación de los profesores un factor determinante en una etapa inicial de reformas educativas en países en vías de desarrollo, es indispensable que se evalúe a más profundidad estos temas a fin de tomar las medidas correctivas necesarias antes de seguir invirtiendo en programas de capacitación de profesores, que quizás no estén preparados para aprovechar e implementar esos nuevos conocimientos y herramientas en clases.

Finalmente, se sugiere por medio del presente trabajo una ampliación del cuestionario censal preparado por el MINEDU. Esa ampliación sería necesaria ya que existen regiones con resultados muy dispares entre ellas y por lo tanto exigen políticas diferenciadas y mediciones diferenciadas, así como factores de predicción del rendimiento que no se están considerando en este momento y que podrían ayudar a explicar los mejores resultados en aquellas regiones que conforman el clúster con desempeño superior en la tasa de aprobados de educación primaria.

Bibliografía

- Aesaert, K., Van Nijlen, D., Vanderlinde, R., Tondeur, J., Devlieger, I., & van Braak, J. (2015). The Contribution of Pupil, Classroom and School Level Characteristics to Primary School Pupils' ICT Competences: A Performance-based Approach. *Journal of Social Science Research, Computers and Education*(87), 55-69. doi:<http://dx.doi.org/10.1016/j.compedu.2015.03.014>
- Brumner, J., & Elacqua, G. (2004). Factores que Inciden en una Educación Efectiva. Evidencia Internacional. *Nota Técnica*. (O. C. Educativas, Ed.) Chile. Obtenido de <http://www.opech.cl/bibliografico/evaluacion/Brunner%20OEA.pdf>.
- CEPAL, & OEI. (2009). *Metas Educativas 2021: estudio de costos*. Santiago de Chile: Naciones Unidas. Recuperado el 5 de abril de 2015, de <http://www.oei.es/metas2021/documento.pdf>
- Educación, M. d. (2012). *Ley 29944 y reglamento*. Obtenido de Normatividad: <http://www.minedu.gob.pe/politicas/docencia/normatividad.php>
- GRADE. (Abril de 2015). Características Socioeconómicas y rendimiento de los estudiantes del Perú. *Análisis y Propuestas(Boletín N.28)*. Lima, Perú: GRADE.
- Guadalupe, C., León, J., & Cueto, S. (2013). *Paper Commissioned for the EFA Monitoring Report 2013/4, Teaching and learning: Achieving quality for all*. GRADE. UNESCO. Recuperado el 1 de mayo de 2015
- Honaker, J., & King, G. (2010). What to do about missing values in time series cross-section data. En H. U. Editorial (Ed.). *Vol. 54, No. 2* , págs. 561–581. *American Journal of Political Science*. Recuperado el 14 de junio de 2015, de <http://gking.harvard.edu/files/gking/files/pr.pdf>
- INEI - Instituto Nacional de Estadística e Informática. (2015). *Estadísticas Sociales. Indicadores*. Lima, Perú: INEI. Recuperado el 5 de Abril de 2015, de <http://www.inei.gob.pe/estadisticas/indice-tematico/sociales/>
- Kovacevic , M., & Calderon, M. (2014). *UNDP's Multidimensional Poverty Index: 2014 Specifications*. UNDP, Development Report Office, New York. Obtenido de http://hdr.undp.org/sites/default/files/specifications_for_computation_of_the_mpi.pdf
- Ministerio de Educación. (2015). *ESCALE - Estadísticas de la Calidad Educativa. Tendencias*. Lima, Perú. Recuperado el 5 de Mayo de 2015, de [Magnitudes: http://escale.minedu.gob.pe/magnitudes](http://escale.minedu.gob.pe/magnitudes)
- Morrell, D. (Febrero de 2011). *Who is Poor?* (H. Magazine, Editor) Recuperado el 21 de junio de 2015, de [Measurements: http://harvardmagazine.com/2011/01/who-is-poor](http://harvardmagazine.com/2011/01/who-is-poor)
- Ntoumanis, N. (14 de agosto de 2013). Multiple linear regression-Assumptions. (Youtube, Recopilador) Recuperado el 5 de abril de 2015, de <https://www.youtube.com/watch?v=DB-oKeNxjFs>

- OECD. (2011). *La medición del aprendizaje de los alumnos: mejores prácticas para evaluar el valor agregado de las escuelas*. Secretaría General de la OECD, Mexico D.F. Recuperado el 5 de Abril de 2015, de <http://www.oecd.org/edu/school/47871357.pdf>
- Pérez, C. (2009). *Modelos Econométricos con SPSS* (1 ed.). Madrid: Garceta.
- Pérez, C. (2012). *Econometría Básica. Aplicaciones con Eviews, Stata, Sas y SPSS* (1 ed.). Madrid, España: Garceta.
- Salinas, A., & Sánchez, J. (2009). Digital inclusion in Chile: Internet in rural schools. (I. J. Development, Ed.) *Journal of Social Science Research*(29), 573-582. doi:doi:10.1016/j.ijedudev.2009.04.003
- Sánchez-Cabezudo Sancho, J. (1986). *Predictores del rendimiento académico*. Universidad Autónoma de Madrid, Departamento de Psicología Biológica y de la Salud. Madrid: Facultad de Psicología. Recuperado el 8 de Julio de 2015
- Shong Chok, N. (2010). (U. o. Pittsburg, Ed.) Recuperado el 5 de junio de 2015, de Pearson's versus Spearman's and Kendall's correlations coefficients for continuous data: http://d-scholarship.pitt.edu/8056/1/Chokns_etd2010.pdf
- Tribunal Constitucional del Perú. (1993). *Constitución Política del Perú*. Recuperado el 4 de Diciembre de 2014
- UNDP. (2014). *Índice de Pobreza Multidimensional (IPM)*. Obtenido de Datos: <http://hdr.undp.org/es/content/indice-de-pobreza-multidimensional-ipm>
- Universidad de Valencia. (s.f.). *Análisis de Clústers*. Recuperado el 12 de Junio de 2015, de Proyecto CEACES: <http://www.uv.es/ceaces/multivari/cluster/CLUSTER2.htm>
- Vasquez, E. (2011). *El Perú de los pobres no visibles para el Estado*. Recuperado el 10 de noviembre de 2014, de Universidad del Pacífico - Documento de Discusión: <http://srvnetappseg.up.edu.pe/siswebciup/Files/DD1204%20-%20Vasquez.pdf>
- Vasquez, E. (2014). Las familias en las políticas y programas sociales del gobierno de Ollanta Humala desde la perspectiva de la pobreza multidimensional: 2012. En C. d. Pacífico (Ed.), *VII Convención de Familias* (pág. 21). Lima: Ministerio de la Mujer y Poblaciones Vulnerables.
- Wainer, J., Vieira, P., & Melguizo, T. (2014). The association between having access to computers and Internet and educational achievement for primary students in Brazil. *Journal of Social Science Research, Computers and Education*(80), 68-76. doi:<http://dx.doi.org/10.1016/j.compedu.2014.08.007>