

**DIRECCIÓN DE PROMOCIÓN DEL BIENESTAR Y RECONOCIMIENTO DOCENTE  
MINISTERIO DE EDUCACIÓN DEL PERÚ**

**INFORME FINAL  
ORDEN DE SERVICIO N° 0006230 - 2018  
026 PROGRAMA DE EDUCACION BÁSICA PARA TODOS**

Recopilación y sistematización de información para diagnóstico del perfil docente para el diseño efectivo de estrategias comunicacionales y de políticas de bienestar, reconocimiento y atracción ejecutados por la DIBRED

Daniel Ita Nagy  
Octubre de 2018

# **Informe de la recopilación y sistematización de información para diagnóstico del perfil docente para el diseño efectivo de estrategias comunicacionales y de políticas de bienestar, reconocimiento y atracción ejecutados por la DIBRED**

## **I. Introducción**

A continuación se presenta el informe que busca, a través de la recopilación y sistematización de información presente en la Encuesta Nacional de Docentes (ENDO), desarrollar una tipología que contribuya al entendimiento del perfil docente y pueda servir de insumo para desarrollar estrategias de comunicación, políticas de bienestar, reconocimiento y atracción hacia ellos, y en general, conocerlos mejor y hacer más efectivo el acercamiento a los mismos.

La primera parte del informe detalla el plan de investigación que se siguió para lograr la caracterización tipológica, en coordinación siempre con la Dirección de Promoción del Bienestar y Reconocimiento Docente (en adelante, DIBRED). A continuación se hace una breve revisión de la literatura nacional e internacional relacionada a experiencias de desarrollo de tipologías docentes, y las herramientas estadísticas utilizadas para generar tipologías, en particular herramientas de segmentación o clustering. Se justificará además la herramienta seleccionada para esta investigación. A continuación se desarrollará la metodología utilizada, que incluye tanto cada una de las etapas realizadas, las decisiones tomadas en cada una de las etapas, los análisis estadísticos multivariados empleados, las fuentes de datos consultadas, las variables elegidas para la tipología y un análisis de resultados que se presenta en formato Power Point, adjunto al presente documento. También se consigna adjunto a este informe la base de datos con la variable que indica la tipología a la cual se ha clasificado a cada docente presente en la ENDO, para posterior explotación de datos que requiera la DIBRED, así como las sintaxis de los software utilizados (SPSS y R) por si se desea replicar lo elaborado en el presente documento. Finalmente se consigna la bibliografía utilizada para esta investigación.

## II. Plan de investigación

Para el desarrollo de la presente investigación se planteó un plan de trabajo, coordinado y aceptado por la DIBRED, que consistió en 7 etapas que se detallan a continuación:

Etapa 1: Revisión de cuestionarios y bases de datos ENDO 2014 y 2016

En esta etapa se revisaron los cuestionarios correspondientes a la encuesta nacional de docentes existentes a la fecha, correspondientes a los años 2014 y 2016, así como las bases de datos, para familiarizarse con ellas, revisar las variables existentes, el formato en que fueron tabuladas y procesadas (SPSS) entre otros.

Etapa 2: Identificación de variables candidatas para la tipología, tanto de segmentación como de perfilamiento

En esta etapa se revisaron las variables existentes en la ENDO 2014 y 2016, se eligieron las más pertinentes para la investigación y se clasificaron ya sea como posibles variables para perfilar, es decir, describir las características de las tipologías identificadas, o como posibles variables de clustering, para generar las tipologías. Esta clasificación se presenta más adelante en el presente documento.

Etapa 3: Adaptación y/o transformación de variables

En esta etapa se adaptaron las variables elegidas para que puedan ser utilizadas para los fines de clustering (por ejemplo, creando nuevos códigos dentro de ellas o unificando y juntando respuestas para mejor uso) y se crearon también nuevas variables (por ejemplo, nivel socioeconómico). Las sintaxis utilizadas para la creación y adaptación de estas variables también se anexan al presente documento.

Etapa 4: Elaboración de base de datos unificada con variables de interés

En esta etapa se desarrolló una base de datos limpia que pueda ser utilizada por el software de análisis estadístico R, conteniendo las variables candidatas de clustering, que fueron el insumo para el desarrollo de las tipologías y de las cuales se hizo la selección de aquellas que mejor clasifican a los docentes.

Etapa 5: Generación de modelos, verificación de ajuste y pertinencia

En esta etapa se generaron las distintas alternativas de tipologías a través de modelos de clustering hasta llegar a dos soluciones estables y pertinentes, que fueron presentadas a la DIBRED para su discusión y elección. Las soluciones contenían las variables más importantes que caracterizaban cada tipología y una recomendación del consultor sobre cuál elegir.

## Etapa 6: Elección de modelo y desarrollo de perfilamiento

Con la tipología elegida, en esta etapa se perfilaron los segmentos y se desarrolló un informe en formato Power Point sobre las características de los mismos, adjunto al presente documento.

## Etapa 7: Elaboración de informe final

Finalmente, como última etapa, se desarrolló el presente informe metodológico conteniendo las distintas etapas del proceso, revisión bibliográfica, justificación metodológica, entre otros.

El cronograma desarrollado con cada una de las etapas descritas se presenta en el cuadro 1.

Cuadro 1: Cronograma del plan de investigación

Semanas	1	2	3	4	5	6
Revisión de cuestionarios y bases ENDO 2014 y 2016	X					
Identificación de variables para tipología como perfilamiento	X					
Adaptación y/o transformación de variables		X	X			
Elaboración de base de datos unificada			X			
Generación de modelos, verificación de ajuste y pertinencia				X		
Elección de modelo y desarrollo de perfilamiento				X	X	
Elaboración de informe final					X	X

### III. Breve revisión bibliográfica sobre investigación en tipología docente y análisis de clustering

#### Tipología docente

En general existe poca investigación específica relacionada a segmentar o clasificar docentes en función a sus características, ya sea actitudinales, sus valores, concepción de la educación, prácticas pedagógicas, atribuciones de logro, entre otros, es decir, relacionada al desarrollo de tipologías docentes bajo técnicas de clustering. Hace ya cerca de cuarenta años, Shavelson (1979) reflexionaba sobre lo poco usada pero importante que es la técnica de análisis de cluster para la investigación en educación, y proponía líneas de investigación usando esta técnica en áreas como educación superior, prácticas de enseñanza, estructura curricular e incluso estructuras cognitivas. Sin embargo, aun en años recientes este tipo de investigación sigue siendo limitada, y centrada en algunos pocos temas.

Entre investigaciones recientes desarrolladas a nivel internacional, Rahmat (2017) realizó en una muestra en Indonesia un análisis de cluster de estilos de liderazgo, identificando a través de clusters jerárquicos cuatro tipos de segmentos o tipologías, en función a liderazgos transformacionales o transaccionales. El análisis de cluster jerárquico tiene sus limitaciones, pero suele utilizarse cuando se aplican escalas, como fue el caso de este autor.

Otra iniciativa fue desarrollada por Ashish y otros (2015), quienes revisaron una serie de aplicaciones de algoritmos de data mining para investigación educativa, enfocados sobre todo en explicar el desempeño de estudiantes (potencial, performance académico, perfiles de aprendizaje, predicción de conducta) y, en el caso de docentes, en entender perfiles de estilos de docencia. Como se puede observar, esta investigación también se centra en estilos de docencia, antes que características sociodemográficas o actitudinales de los docentes. Otro ejemplo de este tipo de investigación es la desarrollada por Ling Koh y Sing Chai (2013), quienes estudiaron a través de análisis de cluster las percepciones hacia los contenidos pedagógicos y la tecnología de un grupo de docentes.

Otras investigaciones similares a estas utilizaron en cambio técnicas cualitativas, con el fin de identificar tipologías. En el caso de nuestra investigación, las metodologías cualitativas podrían utilizarse ya para etapas posteriores, para entender mejor otras características de los docentes clasificados en las tipologías estudiadas. Relacionado a técnicas cualitativas, Tack y Vanderlinde (2014) desarrollaron una tipología centrada en la disposición a la investigación, de involucrarse, engancharse y buscar oportunidades de investigación. Gerda y otros (2014) desarrollaron en Bélgica una tipología basada en las motivaciones de los docentes hacia la enseñanza, también con técnicas cualitativas.

Uno de los temas que está surgiendo bastante en la investigación relacionada a tipologías es las actitudes hacia el uso de la tecnología en la práctica docente, y en general, las actitudes hacia las TICs. Por ejemplo, Mama-Timotheou y Hennessy S. (2013) clasificaron a docentes en tipologías en función a los usos de la tecnología y su relación con las prácticas y creencias hacia las TICs. De forma similar, Malbernat, R. (2013) realizó una investigación clasificando a los docentes según su actitud hacia la innovación y la educación virtual e uso de TICs, usando algoritmos de K-medias.

En la revisión de la literatura efectuada, en Latinoamérica la investigación en tipología docente también ha estado más centrada en estilos de manejo de la clase y estilos de enseñanza que en clasificar a los docentes bajo características personales, como se realiza por ejemplo en la investigación del consumidor o el marketing.

Rojas Bravo (2013) investigó en Uruguay tipologías de docentes en función a contenidos de violencia en sus prácticas pedagógicas y su relación con el clima de aula, aunque en base a técnicas de observación (es decir, técnicas cualitativas). Domínguez y Martínez (2012) desarrollaron para una universidad en México una tipología de la docencia en base a prácticas docentes (planeación, conducción y evaluación) a través de una escala estructurada y aplicando el algoritmo de k-medias (cluster por centroide) identificando hasta 5 tipos de conglomerados y que además fueron perfilados por otras variables como edad, experiencia académica, grado

académico, estudios de postgrado. Esta experiencia es quizás la más cercana a la que se pretende desarrollar en esta investigación, sobre todo porque busca perfilar al docente.

En el Perú, Borgobello y otros (2010) investigaron la relación entre los estilos docentes y tipos de clase a través de técnicas de observación, y cómo se pueden observar distintos tipos de profesores en esta relación. Nuevamente, se desarrolló una tipología pero en base a técnicas cualitativas y centradas en estilo de docencia. Ayala (2016) investigó también en Perú la relación entre tipos de liderazgo y desempeño en La Oroya, centrando la generación de tipologías a través de esta relación pero basada en técnicas estadísticas correlacionales.

En general, los estudios asociados al docente en el Perú han estado más centrados en describirlo, pero no clasificarlo. Entre los temas más desarrollados están el conocer y describir la carrera docente, el perfil del docente, su satisfacción, la percepción que se tiene de su trabajo y la que tienen de él o ella sus alumnos, su relaciono con los salarios, los motivos de la elección de carrera, entre otros (Díaz y Ñopo, 2016). Guadalupe y otros (2017) utilizaron, al igual que esta investigación, datos de la ENDO 2014, para un muy buen resumen, pero centrado en describir al docente y conocerlo a profundidad, pero como un único perfil, antes que generar o proponer tipos de docentes.

Otros investigadores han estudiado a profundidad las actitudes de los docentes hacia su profesión y cuánto la valoran (Cuenca y Portocarrero , 2003; Cuenca y Carrillo (2017)) pero nuevamente con el objetivo de entenderlos antes que clasificarlos. O también a entender la adecuada forma de evaluar su desempeño (Cuenca y otros, 2011).

En general en el Perú no se han encontrado estudios que desarrollen tipologías de docentes a través de estudios de clustering, centrados en las características propias del docente.

### **Técnicas de clustering**

En general los métodos más conocidos son el método jerárquico provisto por el SPSS o el método de K-medias, que remiten a reglas de distancias, donde los sujetos más próximos en sus valores promedios son agrupados como parte de un segmento o cluster. Pero ambos métodos tienen falencias en cuanto no permiten tomar decisiones con rigurosidad estadística, y se basan mas bien en reglas heurísticas (reglas de conveniencia) y en modelos rígidos o de naturaleza inductiva. El método jerárquico por ejemplo no presenta reglas estadísticas de cuántos grupos explican bien los datos observados y basan sus decisiones en criterios gráficos o de conveniencia, y es además muy sensible a casos extremos y al método de elección de la estimación de las distancias (Orme y Johnson, 2008). El método de K-medias solo puede ser utilizado con variables cuantitativas, y la cantidad de grupos a elegir parte del propio investigador. Los grupos son estimados a partir de casos elegidos al azar, a los cuáles se les asigna los casos próximos. Luego, se recalculan los centros (conocidos como centroides) de estos nuevos grupos y se evalúa si hay casos que puedan cambiar de grupo en función a su proximidad a este nuevo centro. Este proceso se realiza iterativamente hasta que ningún caso cambie de grupo. Sin embargo, se ha demostrado que esta dependencia en la elección de puntos de inicio al azar puede llevar, por simple aleatoriedad no controlada, a resultados pobres y no representativos (Orme y Johnson, 2008) y que los supuestos

en que se basa el algoritmo de K-medias (que las varianzas error no varían entre clusters y son iguales entre los indicadores) tienen como resultado en un plano tridimensional grupos siempre de forma esférica, lo cual hace a veces imposible detectar otras configuraciones (Magidson y Vermunt, 2002; Magidson y Vermunt, 2002b). Una alternativa de uso reciente en el análisis de clusters, y que además puede ser utilizado para realizar análisis factorial de variables dicotómicas o regresión con variables nominales, son los modelos de clases latentes.

Los modelos de clases latentes fueron desarrollados en base a los trabajos de Lazarsfeld y Henry (1968), quienes buscaban una forma de estimar variables latentes a partir de encuestas con ítems dicotómicos. En el análisis factorial las variables latentes se asumen como continuas, al igual que las variables observadas, en cambio en los modelos de clases latentes las variables no observables se asumen como discretas o categóricas (Magidson y Vermunt, 2002; Magidson y Vermunt, 2003; Vermunt y Magidson, 2005).

A diferencia de los modelos lineales tradicionales (como la regresión, análisis log-lineales, etc.), que describen la relación entre variables observables, los modelos de clases latentes incluyen en su estimación una o más variables discretas no observables. Estas variables se interpretan en términos del análisis de segmentación como clusters o segmentos (Magidson y Vermunt, 2003; Vermunt y Magidson, 2005).

La segmentación por clases latentes, presenta amplias ventajas frente a los modelos de segmentación tradicional descritos (modelos jerárquicos y no jerárquicos, como el k-means). Una de las más importantes es que se trata de un método de segmentación basado en modelos, es decir, permiten contrastes de hipótesis de ajuste (ajuste de modelo a los datos, a diferencia de la aproximación de la segmentación K-medias que es un método inductivo y donde la cantidad de segmentos a estimar no puede ser contrastada estadísticamente), así como valorar si el aporte de cada variable es significativo, e incluso comparar el ajuste a los datos observados de un modelo basado en  $x$  clusters respecto a otro basado en  $x + n$  clusters. Estos modelos no descansan en los supuestos tradicionales de linealidad, distribución normal, etc., por lo que se encuentran menos afectados a posibles sesgos que provengan de datos que no cumplan estos supuestos (Vermunt y Magidson, 2005).

Otra diferencia importante con métodos tradicionales de segmentación es que la clasificación de los casos en los segmentos se basa en probabilidades estimadas directamente por el modelo, antes que por distancias que pueden ser relativas. Es decir, dos personas pertenecen al mismo segmento porque sus respuestas son generadas por la misma distribución de probabilidades. Cada caso es asignado al cluster al que sea más probable que pertenezca en base a sus respuestas.

Estos modelos permiten, además, incorporar variables con cualquier nivel de medida y considerando en su estimación su distribución apropiada. Esto permite incluir en los análisis variables que antes sólo podían ser incorporadas en los perfiles, es decir, luego de generados los clusters. Las variables categóricas pueden ser parte de los criterios mismos de segmentación desde su estimación. Esto se debe a que los modelos se basan en una estructura de probabilidades de modelos mixtos, que definen la relación entre las variables latentes y las observadas.

En general las técnicas de segmentación o clustering han sido más utilizadas en investigación del consumo, marketing y comunicación, para poder identificar grupos de personas con características comunes a las cuales se puedan dirigir de manera más eficiente campañas de comunicación (por ejemplo, enfocadas en los valores que comparten), productos o servicios (centrados en las necesidades de consumo que hayan manifestado, o grupos donde la competencia en la categoría es menor y por lo tanto el acceso es más sencillo, o también grupos de mayor capacidad económica y por lo tanto de alto interés comercial), o para entender sus motivaciones en general. Y aunque estas investigaciones suelen ser de carácter confidencial y son desarrolladas por empresas para su uso privado, algunas son de carácter público o al menos son conocidas o empleadas en el lenguaje común tanto comercial, como educacional o incluso político. Por ejemplo, la estimación de los niveles socioeconómicos (NSE), basada en criterios usados por ESOMAR (European Society for Opinion and Marketing Research), entre ellos el nivel educativo del jefe de hogar, junto a otras variables de corte como tenencia de bienes y características de la vivienda. La Asociación de Empresas de Investigación de Mercados (APEIM) que agrupa a las principales agencias de investigación de mercados peruanas es quien estima actualmente la fórmula, y utiliza como información la Encuesta Nacional de Hogares de INEI de naturaleza pública. Esta clasificación o tipología puede ser aplicada a cualquier ámbito, siempre y cuando se consignen las preguntas requeridas.

Otro ejemplo de tipología muy difundido en el ámbito local son los Estilos de Vida, desarrollados por Rolando Arellano (2003), que clasifica a los peruanos en seis segmentos en función a su “personalidad grupal” como lo define el autor. Esta clasificación agrupa actitudes, tendencias sociales, intereses y acceso a recursos, que se explica como:

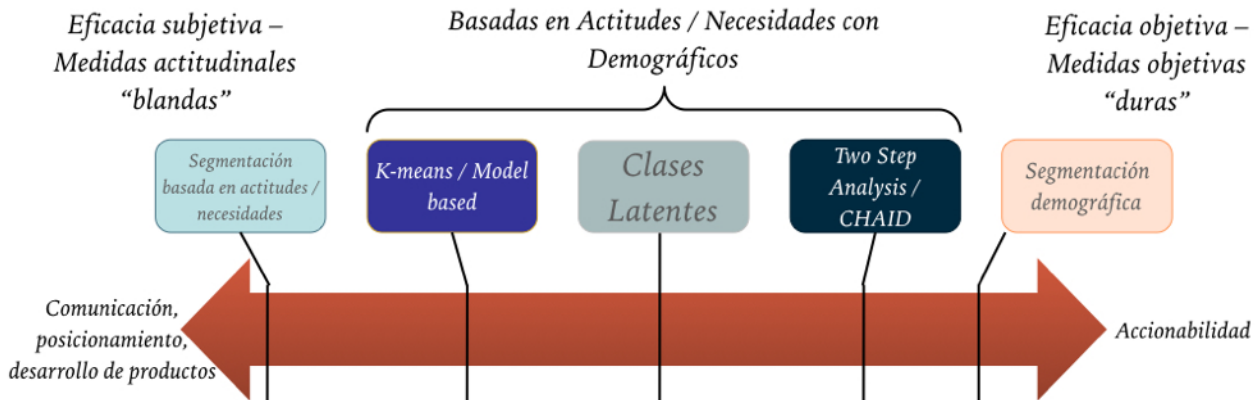
*“una manera de ser y de comportarse compartida por un grupo significativo de personas de una sociedad, que se parecen entre sí por sus características socio-demográficas, psicológicas, de comportamiento, de equipamiento e infraestructura, entre otras” (Arellano, 2003)*

Parte de los segmentos que identificó el autor ya son parte de nuestro hablar cotidiano (los progresistas, las modernas, los austeros, etc.) y es uno de los primeros en identificar el segmento que está definiendo la clase media actual en el país, los progresistas o emprendedores.

Relacionado a lo anterior, los métodos de estimación de clusters (o métodos de segmentación) se pueden clasificar según el objetivo que persiguen (ver gráfico 1).



Gráfico 1: Técnicas de estimación de clusters



Por ejemplo, la segmentación por NSE, de naturaleza más objetiva y accionable, utiliza métodos de estimación con variables “duras”, que permiten clasificar a las personas de manera excluyente (sólo se puede pertenecer a una categoría). Son de eficacia objetiva pues permiten ubicar a las personas de manera sencilla en uno de los grupos y derivar luego ciertas características de los mismos con facilidad. Pero al ser medidas más objetivas pierden profundidad en su descripción, y en su capacidad explicativa son de uso limitado.

La segmentación por estilos de vida, en cambio, hace uso de variables actitudinales, necesidades psicológicas, de consumo, etc., y se ubica más en el otro lado del espectro de tipos de segmentación, de eficacia subjetiva y basado en variables “blandas”. Este tipo de segmentación, de mayor profundidad, es muy útil para objetivos de comunicación, desarrollo de servicios y productos, entre otros, pero dado que se trata de variables no observables (ej. actitud hacia la vida) no permiten ubicar a las personas con tanta facilidad en categorías excluyentes. En este tipo de segmentación se habla más de probabilidades de pertenencia, de tendencias a estar en un grupo u otro y compartir sus características que a criterios absolutos. Se sacrifica, sin embargo, esa objetividad por una comprensión subjetiva más profunda.

Dado los objetivos planteados en la presente consultoría, se ha optado por una segmentación de eficacia subjetiva pero combinada con variables objetivas, por lo que el mejor método de estimación de las tipologías es del tipo clases latentes, o modelos mixtos finitos.

## **IV. Metodología**

### **1. Objetivos de la investigación:**

- Contribuir al diseño efectivo de estrategias comunicacionales y de políticas de bienestar, reconocimiento y atracción ejecutados por la DIBRED
- Identificar variables para la segmentación de la población docente que contribuyan a entender el bienestar docente, la percepción y/o necesidad de reconocimiento, la valoración de la carrera docente y su atractivo
- Lograr tipologías que sean claras de entender y visualizar, que permitan implementar estrategias comunicacionales y políticas públicas a partir de ellas.

### **2. Metodología:**

A continuación se presenta la secuencia metodológica utilizada para lograr la caracterización tipológica de los docentes en función a los resultados de la ENDO 2016.

#### **2.1. Revisión de instrumentos de cuestionarios y bases de datos, e identificación de variables candidatas para la tipología, tanto de segmentación como de perfilamiento**

Se revisaron los cuestionarios y bases de datos de la ENDO 2014 y 2016 y se elaboró una matriz identificando las variables pertinentes tanto para generar los segmentos o tipologías como para caracterizarlos, en función a los objetivos de la investigación. Para elegir las variables de segmentación y perfilamiento se tomaron en cuenta los siguientes criterios:

- Que no tengan filtros, es decir, que hayan sido respondidas por toda la población (sea del sector público o del sector privado)
- Que sean preguntas de opción cerrada. Pueden ser de tipo categórico nominal, ordinal, intervalo o continuas.
- Que contribuyan a los objetivos de la DIBRED: reconocimiento, valoración y atracción (por ejemplo, valoración de instituciones, dificultades en la carrera, satisfacción con la vida, con la profesión, etc.).
- Que permitan tener una imagen clara del docente perteneciente al segmento (género, edad, región, escala magisterial, NSE, entre otros).
- Que estén presentes en ENDO 2014 y ENDO 2016, y que se piensen aplicar en futuras mediciones

Respecto al último punto, luego de la revisión de las variables, se encontró que más del 50% de las variables presentes en la ENDO 2016 no estaban presentes en la ENDO 2014, en particular, las variables claves de segmentación o generación de tipologías, como se aprecia en el cuadro 2. Además, varias de las preguntas fueron modificadas, ya sea en su formulación o en sus opciones de respuesta, así como en el orden en que fueron presentadas, lo cual tiene impacto en los resultados obtenidos y su compatibilidad respecto al 2016. En general un 68% de las variables no eran comparables.

Utilizar sólo variables comunes a ambos estudios iba a limitar seriamente la descripción y riqueza de las tipologías, además de que la encuesta a realizarse en el presente año 2018 iba a ser muy similar a la del 2016, lo que volvía los resultados del 2014 aún más desfasados. En coordinación con la DIBRED se optó por prescindir de los resultados del año 2014 y centrar el análisis de tipologías en la base de datos de 2016, además de mostrar resultados más próximos a la situación actual del docente.

Como se observa también en el cuadro 2, las variables elegidas cubren los diversos temas establecidos como criterios de elección.

Las variables identificadas para generar los clusters se presentan en el cuadro 3. Estas variables fueron los insumos utilizados para realizar las tipologías, cuyo resultado se presentan más adelante. Las variables fueron clasificadas, luego de un análisis de contenido y por sus características de medición comunes, en dos grandes grupos:

- Variables orientadas a evaluar la vocación docente, a su satisfacción actual y percepción a futuro en su carrera docente y cómo perciben a las IIEE públicas tanto en la actualidad como a futuro. Estas variables son de tipo categórica o nominal (ejemplo, de respuestas sí o no, de elección de alternativas de opción cerrada) o de naturaleza ordinal pero simple (de máximo tres alternativas como mejor, igual, peor).
- Variables orientadas a evaluar la satisfacción con la vida, con el trabajo y la confianza que tienen hacia las instituciones. Estas variables son de tipo intervalo, tipo Likert, de 4 o 5 opciones de respuesta.

Los ejercicios de clustering para llegar a las tipologías desarrolladas en esta investigación consideraron estos dos grupos de variables como insumos diferenciales, además de otras variables comunes de tipo demográfico como sexo, edad, NSE, nivel en que enseña, área de la IIEE, gestión de la IIEE, entre otros (ver cuadro 3).

Cabe señalar que algunas variables fueron descartadas como generadores de clusters durante el proceso de generación de tipologías, como por ejemplo las opiniones sobre los cambios introducidos por la LRM (que solo fueron aplicadas al sector público o los que conocían la LRM), la tenencia de hijos menores de 12 años (no generaba diferencias entre segmentos), forma de capacitación preferida (virtual, presencial o semipresencial, porque tampoco contribuía a diferenciar segmentos) entre otros.

Cuadro 2: Clasificación variables ENDO 2014 y 2016

Tema	Descriptivo	Uso	Pregunta 2016	Pregunta 2014	Comparabilidad con 2014
Demográfico	Sexo	Cluster	109_c1	68	SI
Demográfico	Edad	Cluster	109_d1	69	SI
IE	Área	Cluster	4	area	SI
IE	Gestión	Cluster	7	gestion	SI
IE	Nivel que enseña	Cluster	9	nivel	SI
NSE	NSE	Cluster	NSE	No tiene	NO
Atractivo carrera docente	Razón para ser docente	Cluster	301	1	SI
Trabajo actual	Condición actual	Cluster	306	21	NO
Valoración carrera docente	Satisfacción con trabajo actual	Cluster	321	No tiene	NO
Valoración carrera docente	Volvería a ser docente	Cluster	323	4	SI
Atractivo carrera docente	Futuro laboral	Cluster	325	63	SI
Valoración carrera docente	Hijos docentes	Cluster	801	3	SI
Reconocimiento	Confianza en instituciones	Cluster	805	55	NO
Valoración carrera docente	Opinión LRM	Cluster	811	64	SI
Valoración carrera docente	Satisfacción con LRM	Cluster	812	No tiene	NO
Reconocimiento	Satisfacción con la vida	Cluster	818	No tiene	NO
Reconocimiento	Satisfacción con labor docente	Cluster	819	30	NO
Valoración carrera docente	Logro aprendizaje IIEE públicas	Cluster	821	54	SI
Valoración carrera docente	Mejoría de IIEE públicas	Cluster	822	52	SI
Valoración carrera docente	Futuro de IIEE públicas	Cluster	823	53	SI
Atractivo carrera docente	Condiciones trabajar en áreal rural	Perfil	504	No tiene	NO
Atractivo carrera docente	Iniciativas para docentes	Perfil	814_1 a 814_9	No tiene	NO
Atractivo carrera docente	Dejaría pública a privada	Perfil	817	No tiene	NO
Capacitación	Acciones formativas	Perfil	601	No tiene	NO
Capacitación	Programas formativos	Perfil	602	No tiene	NO
Capacitación	Uso de TICs	Perfil	604	No tiene	NO
Capacitación	Visita acompañante pedagógico	Perfil	605	No tiene	NO
Capacitación	Preferencia capacitación	Perfil	607	No tiene	NO
Capacitación	Importancia temas	Perfil	608	13	NO
Capacitación	Intencion de desarrollo academico	Perfil	609	16	NO
Capacitación	Mejoras práctica docente	Perfil	902	40	NO
Capacitación	Modalidad preferida	Perfil	No tiene	14	NO
Capacitación	Características preferidas capacitación	Perfil	No tiene	15	NO
Comunicación	Medios de comunicación utilizados	Perfil	701_1 a 701_8	No tiene	NO
Comunicación	Medios informa actividad docente	Perfil	702	62	NO
Comunicación	Tiene celular	Perfil	703	No tiene	NO
Comunicación	Tiene smartphone	Perfil	704	No tiene	NO
Comunicación	Medios de contacto con Minedu	Perfil	705	No tiene	NO
Comunicación	Dónde se conecta a internet	Perfil	706	88	SI
Comunicación	Actitud hacia TIC	Perfil	707	49	NO
Comunicación	Tipos de personas como vecino	Perfil	820	94	SI
Comunicación	Inclusión temas diversos	Perfil	No tiene	51	NO
Demográfico	Estado civil	Perfil	No tiene	71	NO
Demográfico	Departamento nacimiento	Perfil	101_2	70_2	SI
Demográfico	Raza de origen	Perfil	102	No tiene	NO
Demográfico	Tenencia de bienes	Perfil	108	86	SI
Demográfico	Jefe de hogar	Perfil	109_e1	75	SI
Demográfico	Cantidad de personas en el hogar	Perfil	109	74	SI
Demográfico	Medios usados para trasladarse a la IE	Perfil	113	28	NO
Demográfico	Debe dormir fuera para ir a su IE	Perfil	114	26	NO
Demográfico	Hijo menor de 12 años en hogar	Perfil	109	No tiene	NO
Evaluación desempeño	Forma evaluación	Perfil	806	45	NO
Evaluación desempeño	Aspectos a ser evaluados	Perfil	807	44	NO
Evaluación desempeño	Rol director	Perfil	No tiene	47	NO
Formación	Primera carrera estudiada	Perfil	202	5	SI
Formación	Estudios complementarios pedagogía	Perfil	203	No tiene	NO
Formación	Si ha estudiado algo relacionado a Educación	Perfil	208	No tiene	NO
Formación	Tipo institución donde estudió para ser docente	Perfil	208_1	No tiene	NO
Formación	Gestión donde estudió para ser docente	Perfil	208_3	No tiene	NO
Formación	Calidad formación percibida	Perfil	210	9	NO
Formación	Maestría	Perfil	212_1	No tiene	NO
Formación	Doctorado	Perfil	212_2	No tiene	NO
Formación	Lenguas que conoce	Perfil	213_1 a 6	77	SI
IE	Departamento	Perfil	1	region	SI
IE	EIB	Perfil	11	No tiene	NO
IE	Combinacion area gestion nivel	Perfil	estrato	estrato	SI
IE	Instalaciones	Perfil	317	32	SI
IE	Servicios de la IE	Perfil	318_1 a P318_5	No tiene	NO
IE	Uso de materiales para clases	Perfil	No tiene	33	NO
Reconocimiento	Remuneración justa	Perfil	502	No tiene	NO
Reconocimiento	Opinión sociedad trabajo docente	Perfil	803	57	SI
Salud / Bienestar	Enfermedades	Perfil	401	No tiene	NO
Salud / Bienestar	Discapacidad	Perfil	402	72	SI
Salud / Bienestar	Financiamiento salud	Perfil	405	83	NO
Salud / Bienestar	Problemas salud en IE	Perfil	407	34	NO
Salud / Bienestar	Amenazas comunidad educativa	Perfil	No tiene	37_1 a 6	NO
Situación económica	Ingreso total bruto	Perfil	501_A	98	NO
Situación económica	Tiene ocupación secundaria	Perfil	507	23	SI
Situación económica	Ingreso bruto ocupación secundaria	Perfil	508	No tiene	NO
Situación económica	Situación respecto a sus padres	Perfil	511	No tiene	NO
Situación económica	Deudas	Perfil	512	No tiene	NO
Trabajo actual	Escala magisterial	Perfil	308	No tiene	NO
Trabajo actual	Deseo de nombrarse	Perfil	309	No tiene	NO
Trabajo actual	Participación en nombramiento	Perfil	310	No tiene	NO
Trabajo actual	Cantidad alumnos por aula	Perfil	315	No tiene	NO
Trabajo actual	Carga de horas	Perfil	316	31	NO
Trayectoria	Año en que empezó a trabajar	Perfil	302	17	SI
Trayectoria	Años trabajados	Perfil	304_1 a 3	19_1 a 3	SI
Trayectoria	Participación en organizaciones	Perfil	No tiene	93	NO
Valoración carrera docente	Razon no ser docente	Perfil	324	No tiene	NO
Valoración carrera docente	Satisfacción con ingresos	Perfil	509	No tiene	NO
Valoración carrera docente	Factores de éxito académico	Perfil	804	No tiene	NO
Valoración carrera docente	Conocimiento iniciativas educativas	Perfil	813_1 a 813_6	No tiene	NO
Valoración carrera docente	Dificultad quehacer docente	Perfil	904	43	NO
Valoración carrera docente	Iniciativa de innovación	Perfil	905	No tiene	NO
Valoración carrera docente	Razón docente IIEE privada	Perfil	1001	24	SI
Valoración carrera docente	Intención trabajar IIEE pública	Perfil	1002	25	SI
Valoración carrera docente	Reconocimientos IIEE privada	Perfil	1006	102	SI
Valoración carrera docente	Hijos en IE pública	Perfil	No tiene	81	NO
Valores	Calificación acciones controversiales	Perfil	No tiene	95	NO

Cuadro 3: Variables consideradas para la generación de tipologías

Descriptivo	Pregunta ENDO 2016	Orientación
Sexo	109_c1	Demográficas
Edad	109_d1	Demográficas
Área	4	Demográficas
Gestión	7	Demográficas
Nivel que enseña	9	Demográficas
NSE	NSE	Demográficas
Condición actual	306	Demográficas
Razón para ser docente	301	Vocación y futuro docente
Satisfacción con trabajo actual	321	Vocación y futuro docente
Volvería a ser docente	323	Vocación y futuro docente
Futuro laboral	325	Vocación y futuro docente
Hijos docentes	801	Vocación y futuro docente
Opinión LRM	811	Vocación y futuro docente
Logro aprendizaje IIEE públicas	821	Vocación y futuro docente
Mejoría de IIEE públicas	822	Vocación y futuro docente
Futuro de IIEE públicas	823	Vocación y futuro docente
Confianza en MINEDU	805_1	Satisfacción y confianza
Confianza en UGEL	805_4	Satisfacción y confianza
Confianza en SUTEP	805_7	Satisfacción y confianza
Satisfacción con la vida	818_1	Satisfacción y confianza
Satisfacción con la salud	818_2	Satisfacción y confianza
Satisfacción con el trabajo	818_3	Satisfacción y confianza
Satisfacción con condiciones de jubilacion	818_5	Satisfacción y confianza
Satisfacción con relaciones familiares	818_8	Satisfacción y confianza
Satisfacción con logros de estudiantes	819_2	Satisfacción y confianza
Satisfacción con reconocimiento del Minedu	819_4	Satisfacción y confianza
Satisfacción con relación con padres de familia	819_7	Satisfacción y confianza
Satisfacción con el salario	819_9	Satisfacción y confianza

## 2.2. Adaptación, transformación y creación de variables, y base de datos

Luego de la selección de variables se tuvieron que crear y adaptar algunas variables, por ejemplo transformar la variable edad a rangos, o crear la variable de NSE, a través del software estadístico SPSS. La sintaxis para la adaptación y creación de estas variables se adjunta en el Anexo 1.

Con la base de datos adicional a la encuesta ENDO 2016 que consigna la composición familiar del encuestado se crearon nuevas variables, que identificaban la carga familiar del docente: si tiene cónyuge, hijos, hijos menores de 12 años, y familiares a cargo. La base traspuesta en formato .sav se adjunta en archivo digital al presente documento y la sintaxis para el cálculo de las variables usando esa base se adjunta en el Anexo 2.

A continuación se guardaron las variables seleccionadas en una base limpia en formato de archivo separado por comas (.csv) para poder ser utilizadas en el software estadístico R para las siguientes etapas de análisis. Esta base contenía las variables candidatas de clustering, que

fueron el insumo para el desarrollo de las tipologías y de las cuales se hizo la selección de aquellas que mejor clasificarían a los docentes.

En un siguiente momento, y utilizando el software R, para el caso de las variables candidatas a clasificar a los docentes relacionadas a la satisfacción con la vida y el trabajo y la confianza en instituciones se realizaron análisis factoriales exploratorios para identificar las variables que representaban mejor los constructos subyacentes que se buscaba medir. Se realizaron análisis factoriales exploratorios para los ítems que medían:

- Confianza en las instituciones (ítems medidos por las preguntas 805\_1 a 805\_7 del cuestionario ENDO 2016)
- Satisfacción con la vida (preguntas 818\_1 a 818\_9)
- Satisfacción con el trabajo (preguntas 819\_1 a 819\_10)

Se utilizaron en los tres casos análisis factorial exploratorio con el método de estimación de mínimos cuadrados ordinarios (OLS), que ha demostrado ser el que presenta mejor ajuste en diversos estudios de simulación (Lloret-Segura y otros, 2014). Para estimar la cantidad de factores a extraer se utilizaron los criterios recomendados actuales de Very Simple Structure (VSS), MAP y análisis paralelo. La rotación para estimar las saturaciones factoriales fue oblicua, dado que se trata de variables actitudinales o psicológicas y se espera que se obtengan constructos relacionados entre sí (Lloret-Segura y otros, 2014). La sintaxis en el software R utilizada para realizar los análisis factoriales exploratorios se consigna en el Anexo 3.

Para el caso de confianza en las instituciones se identificaron que los ítems de confianza en el Minedu, en la UGEL y en el SUTEP, eran buenos representantes de los factores subyacentes obtenidos, pues mostraban cargas o saturaciones factoriales altas en sus respectivos factores (ver cuadro 4). Además son instituciones reconocidas por todos los docentes.

Cuadro 4: Cargas factoriales ítems de Confianza en las Instituciones

Ítem		Carga factorial		
		Factor1	Factor2	Factor3
p805_1	Minedu	0.69	0.1	-0.07
p805_2	Consejo Nacional de Educación	0.79	-0.08	0.07
p805_3	DRE	0.43	0.34	0.09
p805_4	UGEL	0	0.97	0.01
p805_5	Medios de comunicación	0.1	0.06	0.32
p805_6	Derrama Magisterial	0.05	-0.03	0.75
p805_7	SUTEP	-0.08	0.07	0.61

Para el caso de satisfacción con la vida se seleccionaron los ítems de satisfacción con la vida, condiciones de jubilación y relación con familiares por los mismos criterios descritos líneas arriba (ver cuadro 5). Por solicitud de la DIBRED se añadieron además los ítems de satisfacción con la salud y el trabajo al análisis.

Cuadro 5: Cargas factoriales ítems de Satisfacción con la vida

Ítem		Carga factorial		
		Factor1	Factor2	Factor3
p818_1	Vida	0.92	-0.02	0
p818_2	Salud	0.37	0.25	0.08
p818_3	Trabajo	0.27	0.3	0.15
p818_4	Educación de los hijos	0.11	0.38	0.21
p818_5	Condición de jubilación futura	-0.02	0.64	-0.1
p818_6	Autoestima	0.22	0.1	0.42
p818_7	Reconocimiento de la sociedad	-0.01	0.4	0.22
p818_8	Relaciones familiares	0	-0.04	0.74
p818_9	Actividades recreativas	-0.02	0.34	0.34

Para la satisfacción con el trabajo se seleccionaron los ítems de satisfacción con los logros de los estudiantes, la relación con los padres de familia y el salario (ver cuadro 6). No se consideró el ítem de relación con el Director porque no era un ítem que respondían todos los docentes (los directores de escuela no participaban de esta pregunta). Por solicitud de la DIBRED se añadieron además los ítems de satisfacción con el reconocimiento del Minedu a sus docentes.

Cuadro 6: Cargas factoriales ítems de Satisfacción con el trabajo

Ítem		Carga factorial		
		Factor1	Factor2	Factor3
p819_1	Actividad pedagógica	0.08	0.54	0.03
p819_2	Logros de los estudiantes	-0.07	0.72	0.02
p819_3	Relación con los estudiantes	0.18	0.53	-0.06
p819_4	Reconocimiento del Minedu	0.08	0.12	0.32
p819_5	Relación con colegas	0.46	0.1	0.04
p819_7	Relación con padres de familia	0.74	0	-0.02
p819_8	Ubicación de la IIEE	0.35	0.05	0.14
p819_9	Salario	0	0	0.94
p819_10	Relación con la comunidad	0.67	-0.01	0.02

También se realizó un análisis factorial de los ítems que evaluaban los cambios introducidos por la Ley de la Reforma Magisterial (LRM), que se presenta en el cuadro 7, pero que fue descartado como insumo para la tipología porque sólo fueron respondidos por docentes que conocían la ley y pertenecían al sector público.

Cuadro 7: Cargas factoriales ítems que evalúan cambios introducidos por la LRM

Ítem		Carga factorial	
		Factor1	Factor2
p812_1_r	Nuevas escalas magisteriales	-0.02	0.85
p812_2_r	Meritocracia	0.24	0.43
p812_3_r	Un solo régimen laboral	0.31	-0.12
p812_4_r	Evaluaciones periódicas	0.38	0.16
p812_5_r	Cese a los 65 años	0.2	0.09
p812_6_r	Acceso a cargos	0.6	0.03
p812_7_r	Asignaciones	0.56	-0.07

Adicionalmente, utilizando los códigos modulares de las IIEE a las cuales pertenecen los docentes entrevistados, se complementó la información consignada en la ENDO 2016 con información del padrón de IIEE del Minedu ([http://escale.minedu.gob.pe/uee/-/document\\_library\\_display/GMv7/view/958881](http://escale.minedu.gob.pe/uee/-/document_library_display/GMv7/view/958881)) actualizado al 20 de setiembre de 2018, y con información del centro poblado donde se ubica la IIEE así como su vulnerabilidad, brindado por el Censo Educativo del 2017. En el primer caso se vinculó la información a través del código modular, y en el segundo caso a través del código de local del código modular. Las variables adicionales se consignan también en la base de datos en formato .sav adjunto a este informe, que indica la tipología a la cual se ha clasificado a cada docente presente en la ENDO 2016.

### 2.3. Elección de método estadístico de segmentación

Se consideraron dos métodos estadísticos para generar los clusters: métodos basados en clases latentes y métodos basados en distancias. En el primer caso se consideró el método de clases latentes politómicas (poLCA) que puede utilizar para la estimación variables categóricas, ordinales y continuas. Este método asume que hay categorías o “clases” subyacentes o comunes a las personas que son las que explican sus diferencias (Linzer y Lewis, 2011). Esas clases son las corresponderían a los segmentos a los que pertenecen los docentes. Este método está implementado en el software estadístico de uso libre R.

Un segundo método considerado fue el de cluster ensemble, que realiza una meta segmentación, de estimaciones basadas en distancias. Busca encontrar patrones de clasificación como resultado de múltiples segmentaciones utilizando modelos jerárquicos y de k-medias (Orme y Johnson, 2008). Este método requiere variables cuantitativas, al menos de intervalo, y utiliza un software especializado para su implementación (Cluster Ensemble Analysis Software, CCEA).

Para la elección de la metodología se tuvo en consideración:

- Los tipos de variables elegidos para generar los clusters. Se observó (ver cuadro 3) que las variables eran o bien de tipo categórico o bien de tipo ordinal. Algunas son de intervalo, pero un buen grupo de ellas, al menos las centradas en vocación y futuro docente, no cumplen ese nivel mínimo.



- Facilidad para clasificación posterior de nuevos docentes. El método poLCA estima probabilidades de pertenencia a la tipología o cluster en función a cada variable considerada. Es posible generar una instrucción sencilla con el software R para clasificar nuevos docentes, siempre y cuando respondan las preguntas requeridas, o en su defecto, estimar una función de clasificación reducida a través de métodos como machine learning. El método de cluster ensemble en cambio requiere el uso del software licenciado y solo permitiría clasificar por las escalas de satisfacción y confianza en instituciones.
- Coherencia y claridad de los clusters o segmentos obtenidos. Como el método poLCA permite incluir variables demográficas de tipo categórico en la segmentación o generación de clusters, los perfiles obtenidos son mucho más claros.

Por los motivos expuestos finalmente se optó por estimar las tipologías con el método de clases latentes politómicas (poLCA) a través del software libre R. La sintaxis en el software R utilizada para la estimación de los segmentos se consigna en el Anexo 4. En este anexo se consigna todo el proceso desde la lectura de la base de datos, construcción y adaptación de variables, generación de tipologías y exportación de la pertenencia a clusters por caso para anexarlos luego a la base ENDO 2016 en formato .sav del SPSS.

#### **2.4. Generación de modelos, ajuste y pertinencia**

Luego de definir las variables candidatas a generar las tipologías y el método de estimación adecuado, se procedió a estimar los segmentos o clusters con el software libre R. En primer lugar se transformaron diversas variables a las características del software, como por ejemplo etiquetar los valores perdidos de las variables, consignar el orden adecuado de las escalas en el caso de las variables de intervalo, entre otros.

En un segundo lugar, dado que se requiere que todas las preguntas utilizadas estén respondidas, se imputaron los valores perdidos de las variables de interés para el análisis de cluster.

Como se señaló anteriormente se desarrollaron dos estrategias de clustering en función a las variables consideradas para generar las tipologías (ver sección 2.1): una estrategia centrada en variables orientadas a evaluar la vocación docente, a su satisfacción actual y percepción a futuro en su carrera docente y cómo perciben a las IIEE públicas; y una segunda centrada en variables orientadas a evaluar la satisfacción con la vida, con el trabajo y la confianza que tienen hacia las instituciones. Ambas estrategias consideraban además variables demográficas (ver cuadro 3).

Para el primer caso, que llamaremos tipología centrada en vocación y futuro, se realizó una imputación sólo de las variables requeridas bajo esa estrategia, utilizando el método de imputación multivariada de ecuaciones encadenadas (van Buuren y Groothuis-Oudshoorn, 2011) conocido como *mice* por sus siglas en inglés. Este método, que es parte de los métodos de imputación por regresión estocástica, agrega un puntaje aleatorio a las variables imputadas utilizando el estimado del error cuadrático medio o de las varianzas de los residuos producto de la regresión, para preservar la varianza y simetría de los datos originales. Es el método más recomendado en la actualidad.

Luego de obtener los datos imputados se estimaron distintas alternativas de tipologías, que se consignan en la sintaxis adjunto en el Anexo 4. Se descartaron algunas variables, como ya se había indicado, porque no contribuían a identificar bien las tipologías, y se fueron generando alternativas de más a menos parsimoniosas, siempre y cuando los tamaños de muestra obtenidos en la cantidad de segmentos tuviera un tamaño con base suficiente y que tuvieran sentido y diferenciación respecto a las variables utilizadas.

Durante el proceso se hizo evidente que era necesario generar las tipologías separando la base de datos en docentes pertenecientes a IIEE públicas de IIEE privadas. Esto debido a que cuando se generaban tipologías con toda la base junta, siempre se obtenía una base cuya única característica diferencial era ser docente de una IIEE privada. A pesar de solicitarse mayor cantidad de clusters este segmento permanecía sin dividirse, evitando visualizar dentro de este grupo la posible diversidad existente.

Para la segunda estrategia, que llamaremos tipología centrada en la satisfacción, se siguió el mismo procedimiento que en la estrategia centrada en la vocación y futuro docente. Se realizó una imputación sólo de las variables requeridas bajo esa estrategia, también utilizando el método de imputación multivariada de ecuaciones encadenadas, y se estimaron las alternativas de tipologías pero ya diferenciando entre IIEE públicas y privadas desde un inicio.

Finalmente se desarrollaron dos soluciones estables y pertinentes, tanto para IIEE públicas como para IIEE privadas, que fueron presentadas a la DIBRED para su discusión y elección. Las soluciones presentadas contenían las variables más importantes que caracterizaban cada tipología y una recomendación del consultor sobre cuál elegir, las cuales se presentan en forma resumida a continuación. En el cuadro 8 se presentan las variables utilizadas para generar la tipología de docentes de IIEE públicas centrada en vocación y futuro docente. Los principales resultados como tamaño de los segmentos en porcentajes y docentes estimados se presentan en el gráfico 2 y 3, y un resumen de sus principales características diferenciales en el cuadro 9.

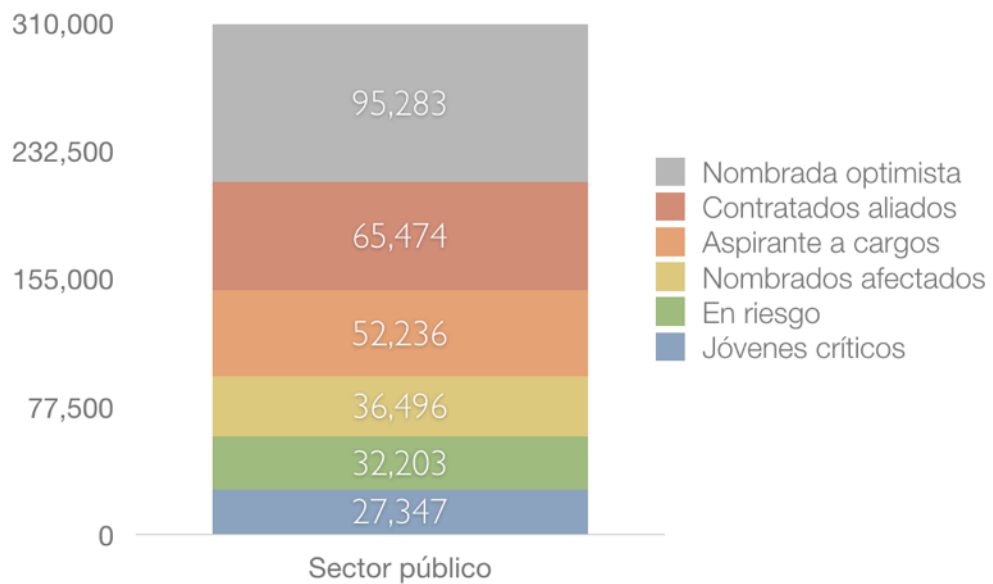
Cuadro 8: Variables para tipología sector público centrada en vocación y futuro docente

Área	Satisfacción con IIEE actual
Nivel que enseña	Volvería a ser docente
Sexo	Hijos docentes
Edad	Futuro laboral
NSE	Opinión LRM
Razón para ser docente	Logro aprendizaje IIEE públicas
Condición actual	Mejoría de IIEE públicas
	Futuro de IIEE públicas

Gráfico 2: Tipología sector público centrada en vocación y futuro docente, tamaño de segmentos en porcentajes



Gráfico 3: Tipología sector público centrada en vocación y futuro docente, tamaño de segmentos en docentes proyectados



Cuadro 9: Variables para tipología sector público centrada en vocación y futuro docente, principales características

	Jóvenes críticos	En riesgo	Nombrados afectados	Nombrado aspirante a cargos	Contratados aliados	Nombrada optimista
Cantidad	27,347	32,203	36,496	52,236	65,474	95,283
%	9	10	12	17	21	31
Edad	Menor de 45	Mayor de 45	Mayor de 45	Mayor de 45	Menor de 45	Entre 36 y 55
Sexo	Mixto	Mixto	Más mujeres	Hombre	Más mujeres	Mujer
Carga familiar	Alta	Promedio	Promedio	Promedio	Alta	Promedio
Área geográfica	Mixto	Más urbano	Más urbano	Mixto	Más rural	Más urbano
Nivel diferencial	Secundaria				Inicial	Inicial
Condición laboral	Más contratados	Nombrado	Nombrado	Nombrado	Contratado	Nombrado
Razón para ser docente	Única opción	Única opción	Vocación	Vocación	Niños	Vocación
Satisfacción con IIEE actual	Baja	Baja	Alta	Alta	Alta	Alta
Elegiría de nuevo ser docente	50%	No	Sí	Sí	Sí	Sí
Gustaría hijos docentes	No	No				
Visión a 5 años	Acompañante / especialista	Dejar la carrera	Seguir en su IIEE	Director	Acompañante / especialista	Seguir en su IIEE
Satisfacción con formación pedagógica	Baja	Baja	Media	Media	Alta	Alta
Opinión LRM	Indiferente	Perjudica	Perjudica	Beneficia	Indiferente	Beneficia
Percepción IIEE públicas	Negativa	Algo negativa	Negativa	Muy positiva	Neutral	Muy positiva
Confianza en instituciones	Baja	Baja	Promedio	Regular	Regular	Regular
Satisfacción con vida y trabajo	Baja	Baja	Promedio	Alta	Alta	Alta
Remuneración que recibe es justa	Muy baja	Muy baja	Muy baja	Muy baja	Muy baja	Muy baja
Tiene alguna otra ocupación	Otra IIEE	Negocios/otro		Negocios		

En el cuadro 10 se presentan las variables utilizadas para generar la tipología de docentes de IEE privadas centradas también en vocación y futuro docente, y los principales resultados en los gráficos 4 y 5, así como un resumen de sus características diferenciales en el cuadro 11.

Cuadro 10: Variables para tipología sector privado centrada en vocación y futuro docente

Nivel que enseña	Hijos docentes
Sexo	Futuro laboral
NSE	Logro aprendizaje IIEE públicas
Razón para ser docente	Mejoría de IIEE públicas
Satisfacción con IIEE actual	Futuro de IIEE públicas
Volvería a ser docente	

Gráfico 4: Tipología sector privado centrada en vocación y futuro docente, tamaño de segmentos en porcentajes

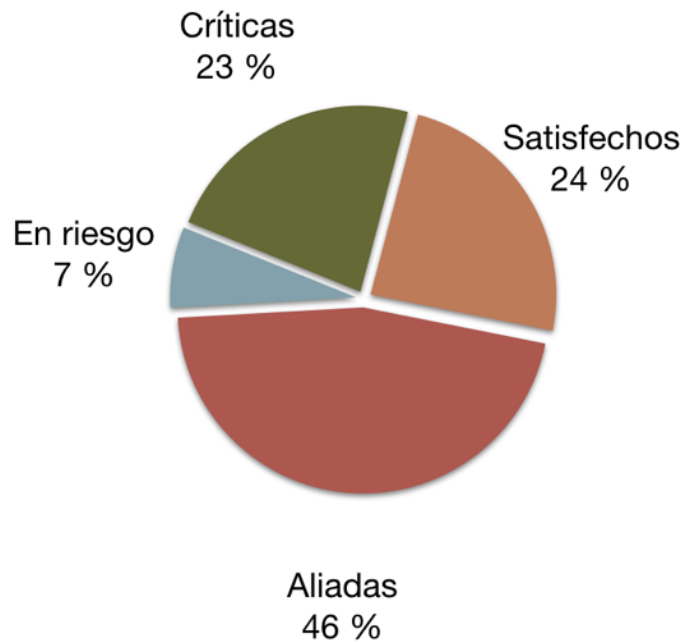
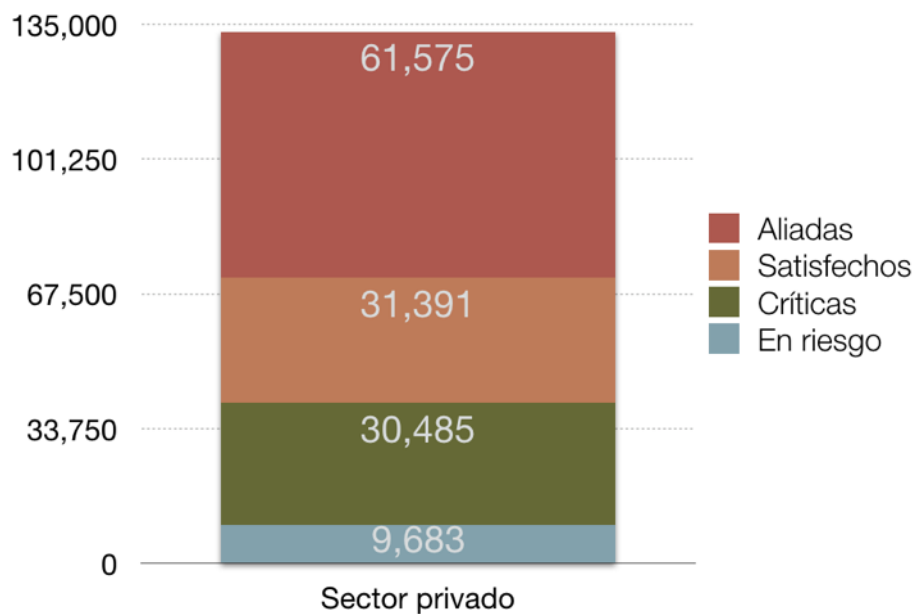


Gráfico 5: Tipología sector privado centrada en vocación y futuro docente, tamaño de segmentos en docentes proyectados



Cuadro11: Variables para tipología sector privado centrada en vocación y futuro docente, principales características

	<b>En riesgo</b>	<b>Críticas</b>	<b>Satisfechos</b>	<b>Aliadas</b>
Cantidad	<b>9,683</b>	<b>30,485</b>	<b>31,391</b>	<b>61,575</b>
%	7	23	24	46
Edad	Promedio	Promedio	Mayor de 45	Promedio
Sexo	Mixto	Mujer	Hombre	Mujer
Carga familiar	Promedio	Promedio	Promedio	Promedio
Área geográfica	Urbano	Urbano	Urbano	Urbano
Nivel diferencial		Inicial	Secundaria	Inicial
Condición laboral	Algunos sin contrato	Algunos sin contrato	Contratado	Contratado
Razón para ser docente	Única opción	Niños	Vocación	Niños
Satisfacción con IIEE actual	Baja	Alta	Alta	Alta
Elegiría de nuevo ser docente	No	Sí	Sí	Sí
Gustaría hijos docentes	No			
Visión a 5 años	Dejar la carrera	Ocupar cargos	Ocupar cargos	Ocupar cargos
Satisfacción con formación pedagógica	Baja	Alta	Media	Alta
Opinión LRM				
Percepción IIEE públicas	Neutral	Negativa	Neutral	Muy positiva
Confianza en instituciones	Muy baja	Neutral	Neutral	Alta
Satisfacción con vida y trabajo	Baja	Muy alta	Muy alta	Muy alta
Remuneración que recibe es justa	Media	Media	Alta	Media
Tiene alguna otra ocupación		Negocios		Negocios

Por otro lado, en el cuadro 12 se presentan las variables utilizadas para generar las tipologías de docentes de IIEE públicas centradas en variables de satisfacción y confianza en instituciones, y en los gráficos 6 y 7 los tamaños de segmentos y cantidades de docentes estimados, así como en el cuadro 13 sus principales características.

Cuadro 12: Variables para tipología sector público centrada en satisfacción y confianza en instituciones

Área	Satisfacción con la vida, salud, trabajo, jubilación y relaciones familiares
Nivel que enseña	Satisfacción con logros de estudiantes, padres de familia, salarior, reconocimiento Minedu
Sexo	Razón para ser docente
Edad	Futuro laboral
NSE	Confianza en instituciones (Minedu, UGEL, SUTEP)
Condición actual	

Gráfico 6: Tipología sector público centrada en satisfacción y confianza en instituciones, tamaño de segmentos en porcentajes

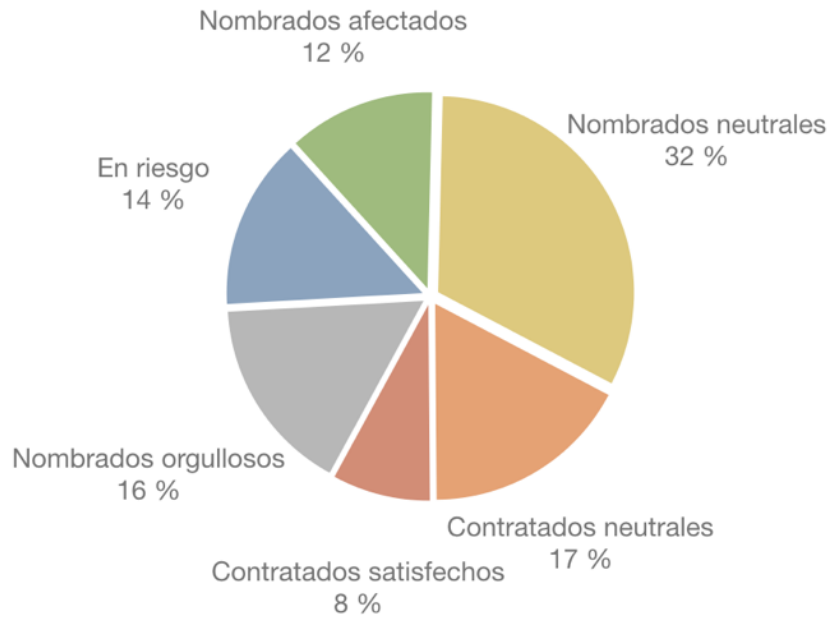
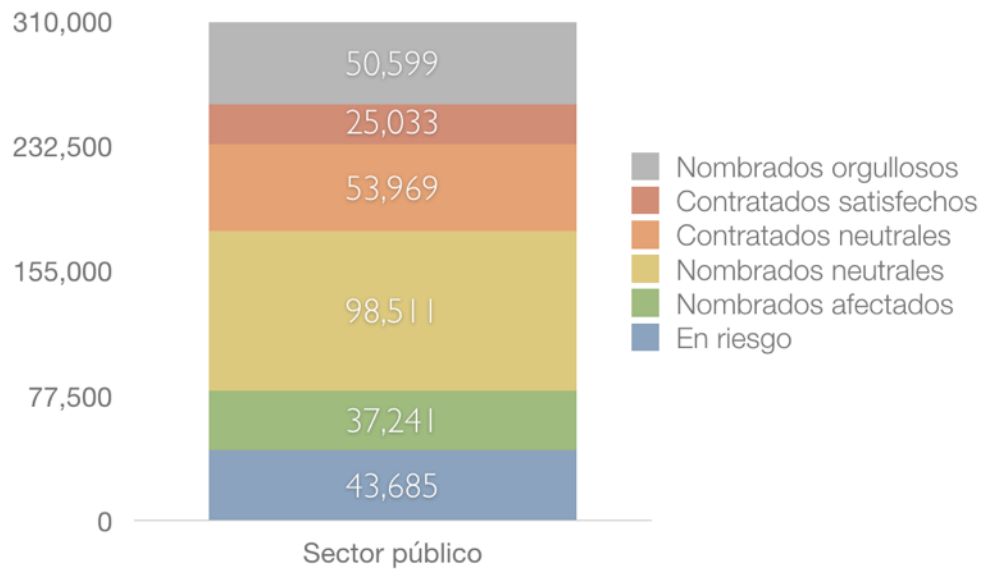


Gráfico 7: Tipología sector público centrada en satisfacción y confianza en instituciones, tamaño de segmentos en docentes proyectados



Cuadro13: Variables para tipología sector público centrada en satisfacción y confianza en instituciones, principales características

	En riesgo	Nombrados afectados	Nombrados neutrales	Contratados neutrales	Contratados satisfechos	Nombrados orgullosos
Cantidad	43,685	37,241	98,511	53,969	25,033	50,599
%	14	12	32	17	8	16
Edad	Entre 36 y 55	Entre 36 y 55	Mayor de 45	Menor de 45	Menor de 45	Mayor de 45
Sexo	Mixto	Mixto	Mixto	Mixto	Mixto	Mixto
Carga familiar	Promedio	Promedio	Promedio	Alta	Alta	Promedio
Área geográfica	Más urbano	Más urbano	Más urbano	Más rural	Más rural	Más urbano
Nivel diferencial	Secundaria	Secundaria	Primaria	Primaria	Inicial	Primaria
Condición laboral	Más nombrado	Nombrado	Nombrado	Contratado	Contratado	Nombrado
Razón para ser docente	Única opción	Vocación	Vocación / niños	Niños	Niños	Vocación
Satisfacción con IIEE actual	Baja	Alta	Alta	Alta	Alta	Alta
Elegiría de nuevo ser docente	60%	Sí	Sí	Sí	Sí	Sí
Gustaría hijos docentes	No	Sí	Sí	Sí	Sí	Sí
Visión a 5 años	Dejar la carrera	Acompañante	Director	Acompañante / especialista	Directivo / acompañante / especialista	Director
Satisfacción con formación pedagógica	Media	Alta	Media	Media	Alta	Alta
Opinión LRM	Perjudica	Perjudica	Perjudica	Indiferente	Indiferente	Beneficia
Percepción IIEE públicas	Negativa	Neutral	Neutral	Neutral	Neutral	Muy positiva
Confianza en instituciones	Muy baja	Muy baja	Neutral	Neutral	Alta	Alta
Satisfacción con vida y trabajo	Baja	Muy alta	Promedio	Promedio	Muy alta	Muy alta
Remuneración que recibe es justa	Muy baja	Muy baja	Muy baja	Muy baja	Muy baja	Muy baja
Tiene alguna otra ocupación			Negocios			

Finalmente, en el cuadro 14 se presentan las variables utilizadas para generar las tipologías de docentes de IIEE privadas centradas en variables de satisfacción y confianza en instituciones, así como los tamaños de segmentos y cantidad de docentes estimados en los gráficos 8 y 9 y sus principales características en el cuadro 15.

Cuadro 14: Variables para tipología sector privado centrada en satisfacción y confianza en instituciones

Nivel que enseña	Satisfacción con la vida, salud, trabajo, jubilación y relaciones familiares
Sexo	Satisfacción con logros de estudiantes, padres de familia, salario, reconocimiento Minedu
NSE	Confianza en instituciones (Minedu, UGEL, SUTEP)
Edad	
Razón para ser docente	
Futuro laboral	



Gráfico 8: Tipología sector privado centrada en satisfacción y confianza en instituciones, tamaño de segmentos en porcentajes

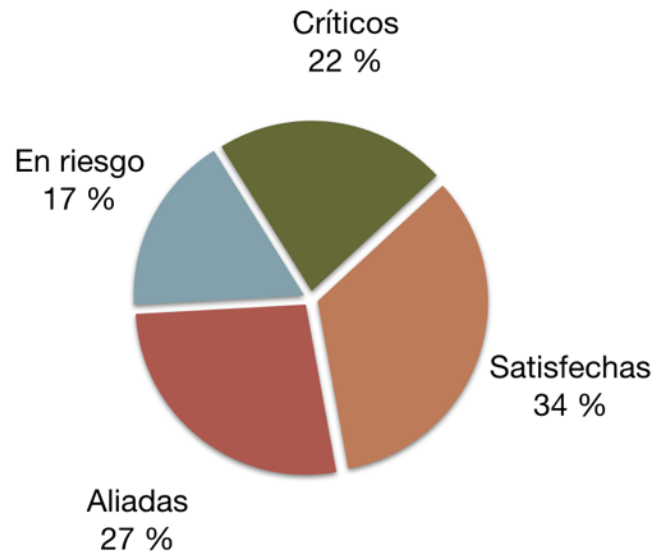
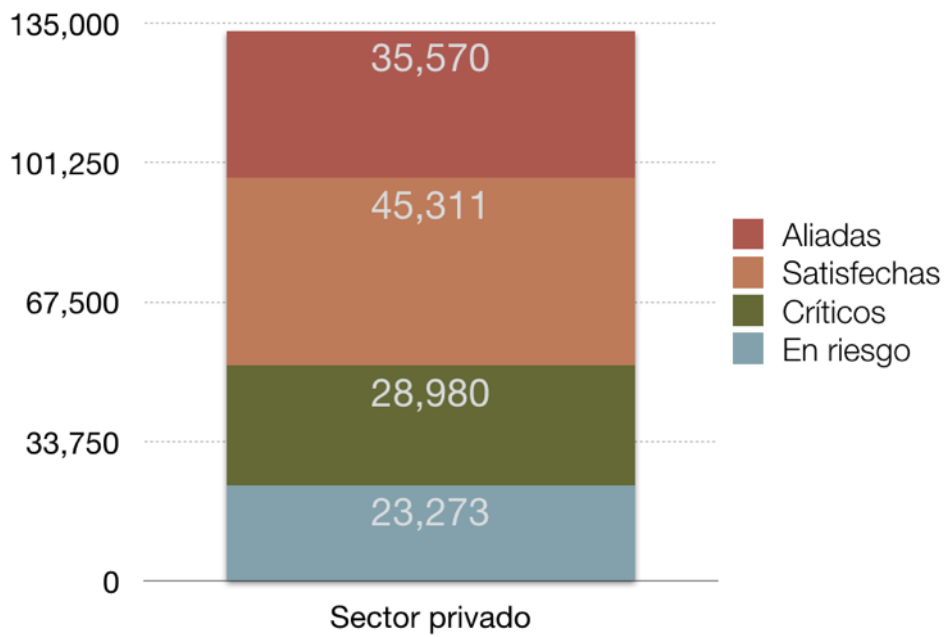


Gráfico 9: Tipología sector privado centrada en satisfacción y confianza en instituciones, tamaño de segmentos en docentes proyectados



Cuadro15: Variables para tipología sector privado centrada en satisfacción y confianza en instituciones, principales características

	En riesgo	Críticos	Satisfechas	Aliadas
Cantidad	23,273	28,980	45,311	35,570
%	17	22	34	27
Edad	Promedio	Mayor de 45	Menor de 35	Promedio
Sexo	Más mujeres	Más hombres	Mujer	Más mujeres
Carga familiar	Promedio	Promedio	Promedio	Promedio
Área geográfica	Urbano	Urbano	Urbano	Urbano
Nivel diferencial		Secundaria	Inicial	
Condición laboral	Algunos sin contrato	Contratado	Algunos sin contrato	Contratado
Razón para ser docente	Vocación	Vocación / estabilidad	Niños / buen ejemplo	Vocación
Satisfacción con IIEE actual	Media	Alta	Alta	Alta
Elegiría de nuevo ser docente	19% no	Sí	Sí	Sí
Gustaría hijos docentes	No			Sí
Visión a 5 años	Dejar la carrera / cambiarse de IIEE		Directivo	Directivo
Satisfacción con formación pedagógica	Media	Media	Alta	Alta
Opinión LRM				
Percepción IIEE públicas	Negativa	Negativa	Positiva	Positiva
Confianza en instituciones	Muy baja	Neutral	Alta	Muy alta
Satisfacción con vida y trabajo	Baja	Media alta	Media alta	Muy alta
Remuneración que recibe es justa	Baja	Alta	Media	Alta
Tiene alguna otra ocupación		Otros		

## 2.5. Perfilamiento de la tipología, análisis de resultados e informe

La solución elegida fue la primera, centrada en la vocación y futuro docente, y su opinión de la LRM (ley de la Reforma Magisterial). Sobre la base de esta última es que se perfilaron los segmentos en función a las diversas variables identificadas previamente y se desarrolló un informe en formato .ppt sobre las características de los mismos, adjunto al presente documento en formato digital y en el Anexo 6.

## 2.6. Clasificación posterior

Para facilitar la clasificación futura de nuevos docentes en las tipologías encontradas, se desarrolló una rutina en el programa R para tal fin. En esta rutina se detallan las bases que se deben leer en el programa, las instrucciones para recrear la tipología, así como el formato en que se deben ingresar las nuevas respuestas, la instrucción para clasificar a los docentes en base a ellas, y guardar en una nueva base la pertenencia al segmento requerido. Todo esto se detalla en la sintaxis adjunta en el Anexo 5.

## V. Bibliografía

Arellano Cueva, R. (2003) Los estilos de vida en el Perú: cómo somos y pensamos los peruanos del siglo XXI. Lima, Arellano Investigaciones de Marketing S. A.

Ayala Cornejo, M. (2016) Tipos de liderazgo y desempeño docente en el nivel secundario de las instituciones educativas. Universidad Nacional Del Centro Del Perú, Escuela De Posgrado, Unidad De Posgrado De La Facultad De Educación. Tesis para optar por el grado de Magister

Borgobello, A., Peralta, N. y N. Rosselló (2010) El estilo docente universitario en relación al tipo de clase y a la disciplina enseñada. Laboratorio de Psicología Social Experimental de IRICE-CONICET. LIBERABIT: Lima (Perú) 16(1): 7-16

Cuenca, R. y Carrillo, S. (2017) Una mirada a la profesión docente en el Perú: futuros docentes, docentes en servicio y formadores de docentes. Ministerio de Educación y UNESCO, Lima

Cuenca, R., Montero, C., Ames, P. y V. Rojas (2011) Hacia una propuesta de criterios de buen desempeño docente : estudios que aportan a la reflexión, al diálogo y a la construcción concertada de una política educativa. CNM y FSM, Lima.

Cuenca, R. y Portocarrero, C. (2003) Actitudes y valoración de los docentes en servicio hacia su profesión. Ministerio de Educación, Lima

Díaz, J. y Ñopo, H. (2016) La carrera docente en el Perú. Investigación para el desarrollo en el Perú. Once balances. Lima: GRADE.

Deepa R. y Ieva S. (2012) Exploring the potential of role play in higher education: development of a typology and teacher guidelines .Innovations in Education and Teaching International Vol. 49, No. 4, November 2012, 427–436

Domínguez, R. y Martínez, L. (2012) Tipología de la docencia en una universidad pública del noroeste de México. Revista Iberoamericana de Evaluación Educativa 2012 - Volumen 5, Número 1e

Dutt, Ashish & Aghabozrgi, Saeed & Ismail, Maizatul Akmar & Mahroeian, Hamidreza. (2015). Clustering Algorithms Applied in Educational Data Mining. International Journal of Information and Electronics Engineering (IJIEE). 5.

Gerda J. Visser-Wijnveen, Ann Stes & Peter Van Petegem (2014) Clustering teachers' motivations for teaching, Teaching in Higher Education, 19:6, 644-656.

Guadalupe, C., León, J., Rodríguez, J. y S. Vargas (2017) Estado de la educación en el Perú Análisis y perspectivas de la educación básica. Grade, Lima.

- Hanne T. & R. Vanderlinde (2014) Teacher Educators' Professional Development: Towards a Typology of Teacher Educators' Researcherly Disposition. *British Journal of Educational Studies*, 62:3, 297-315
- Joyce Hwee Ling Koh & Ching Sing Chai (2013) Teacher clusters and their perceptions of technological pedagogical content knowledge (TPACK) development through ICT lesson design. *Computers & Education* 70 (2014) 222–232
- Linzer, D. y Lewis, J. (2011) poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*. June 2011, Volume 42, Issue 10.
- Lloret-Segura, S. ,Ferrerres-Traver, A., Hernández-Baeza A. y Tomás-Marco, I. (214). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, vol. 30, nº 3 (octubre), 1151-1169
- Magidson J. y Vermunt, J.K. (2001). "Latent Class Factor and Cluster Models, Bi-plots and Related Graphical Displays". Chapter 5 in Becker and Sobel (Eds.) *Sociological Methodology*. Vol. 31, 223-264
- Magidson J. y Vermunt, J.K. (2002). "Latent Class Modeling as a Probabilistic Extension of K-Means Clustering". *Quirk's Marketing Research Review*, N°20, 77-80.
- Magidson J. y Vermunt, J.K. (2002b). "Latent Class Models for Clustering: a Comparison with K-means". *Canadian Journal of Marketing Research*, N°20, 36-43.
- Magidson J. y Vermunt, J.K. (2003). *Latent Class Models*. Belmont: Statistical Innovations
- Magidson J. y Vermunt J.K. (2003b). "A nontechnical introduction to latent class models". *DMA Research Council Journal*.
- Malbernat, R. (2013) Incorporar actividades virtuales en educación superior: Algoritmo de Segmentación de docentes según sus competencias. XV Workshop de Investigadores en Ciencias de la Computación. Paraná, Entre Ríos.
- Mama-Timotheou M. & Hennessy S. (2013) Developing a typology of beliefs and practices concerning classroom use of ICT, *Computers & Education*
- Orme B. y Johnson R. (2008). *Improving K-Means Cluster Analysis: Ensemble Analysis Instead of Highest Reproducibility Replicates*. Sequim: Sawtooth Software 2008
- Rahmat, A. (2017) Clustering in Education. *European Research Studies Journal* Volume XX, Issue 3A, 2017
- Rojas, J. (2013) Clima escolar y tipología docente: la violencia escolar en las prácticas educativas. *Cuadernos de Investigación Educativa*, Vol. 4, N° 19, 2013, Montevideo (Uruguay), 87-104.

Shavelson, R. (1979) Applications of Cluster Analysis in Educational Research: looking for a needle in a haystack. *British Educational Research Journal*, Vol. 5, No. 1, 1979

van Buuren, S. y Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. December 2011, Volume 45, Issue 3.

Vermunt J.K. y Magidson J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont, MA: Statistical Innovations Inc.

## Anexo 1:

### Sintaxis SPSS para adaptación de variables y estimación de NSE de base ENDO 2016

\* RECOD CONDICION ACTUAL 1 NOMBRADO 0 CONTRATADO OTROS.

```
RECODE p306 (1=1) (ELSE=0) INTO p306_r.  
VAL LAB p306_r  
1 'Nombrado'  
0 'Contratado o sin contrato'.  
EXECUTE.
```

\*RECOD EDAD EN RANGOS.

```
RECODE p109_d_01  
(999=SYSMIS)  
(0 thru 35=1)  
(36 thru 45 =2)  
(46 thru 55 =3)  
(56 thru 99 = 4)  
INTO edad.  
EXECUTE.
```

```
VAL LAB edad  
1 'Hasta 35 años'  
2 '36 a 45 años'  
3 '46 a 55 años'  
4 '55 años a mas'  
.
```

\*RECOD OPINION LRM INCLUYENDO PRIVADO.

```
RECODE p811 (SYSMIS=5) (ELSE=Copy) INTO p811_r.  
EXECUTE.
```

```
VAL LAB p811_r  
1 'Lo beneficia'  
2 'Lo perjudica'  
3 'Ni lo beneficia ni perjudica'  
4 'No conoce la ley'  
5 'Sector privado'  
9 'NEP'.
```

\*RECOD OPINION PARTES DE LA LRM INCLUYENDO PRIVADO Y NO CONOCE LA LEY.

```
DO IF p811_r=5.  
COMPUTE p812_1_r=6.  
ELSE IF p811_r=9.  
COMPUTE p812_1_r=9.  
ELSE IF p811_r=4.  
COMPUTE p812_1_r=5.  
ELSE IF p811_r<4.  
COMPUTE p812_1_r=p812_1.  
END IF.  
execute.
```

```
DO IF p811_r=5.  
COMPUTE p812_2_r=6.  
ELSE IF p811_r=9.  
COMPUTE p812_2_r=9.  
ELSE IF p811_r=4.  
COMPUTE p812_2_r=5.  
ELSE IF p811_r<4.  
COMPUTE p812_2_r=p812_2.  
END IF.  
execute.
```

```
DO IF p811_r=5.
COMPUTE p812_3_r=6.
ELSE IF p811_r=9.
COMPUTE p812_3_r=9.
ELSE IF p811_r=4.
COMPUTE p812_3_r=5.
ELSE IF p811_r<4.
COMPUTE p812_3_r=p812_3.
END IF.
execute.
```

```
DO IF p811_r=5.
COMPUTE p812_4_r=6.
ELSE IF p811_r=9.
COMPUTE p812_4_r=9.
ELSE IF p811_r=4.
COMPUTE p812_4_r=5.
ELSE IF p811_r<4.
COMPUTE p812_4_r=p812_4.
END IF.
execute.
```

```
DO IF p811_r=5.
COMPUTE p812_5_r=6.
ELSE IF p811_r=9.
COMPUTE p812_5_r=9.
ELSE IF p811_r=4.
COMPUTE p812_5_r=5.
ELSE IF p811_r<4.
COMPUTE p812_5_r=p812_5.
END IF.
execute.
```

```
DO IF p811_r=5.
COMPUTE p812_6_r=6.
ELSE IF p811_r=9.
COMPUTE p812_6_r=9.
ELSE IF p811_r=4.
COMPUTE p812_6_r=5.
ELSE IF p811_r<4.
COMPUTE p812_6_r=p812_6.
END IF.
execute.
```

```
DO IF p811_r=5.
COMPUTE p812_7_r=6.
ELSE IF p811_r=9.
COMPUTE p812_7_r=9.
ELSE IF p811_r=4.
COMPUTE p812_7_r=5.
ELSE IF p811_r<4.
COMPUTE p812_7_r=p812_7.
END IF.
execute.
```

```
VAL LAB p812_1_r p812_2_r p812_3_r p812_4_r p812_5_r p812_6_r p812_7_r
1 'Totalmente en desacuerdo'
2 'En desacuerdo'
3 'De acuerdo'
4 'Totalmente de acuerdo'
5 'No conoce la ley'
6 'Sector privado'
9 'NEP'.
```

\*\*\*CALCULO DE NSE.

\*BIENES.

```
RECODE p108_05 p108_07 p108_08 p108_09
(1=1)
(ELSE=0)
INTO computadora lavadora refrigeradora microondas.
EXECUTE.
```

```
RECODE p108_11 p108_15
(1=5)
(ELSE=0)
INTO auto domestico.
EXECUTE.
```

```
compute N2=(computadora + lavadora + refrigeradora + microondas)*2.
COMPUTE N3=auto+domestico.
```

\*MATERIAL PISO.

```
DO IF p105=1.
COMPUTE N4=8.
ELSE IF p105=2.
COMPUTE N4=7.
ELSE IF p105=3.
COMPUTE N4=5.
ELSE IF p105=4.
COMPUTE N4=3.
ELSE IF p105=5.
COMPUTE N4=3.
ELSE.
COMPUTE N4=0.
END IF.
EXECUTE.
```

\*PRESTACION SALUD JEFE DE HOGAR - NO EXISTE LA VARIABLE SE REEMPLAZA POR ESTA REFERIDA AL DOCENTE.

```
DO IF P405=1.
COMPUTE N5=2.
ELSE IF P405=2.
COMPUTE N5=0.
ELSE IF P405=3.
COMPUTE N5=6.
ELSE.
COMPUTE N5=0.
END IF.
EXECUTE.
```

\*MATERIAL PAREDES.

```
DO IF P104=1.
COMPUTE N6=6.
ELSE IF P104=2.
COMPUTE N6=4.
ELSE IF P104<8 AND P104>2.
COMPUTE N6=2.
ELSE.
COMPUTE N6=0.
END IF.
EXECUTE.
```

\*BAÑO.



```
DO IF P106=1.
COMPUTE N7=5.
ELSE IF P106=2.
COMPUTE N7=3.
ELSE IF P106>2 AND P106<7.
COMPUTE N7=1.
ELSE.
COMPUTE N7=0.
END IF.
EXECUTE.
```

\*\*\*ESTIMACION PUNTAJE EDUCACION JEFE DE HOGAR.

```
DO IF P110<=2.
compute N1_a=0.
else if p110>2 and p110<6.
compute N1_a=1.
else if p110>5 and p110<8.
compute N1_a=2.
else if p110=8.
compute n1_a=3.
else if p110=9.
compute n1_a=4.
else if p110=10.
compute n1_a=5.
else if p110>10 and p110<13.
compute n1_a=6.
ELSE.
compute n1_a=99.
end if.
execute.
```

```
RECODE N1_a (MISSING=99).
EXECUTE.
```

\*ESTIMACION PUNTAJE DOCENTES JEFES DE HOGAR.

```
*POSTGRADO.
do if p212_1=1 or p212_2=1.
compute postgrado=1.
else.
compute postgrado=0.
end if.
execute.
```

```
*PUNTAJE POR POSGRADO.
do if n1_a=99 and (postgrado=1).
compute n1_b=6.
end if.
```

```
*PUNTAJE POR NO HABER TERMINADO CARRERA.
do IF n1_a=99 and p202=3.
compute n1_b=0.
end if.
execute.
```

```
*PUNTAJE POR UNIVERSIDAD EN PEDAGOGIA.
do if n1_a=99 and postgrado=0 and p208_1=2.
compute n1_b=5.
end if.
execute.
```

```

*PUNTAJE POR INSITUTO EN PEDAGOGIA.
do if n1_a=99 and postgrado=0 and (p208_1=1 OR P208_1>2).
compute n1_b=3.
end if.
execute.

*PUNTAJE DE JEFE DE HOGAR.
do if n1_a<99.
compute n1_b=n1_a.
end if.
execute.

*RESTO SON PROFESORES QUE ESTUDIARON OTRA PROFESION, SE REVISÓ CARRERAS SE ASUME ESTUDIOS
UNIV COMPLETOS.
RECODE n1_b (MISSING=5).
EXECUTE.

*PUNTAJE FINAL POR NIVEL EDUCATIVO DEL JEFE DE HOGAR
compute N1=n1_b.

*PUNTAJE NSE.
compute NSE_PTJE = N1+N2+N3+N4+N5+N6+N7.

RECODE NSE_PTJE
(0 thru 12=1)
(13 thru 19 =2)
(20 thru 22 =3)
(23 thru 28 = 4)
(29 thru 33 = 5)
(34 thru 39 = 6)
(40 thru 47 = 7)
(48 thru 98 = 8)
INTO NSE_1.
EXECUTE.

VAL LAB NSE_1
1 'NSE E'
2 'NSE D'
3 'NSE C2'
4 'NSE C1'
5 'NSE B2'
6 'NSE B1'
7 'NSE A2'
8 'NSE A1'
.

RECODE NSE_1
(1=1)
(2 =2)
(3 thru 4 =3)
(5 thru 6 = 4)
(7 thru 8 = 5)
INTO NSE.
EXECUTE.

VAL LAB NSE
1 'NSE E'
2 'NSE D'
3 'NSE C'
4 'NSE B'
5 'NSE A'
.

RECODE NSE
(1=1)

```

```
(2 =2)  
(3 =3)  
(4 = 4)  
(5 = 4)  
INTO NSE_2.  
EXECUTE.
```

```
VAL LAB NSE_2  
1 'NSE E'  
2 'NSE D'  
3 'NSE C'  
4 'NSE AB'  
.
```

**Anexo 2:**  
**Sintaxis SPSS para cálculo de variables personas que viven en el hogar de base**  
**P109\_ENDO 2016 traspuesta**

```
DO IF p109_d_0<12.
compute menor=1.
else.
compute menor=0.
end if.
execute.
```

```
compute menores=sum(menor.1 to menor.10).
EXECUTE.
```

```
do if menores>0.
compute menor=1.
else.
compute menor=0.
end if.
execute.
freq var menor.
```

```
do if (p109_b_0.1=2
OR p109_b_0.2=2
OR p109_b_0.3=2
OR p109_b_0.4=2
OR p109_b_0.5=2
OR p109_b_0.6=2
OR p109_b_0.7=2
OR p109_b_0.8=2
OR p109_b_0.9=2
OR p109_b_0.10=2).
COMPUTE conyugue=1.
ELSE.
compute conyugue=0.
end if.
execute.
RECODE conyugue (MISSING=0).
freq var conyugue.
```

```
do if (p109_b_0.1=3
OR p109_b_0.2=3
OR p109_b_0.3=3
OR p109_b_0.4=3
OR p109_b_0.5=3
OR p109_b_0.6=3
OR p109_b_0.7=3
OR p109_b_0.8=3
OR p109_b_0.9=3
OR p109_b_0.10=3).
COMPUTE hijos=1.
ELSE.
compute hijos=0.
end if.
execute.
RECODE hijos (MISSING=0).
freq var hijos.
```

```
do if (p109_b_0.1=4
OR p109_b_0.2=4
OR p109_b_0.3=4
OR p109_b_0.4=4
OR p109_b_0.5=4
OR p109_b_0.6=4
OR p109_b_0.7=4
OR p109_b_0.8=4
OR p109_b_0.9=4
OR p109_b_0.10=4).
COMPUTE padres=1.
ELSE.
```

```

compute padres=0.
end if.
execute.
RECODE padres (MISSING=0).
freq var padres.

do if (p109_b_0.1=5
OR p109_b_0.2=5
OR p109_b_0.3=5
OR p109_b_0.4=5
OR p109_b_0.5=5
OR p109_b_0.6=5
OR p109_b_0.7=5
OR p109_b_0.8=5
OR p109_b_0.9=5
OR p109_b_0.10=5).
COMPUTE familiar=1.
ELSE.
compute familiar=0.
end if.
execute.
RECODE familiar (MISSING=0).
freq var familiar.

do if (p109_b_0.1=6
OR p109_b_0.2=6
OR p109_b_0.3=6
OR p109_b_0.4=6
OR p109_b_0.5=6
OR p109_b_0.6=6
OR p109_b_0.7=6
OR p109_b_0.8=6
OR p109_b_0.9=6
OR p109_b_0.10=6).
COMPUTE otro_no_f=1.
ELSE.
compute otro_no_f=0.
end if.
execute.
RECODE otro_no_f (MISSING=0).
freq var otro_no_f.

```

### Anexo 3:

## Sintaxis R de análisis factorial exploratorio de ítems de confianza hacia instituciones, satisfacción con la vida y con el trabajo

```
#ACTIVA LOS PAQUETES EN R
library(psych)
library(GPArotation)
library(dplyr)
library(MVN)

getwd()
#COLOCAR RUTA DONDE SE ENCUENTRA LA BASE DE DATOS EN FORMATO CSV
#setwd("/Users/Tipologia docentes/Bases")

#LEE BASE DE DATOS
endo_2016_escalas <- read.csv("endo_2016_escalas.csv")

summary(endo_2016_escalas)

#CONVERTIR CODIGOS A MISSING FORMATO R
escalas<-endo_2016_escalas[c(2:27,35)]
lrm<-endo_2016_escalas[28:34]
cuestionario<-endo_2016_escalas[1]

str(escalas)
str(lrm)

escalas[escalas==9] <- NA
lrm[lrm==5] <- NA
lrm[lrm==6] <- NA
lrm[lrm==9] <- NA

endo_2016_escalas<-cbind(cuestionario,escalas,lrm)

#####FACTORIAL ESCALA CONFIANZA#####
confianza <- endo_2016_escalas[2:8]
head(confianza)

#####INSPECCIONAR NORMALIDAD MULTIVARIANTE
confianza_500<- sample_n(confianza, 500, replace = FALSE)
confianza_500 <- na.omit(confianza_500)

mvn(data = confianza_500, univariatePlot = "histogram")

#GRÁFICOS DE NORMALIDAD MULTIVARIANTE
par(mfrow=c(1,1))
mvn(data = confianza_500, multivariatePlot = "qq", desc = FALSE)

#TEST DE NORMALIDAD MULTIVARIANTE
mvn(data = confianza_500, mvnTest = "hz")

#####ANALISIS FACTORIAL

#CURTOSIS Y ASIMETRIA
describe(confianza)

#CORRELACIONES Y MAPA DE CALOR
matriz_confianza<-lowerCor(confianza)
png('confianza.png')
cor.plot(confianza,numbers=TRUE,main="Heatmap")
dev.off()

#INDICE DE ADECUACION DE Kaiser-Meyer-Olkin Y ALPHA
KMO(confianza)
alpha(confianza)

#CANTIDAD DE FACTORES
nfactors(confianza)
```

```

fa.parallel(confianza, main="Parallel Analysis - Ejemplo", cor="cor")

#RESTAURA LA PANTALLA A UN SOLO RECUADRO
par(mfrow=c(1,1))

#ESTIMACION 3 FACTORES
fa_confianza <- fa(confianza,3)
fa_confianza

plot(fa_confianza)

#COLOCAR RUTA DONDE SE DESEE GRABAR OUTPUTS
#setwd("/Users/Tipologia docentes/R")
write.csv(fa_confianza$loadings,"fa_confianza_loadings.csv")
write.csv(cbind(fa_confianza$communalities,fa_confianza
$uniquenesses),"fa_confianza_comunalidad.csv")
write.csv(fa_confianza$Vaccounted,"fa_confianza_eig.csv")
write.csv(fa_confianza$Phi, "fa_confianza_corr.csv")

#####FACTORIAL SATISFACCION CON LA VIDA#####
vida <- endo_2016_escalas[9:17]
head(vida)

#####INSPECCIONAR NORMALIDAD MULTIVARIANTE
vida_500<- sample_n(vida, 500, replace = FALSE)
vida_500 <- na.omit(vida_500)

mvn(data = vida_500, univariatePlot = "histogram")

#GRÁFICOS DE NORMALIDAD MULTIVARIANTE
par(mfrow=c(1,1))
mvn(data = vida_500, multivariatePlot = "qq", desc = FALSE)

#TEST DE NORMALIDAD MULTIVARIANTE
mvn(data = vida_500, mvnTest = "hz")

#####ANALISIS FACTORIAL

#CURTOSIS Y ASIMETRIA
describe(vida)

#CORRELACIONES Y MAPA DE CALOR
matriz_vida<-lowerCor(vida)
png('vida.png')
cor.plot(vida,numbers=TRUE,main="Heatmap")
dev.off()

#INDICE DE ADECUACION DE Kaiser-Meyer-Olkin Y ALPHA
KMO(vida)
alpha(vida)

#CANTIDAD DE FACTORES
nfactors(vida)
fa.parallel(vida, main="Parallel Analysis - Ejemplo", cor="cor")

#RESTAURA LA PANTALLA A UN SOLO RECUADRO
par(mfrow=c(1,1))

#ESTIMACION 3 FACTORES
fa_vida <- fa(vida,3)
fa_vida

plot(fa_vida)

#COLOCAR RUTA DONDE SE DESEE GRABAR OUTPUTS
#setwd("/Users/Tipologia docentes/R")
write.csv(fa_vida$loadings,"fa_vida_loadings.csv")
write.csv(cbind(fa_vida$communalities,fa_vida$uniquenesses),"fa_vida_comunalidad.csv")

```

```

write.csv(fa_vida$Vaccounted,"fa_vida_eig.csv")
write.csv(fa_vida$Phi, "fa_vida_corr.csv")

#####FACTORIAL SATISFACCION CON DOCENCIA#####
#SE ELIMINO ITEM 6 PORQUE NO APLICA A TODOS LOS DOCENTES
docencia <- endo_2016_escalas[,c(18:22,24:27)]
head(docencia)

#####INSPECCIONAR NORMALIDAD MULTIVARIANTE
docencia_500<- sample_n(docencia, 500, replace = FALSE)
docencia_500 <- na.omit(docencia_500)

mvn(data = docencia_500, univariatePlot = "histogram")

#GRÁFICOS DE NORMALIDAD MULTIVARIANTE
par(mfrow=c(1,1))
mvn(data = docencia_500, multivariatePlot = "qq", desc = FALSE)

#TEST DE NORMALIDAD MULTIVARIANTE
mvn(data = docencia_500, mvnTest = "hz")

#####ANALISIS FACTORIAL

#CURTOSIS Y ASIMETRIA
describe(docencia)

#CORRELACIONES Y MAPA DE CALOR
matriz_docencia<-lowerCor(docencia)
png('docencia.png')
cor.plot(docencia,numbers=TRUE,main="Heatmap")
dev.off()

#INDICE DE ADECUACION DE Kaiser-Meyer-Olkin Y ALPHA
KMO(docencia)
alpha(docencia)

#CANTIDAD DE FACTORES
nfactors(docencia)
fa.parallel(docencia, main="Parallel Analysis - Ejemplo", cor="cor")

#RESTAURA LA PANTALLA A UN SOLO RECUADRO
par(mfrow=c(1,1))

#ESTIMACION 3 FACTORES
fa_docencia <- fa(docencia,3)
fa_docencia

plot(fa_docencia)

#COLOCAR RUTA DONDE SE DESEE GRABAR OUTPUTS
#setwd("/Users/Tipologia docentes/R")
write.csv(fa_docencia$loadings,"fa_docencia_loadings.csv")
write.csv(cbind(fa_docencia$communalities,fa_docencia
$uniquenesses),"fa_docencia_comunalidad.csv")
write.csv(fa_docencia$Vaccounted,"fa_docencia_eig.csv")
write.csv(fa_docencia$Phi, "fa_docencia_corr.csv")

#####FACTORIAL LRM#####
##### DESCARTADO LUEGO COMO ITEM PARA SEGMENTAR#####
escala_lrm <- endo_2016_escalas[29:35]
head(escala_lrm)

#####INSPECCIONAR NORMALIDAD MULTIVARIANTE
escala_lrm_500<- sample_n(escala_lrm, 500, replace = FALSE)
escala_lrm_500 <- na.omit(escala_lrm_500)

mvn(data = escala_lrm_500, univariatePlot = "histogram")

```



```

#GRÁFICOS DE NORMALIDAD MULTIVARIANTE
par(mfrow=c(1,1))
mvn(data = escala_lrm_500, multivariatePlot = "qq", desc = FALSE)

#TEST DE NORMALIDAD MULTIVARIANTE
mvn(data = escala_lrm_500, mvnTest = "hz")

#####ANALISIS FACTORIAL

#CURTOSIS Y ASIMETRIA
describe(escala_lrm)

#CORRELACIONES Y MAPA DE CALOR
matriz_escala_lrm<-lowerCor(escala_lrm)
png('escala_lrm.png')
cor.plot(escala_lrm,numbers=TRUE,main="Heatmap")
dev.off()

#INDICE DE ADECUACION DE Kaiser-Meyer-Olkin Y ALPHA
KMO(escala_lrm)
alpha(escala_lrm)

#CANTIDAD DE FACTORES
nfactors(escala_lrm)
fa.parallel(escala_lrm, main="Parallel Analysis - Ejemplo", cor="cor")

#RESTAURA LA PANTALLA A UN SOLO RECUADRO
par(mfrow=c(1,1))

#ESTIMACION 3 FACTORES
fa_escala_lrm <- fa(escala_lrm,2)
fa_escala_lrm

plot(fa_escala_lrm)

#COLOCAR RUTA DONDE SE DESEE GRABAR OUTPUTS
#setwd("/Users/Tipologia docentes/R")
write.csv(fa_escala_lrm$loadings,"fa_escala_lrm_loadings.csv")
write.csv(cbind(fa_escala_lrm$communalities,fa_escala_lrm
$uniquenesses),"fa_escala_lrm_comunalidad.csv")
write.csv(fa_escala_lrm$Vaccounted,"fa_escala_lrm_eig.csv")
write.csv(fa_escala_lrm$Phi, "fa_escala_lrm_corr.csv")

```

## Anexo 4: Sintaxis R estimación de tipologías

```
library(poLCA)
library(mice)
library(car)
library(gmodels)
library(dplyr)

getwd()
#COLOCAR RUTA DONDE SE ENCUENTRA LA BASE DE DATOS EN FORMATO CSV endo_2016_seleccion.csv
#setwd("/Users/Tipologia docentes/Bases")

#PREVIAMENTE SE QUITARON ACENTOS COMAS Y PUNTOS DE CODIGOS Y SE CONVIRTIO MISSING CASES A
999
endo_2016<-read.csv("endo_2016_seleccion.csv")
str(endo_2016)

summary(endo_2016)

endo_2016_sin_cuest<-endo_2016[2:60]
endo_2016_cuest<-endo_2016[1]

#CONVERTIR CODIGOS 999 A MISSING FORMATO R
endo_2016_sin_cuest[endo_2016_sin_cuest==999] <- NA

endo_2016<-cbind(endo_2016_cuest,endo_2016_sin_cuest)

#QUITAR LOS NIVELES NO UTILIZADOS
endo_2016<-droplevels(endo_2016)

#ORDENAR LOS NIVELES DE FACTORES
endo_2016$p805_1 <- factor(endo_2016$p805_1, ordered=TRUE, levels = c("Nada", "Poco",
"Regular", "Bastante"))
endo_2016$p805_2 <- factor(endo_2016$p805_2, ordered=TRUE, levels = c("Nada", "Poco",
"Regular", "Bastante"))
endo_2016$p805_3 <- factor(endo_2016$p805_3, ordered=TRUE, levels = c("Nada", "Poco",
"Regular", "Bastante"))
endo_2016$p805_4 <- factor(endo_2016$p805_4, ordered=TRUE, levels = c("Nada", "Poco",
"Regular", "Bastante"))
endo_2016$p805_5 <- factor(endo_2016$p805_5, ordered=TRUE, levels = c("Nada", "Poco",
"Regular", "Bastante"))
endo_2016$p805_6 <- factor(endo_2016$p805_6, ordered=TRUE, levels = c("Nada", "Poco",
"Regular", "Bastante"))
endo_2016$p805_7 <- factor(endo_2016$p805_7, ordered=TRUE, levels = c("Nada", "Poco",
"Regular", "Bastante"))

endo_2016$p818_1 <- factor(endo_2016$p818_1, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p818_2 <- factor(endo_2016$p818_2, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p818_3 <- factor(endo_2016$p818_3, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p818_4 <- factor(endo_2016$p818_4, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p818_5 <- factor(endo_2016$p818_5, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p818_6 <- factor(endo_2016$p818_6, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p818_7 <- factor(endo_2016$p818_7, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p818_8 <- factor(endo_2016$p818_8, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p818_9 <- factor(endo_2016$p818_9, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))

endo_2016$p819_1 <- factor(endo_2016$p819_1, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p819_2 <- factor(endo_2016$p819_2, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p819_3 <- factor(endo_2016$p819_3, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
```

```

endo_2016$p819_4 <- factor(endo_2016$p819_4, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p819_5 <- factor(endo_2016$p819_5, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p819_6 <- factor(endo_2016$p819_6, ordered=TRUE, levels = c("I.I.E.E Unidocente
/Director","Nada satisfecho", "Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p819_7 <- factor(endo_2016$p819_7, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p819_8 <- factor(endo_2016$p819_8, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p819_9 <- factor(endo_2016$p819_9, ordered=TRUE, levels = c("Nada satisfecho",
"Poco satisfecho", "Satisfecho", "Muy satisfecho"))
endo_2016$p819_10 <- factor(endo_2016$p819_10, ordered=TRUE, levels = c("Nada
satisfecho", "Poco satisfecho", "Satisfecho", "Muy satisfecho"))

endo_2016$p821 <- factor(endo_2016$p821, ordered=TRUE, levels = c("Peores", "Iguales",
"Mejores"))
endo_2016$p822 <- factor(endo_2016$p822, ordered=TRUE, levels = c("Peores", "Iguales",
"Mejores"))
endo_2016$p823 <- factor(endo_2016$p823, ordered=TRUE, levels = c("Peores", "Iguales",
"Mejores"))

endo_2016$edad <- factor(endo_2016$edad, ordered=TRUE, levels = c("Hasta 35 anos", "36 a
45 anos", "46 a 55 anos","55 anos a mas"))

endo_2016$NSE <- factor(endo_2016$NSE, ordered=TRUE, levels = c("NSE E", "NSE D", "NSE
C","NSE B", "NSE A"))

endo_2016$p801 <- factor(endo_2016$p801, ordered=TRUE, levels = c("No me gustaria", "Me
daria lo mismo", "Si me gustaria"))

endo_2016$p812_1_r <- factor(endo_2016$p812_1_r, ordered=TRUE, levels = c("Sector
privado", "No conoce la ley", "Totalmente en desacuerdo","En desacuerdo","De
acuerdo","Totalmente de acuerdo"))
endo_2016$p812_2_r <- factor(endo_2016$p812_2_r, ordered=TRUE, levels = c("Sector
privado", "No conoce la ley", "Totalmente en desacuerdo","En desacuerdo","De
acuerdo","Totalmente de acuerdo"))
endo_2016$p812_3_r <- factor(endo_2016$p812_3_r, ordered=TRUE, levels = c("Sector
privado", "No conoce la ley", "Totalmente en desacuerdo","En desacuerdo","De
acuerdo","Totalmente de acuerdo"))
endo_2016$p812_4_r <- factor(endo_2016$p812_4_r, ordered=TRUE, levels = c("Sector
privado", "No conoce la ley", "Totalmente en desacuerdo","En desacuerdo","De
acuerdo","Totalmente de acuerdo"))
endo_2016$p812_5_r <- factor(endo_2016$p812_5_r, ordered=TRUE, levels = c("Sector
privado", "No conoce la ley", "Totalmente en desacuerdo","En desacuerdo","De
acuerdo","Totalmente de acuerdo"))
endo_2016$p812_6_r <- factor(endo_2016$p812_6_r, ordered=TRUE, levels = c("Sector
privado", "No conoce la ley", "Totalmente en desacuerdo","En desacuerdo","De
acuerdo","Totalmente de acuerdo"))
endo_2016$p812_7_r <- factor(endo_2016$p812_7_r, ordered=TRUE, levels = c("Sector
privado", "No conoce la ley", "Totalmente en desacuerdo","En desacuerdo","De
acuerdo","Totalmente de acuerdo"))

#RECODIFICAR NSE UNIENDO AB
attach(endo_2016)
endo_2016$NSE_2[NSE == "NSE E"] <- "NSE E"
endo_2016$NSE_2[NSE == "NSE D"] <- "NSE D"
endo_2016$NSE_2[NSE == "NSE C"] <- "NSE C"
endo_2016$NSE_2[NSE == "NSE B"] <- "NSE AB"
endo_2016$NSE_2[NSE == "NSE A"] <- "NSE AB"
detach(endo_2016)

endo_2016$NSE_2<-as.ordered(endo_2016$NSE_2)
endo_2016$NSE_2 <- factor(endo_2016$NSE_2, levels=rev(levels(endo_2016$NSE_2)))

NSE_2<-endo_2016$NSE_2

#AÑADIR VARIABLE ESTIMADA DE PRESENCIA DE MENORES DE 12 AÑOS EN EL HOGAR
#COLOCAR RUTA DONDE SE ENCUENTRA LA BASE DE DATOS EN FORMATO CSV endo2016_menores.csv
#setwd("/Users/Tipologia docentes/Bases")

```

```

endo_2016_menores<-read.csv("endo2016_menores.csv")

endo_2016 <- endo_2016[order(endo_2016$cuestionario),]
endo_2016_menores <- endo_2016_menores[order(endo_2016_menores$cuestionario),]
endo_2016 <- merge(endo_2016,endo_2016_menores,by="cuestionario")

menor<-endo_2016$menor
cuestionario<-endo_2016$cuestionario

Sys.setlocale("LC_ALL", "C")

#EN ANALISIS FACTORIAL PREVIO SE ELIGIERON LOS ITEMS DE CONFIANZA EN INSTITUCIONES
# p805_1 MINEDU
# p805_4 UGEL
# p805_7 SUTEP

#EN ANALISIS FACTORIAL PREVIO SE ELIGIERON LOS ITEMS DE SATISFACCION CON DOCENCIA
# p819_2 LOGROS DE ESTUDIANTES
# p819_7 PADRES DE FAMILIA
# p819_9 SALARIO

#EN ANALISIS FACTORIAL PREVIO SE ELIGIERON LOS ITEMS DE SATISFACCION CON VIDA
# p818_1 VIDA
# p818_5 CONDICIONES DE JUBILACION
# p818_8 RELACIONES FAMILIARES

#EN ANALISIS FACTORIAL PREVIO SE ELIGIERON LOS ITEMS DE OPINION LRM
# p812_1_r ESCALAS MAGISTERIALES
# p812_6_r ACCESO A CARGOS

#POR SOLICITUD DE MINEDU SE AÑADIERON
# p818_2 Satisfacción con salud
# p818_3 Satisfacción con trabajo
# p819_4 Satisfacción reconocimiento Minedu

##### MODELOS BASADOS EN VOCACION Y MIRADA DE FUTURO #####

#####IMPUTAR VALORES PERDIDOS
imp <- mice(endo_2016[c("NSE", "p04",
"p07", "p09", "p109_c_01", "p301", "p321", "p323", "p325",
"p801", "p821", "p822", "p823", "p306_r", "edad",
"p811_r", "p812_1_r", "p812_6_r")], seed = 100, m=10)

#COMPLETAR VALORES PERDIDOS CON IMPUTACIONES
endo_2016_imp <- complete(imp)
str(endo_2016_imp)

endo_2016_imp <-cbind(endo_2016_imp,NSE_2,menor,cuestionario)
endo_2016_imp$menor<-as.factor(endo_2016_imp$menor)

#GRABAR DATOS IMPUTADOS
write.csv(endo_2016_imp,"endo_2016_imputado.csv")

##### HISTORIAL DE MODELOS ANALIZADOS

#PRIMER MODELO - 5 SEGMENTOS

modelo1 <- cbind(NSE_2, p04, p07,p09,p109_c_01,p301,p321,p323,
p801,p821,p822,p823,p306_r,edad,p811_r,p812_1_r,p812_6_r)~1

cluster1 <- poLCA(modelo1,endo_2016_imp,nclass=5,nrep=10)

#SEGUNDO MODELO - 6 SEGMENTOS
cluster2 <- poLCA(modelo1,endo_2016_imp,nclass=6,nrep=10)

table <- table(cluster1$predclass,cluster2$predclass)
prop.table(table, 2)%>% round(2)

#TERCER MODELO - 5 SEGMENTOS - SE RETIRAN P812_1 Y P812_6 PORQUE CONVERGEN EN 1 CLUSTER
#SE AGREGO VARIABLE SI TIENE HIJOS MENORES DE 12 ANOS EN HOGAR
modelo3 <- cbind(NSE_2, p04, p07,p09,p109_c_01,p301,p321,p323,

```

```

p801,p821,p822,p823,p306_r,edad,p811_r,menor)~1

cluster3 <- poLCA(modelo3,endo_2016_imp,nclass=5,nrep=10)

#CUARTO MODELO - 6 SEGMENTOS - SE RETIRAN P812_1 Y P812_6 PORQUE CONVERGEN EN 1 CLUSTER
#SE AGREGO VARIABLE SI TIENE HIJOS MENORES DE 12 ANOS EN HOGAR
#SE BUSCO SEPARAR SEGMENTO COLEGIO PRIVADO

cluster4 <- poLCA(modelo3,endo_2016_imp,nclass=6,nrep=10)

#QUINTO MODELO - 7 SEGMENTOS - SE RETIRAN P812_1 Y P812_6 PORQUE CONVERGEN EN 1 CLUSTER
#SE AGREGO VARIABLE SI TIENE HIJOS MENORES DE 12 ANOS EN HOGAR
#SE BUSCO SEPARAR AUN MAS DOCENTES DE COLEGIO PRIVADO
cluster5 <- poLCA(modelo3,endo_2016_imp,nclass=7,nrep=10)

#SEXTO MODELO - 6 SEGMENTOS - SE RETIRAN P812_1 Y P812_6 PORQUE CONVERGEN EN 1 CLUSTER
#SE RETIRA VARIABLE SI TIENE HIJOS MENORES DE 12 ANOS EN HOGAR PORQUE NO AYUDA A SEPARAR
modelo6 <- cbind(NSE_2, p04, p07,p09,p109_c_01,p301,p321,p323,
                p801,p821,p822,p823,p306_r,edad,p811_r)~1

cluster6 <- poLCA(modelo6,endo_2016_imp,nclass=6,nrep=10)

#SEPTIMO MODELO - 7 SEGMENTOS - SE RETIRAN P812_1 Y P812_6 PORQUE CONVERGEN EN 1 CLUSTER
#SE RETIRA VARIABLE SI TIENE HIJOS MENORES DE 12 ANOS EN HOGAR PORQUE NO AYUDA A SEPARAR
modelo6 <- cbind(NSE_2, p04, p07,p09,p109_c_01,p301,p321,p323,
                p801,p821,p822,p823,p306_r,edad,p811_r)~1

cluster7 <- poLCA(modelo6,endo_2016_imp,nclass=7,nrep=10)

## MODELOS ANTERIORES NO SE LOGRA SEPARAR DOCENTES DE IIEE PRIVADOS
## SE DECIDE SEPARAR LOS ANALISIS POR IIEE PUBLICOS Y PRIVADOS

#FILTRAR BASE SOLO PUBLICA
endo_2016_imp_pub <- endo_2016_imp[ which(endo_2016_imp$p07=='Publica'), ]
summary(endo_2016_imp_pub)
endo_2016_imp_pub<-droplevels(endo_2016_imp_pub)

#OCTAVO MODELO - 5 SEGMENTOS SOLO PUBLICAS
cluster8 <- poLCA(modelo6,endo_2016_imp_pub,nclass=5,nrep=10)

#NOVENO MODELO - 6 SEGMENTOS SOLO PUBLICAS
#SEPARA MEJOR CAUSAS DE INSATISFACCION DOCENTE
cluster9 <- poLCA(modelo6,endo_2016_imp_pub,nclass=6,nrep=5)

#DECIMO MODELO - 6 SEGMENTOS SOLO PUBLICAS
#SE AÑADIERON 2 VARIABLES: P325 EXPECTATIVAS A FUTURO Y P607 FORMA DE CAPACITACION
modelo7 <- cbind(NSE_2, p04, p07,p09,p109_c_01,p301,p321,p323,
                p801,p821,p822,p823,p306_r,edad,p811_r,
                p325,p607)~1

cluster10 <- poLCA(modelo7,endo_2016_imp_pub,nclass=6,nrep=5)

#UNDECIMO MODELO - 6 SEGMENTOS SOLO PUBLICAS
#SE RETIRA P607 FORMA DE CAPACITACION NO CONVERGE
#####
#          MODELO FINAL          #
#          IIEE PUBLICAS          #
#          6 SEGMENTOS           #
#####

mod_pub_final <- cbind(NSE_2, p04, p07,p09,p109_c_01,p301,p321,p323,
                    p801,p821,p822,p823,p306_r,edad,p811_r,
                    p325)~1

cluster_pub_final<- poLCA(mod_pub_final,endo_2016_imp_pub,nclass=6,nrep=10,maxiter =
5000)

tipologia_pub<-cluster_pub_final$predclass

#DATA UNIDA CON CLUSTERS
endo_2016_pub <- endo_2016[ which(endo_2016$p07=='Publica'), ]
endo_2016_pub_clusters <- cbind(tipologia_pub,endo_2016_pub)

```

```

str(endo_2016_pub_clusters)

## MODELOS PARA IIEE PRIVADAS

#SELECCION DE IIEE PRIVADAS
endo_2016_imp_pri <- endo_2016_imp[ which(endo_2016_imp$p07=='Privada'), ]
summary(endo_2016_imp_pri)
endo_2016_imp_pri<-droplevels(endo_2016_imp_pri)

#MODELO SOLO PRIVADAS - 3 SEGMENTOS
#SE RETIRAN P811_r p306_r p07 P04 PORQUE NO APLICAN
#SE RETIRA p109_c_01 EDAD PORQUE NO AYUDA A SEPARAR
#SE RETIRA VARIABLE SI TIENE HIJOS MENORES DE 12 ANOS EN HOGAR PORQUE NO AYUDA A SEPARAR
modelo8 <- cbind(NSE_2,p09,p301,p321,p323,
                p801,p821,p822,p823)~1

cluster1_pri <- poLCA(modelo8,endo_2016_imp_pri,nclass=3,maxiter=5000, nrep=10)

#####
#          MODELO FINAL          #
#          IIEE PRIVADAS          #
#          4 SEGMENTOS          #
#####
#MODELO SOLO PRIVADAS - 4 SEGMENTOS
#SE RETIRAN P811_r p306_r p07 P04 PORQUE NO APLICAN
#SE RETIRA EDAD PORQUE NO AYUDA A SEPARAR
#SE RETIRA VARIABLE SI TIENE HIJOS MENORES DE 12 ANOS EN HOGAR PORQUE NO AYUDA A SEPARAR
#SE AÑADE P325 y p109_c_01
mod_pri_final <- cbind(NSE_2,p09,p109_c_01,p301,p321,p323,
                    p801,p821,p822,p823,
                    p325)~1

cluster_pri_final <- poLCA(mod_pri_final,endo_2016_imp_pri,nclass=4,maxiter=5000,
nrep=10)

tipologia_pri<-cluster_pri_final$predclass

#DATA UNIDA CON CLUSTERS
endo_2016_pri <- endo_2016[ which(endo_2016$p07=='Privada'), ]
endo_2016_pri_clusters <- cbind(tipologia_pri,endo_2016_pri)

str(endo_2016_pri_clusters)

#GUARDAR ARCHIVOS CON CLUSTERS FINALES
write.csv(endo_2016_pub_clusters,"endo_2016_pub_clusters_23102018.csv")
write.csv(endo_2016_pri_clusters,"endo_2016_pri_clusters_23102018.csv")

#####
#####
# MODELOS DESCARTADOS LUEGO DE COMPARARLOS CON MODELOS ANTERIORES

#IMPUTAR VALORES PERDIDOS
imp_2 <- mice(endo_2016[c("NSE", "p04", "p07", "p09", "p109_c_01", "edad", "p306_r",
                        "p805_1", "p805_4", "p805_7", "p818_1", "p818_5", "p818_8",
                        "p819_2", "p819_7", "p819_9",
                        "p818_2", "p818_3", "p819_4")], seed = 100, m=10)

#COMPLETAR VALORES PERDIDOS CON IMPUTACIONES
endo_2016_imp_2 <- complete(imp_2)

p301 <-endo_2016_imp$p301
p325 <-endo_2016_imp$p325

```

```

summary(endo_2016_imp_2)

endo_2016_imp_2 <- cbind(endo_2016_imp_2, NSE_2, menor, cuestionario, p301, p325)
endo_2016_imp_2$menor <- as.factor(endo_2016_imp_2$menor)

write.csv(endo_2016_imp_2, "endo_2016_imputado_2.csv")

#PRIMER MODELO - 5 SEGMENTOS
modelo_a <- cbind(NSE_2, p04, p07, p09, p109_c_01, edad, p306_r,
                p805_1, p805_4, p805_7, p818_1, p818_5, p818_8,
                p819_2, p819_7, p819_9,
                p818_2, p818_3, p819_4)~1

cluster_a <- polCA(modelo_a, endo_2016_imp_2, nclass=5, nrep=10)
cluster_a

#SEGUNDO MODELO - 6 SEGMENTOS
cluster_a2 <- polCA(modelo_a, endo_2016_imp_2, nclass=6, nrep=10)
cluster_a2

#TERCER MODELO - 6 SEGMENTOS CON P325 Y P301
modelo_b <- cbind(NSE_2, p04, p07, p09, p109_c_01, edad, p306_r,
                p805_1, p805_4, p805_7, p818_1, p818_5, p818_8,
                p819_2, p819_7, p819_9,
                p818_2, p818_3, p819_4, p325, p301)~1

cluster_a3 <- polCA(modelo_b, endo_2016_imp_2, nclass=6, nrep=10)

#####TAMBIEN SE TUVO QUE SEPARAR IIEE PUBLICAS DE IIEE PRIVADAS

#####SOLO PUBLICAS

endo_2016_imp2_pub <- endo_2016_imp_2[ which(endo_2016_imp_2$p07=='Publica'), ]
summary(endo_2016_imp2_pub)

#SOLO PUBLICAS - 5 SEGMENTOS CON P325 Y P301
modelo_b <- cbind(NSE_2, p04, p07, p09, p109_c_01, edad, p306_r,
                p805_1, p805_4, p805_7, p818_1, p818_5, p818_8,
                p819_2, p819_7, p819_9,
                p818_2, p818_3, p819_4, p325, p301)~1

cluster_a4 <- polCA(modelo_b, endo_2016_imp2_pub, nclass=5, nrep=10)

#SOLO PUBLICAS - 6 SEGMENTOS CON P325 Y P301
cluster_a5 <- polCA(modelo_b, endo_2016_imp2_pub, nclass=6, nrep=10)

#####SOLO PRIVADAS

endo_2016_imp2_pri <- endo_2016_imp_2[ which(endo_2016_imp_2$p07=='Privada'), ]
str(endo_2016_imp2_pri)
endo_2016_imp2_pri <- droplevels(endo_2016_imp2_pri)

#SOLO PRIVADAS - 4 SEGMENTOS CON P325 Y P301
#SE RETIRAN p07 P04 p306_r PORQUE NO APLICAN
#SE AÑADE P325 y p109_c_01
modelo_c <- cbind(NSE_2, p09, p109_c_01, edad,
                p805_1, p805_4, p805_7, p818_1, p818_5, p818_8,
                p819_2, p819_7, p819_9,
                p818_2, p818_3, p819_4, p325, p301)~1

cluster_a_pri <- polCA(modelo_c, endo_2016_imp2_pri, nclass=4, nrep=10)

#SOLO PRIVADAS - 3 SEGMENTOS CON P325 Y P301
cluster_a2_pri <- polCA(modelo_c, endo_2016_imp2_pri, nclass=3, nrep=10)

```

## Anexo 5: Sintaxis R para clasificación de nuevos casos

```
#####  
# CODIGO PARA CLASIFICAR NUEVOS CASOS DOCENTES IIEE PUBLICAS #  
#####  
  
#LEER BASE DE DATOS CON CASOS IMPUTADOS  
#setwd("/Users/Tipologia docentes/Bases")  
endo_2016_clasif<-read.csv("endo_2016_imputado.csv")  
str(endo_2016_clasif)  
  
#FILTRAR BASE SOLO PUBLICA  
endo_2016_clasif_pub <- endo_2016_clasif[ which(endo_2016_clasif$p07=='Publica'), ]  
summary(endo_2016_clasif_pub)  
  
endo_2016_clasif_pub<-droplevels(endo_2016_clasif_pub)  
  
modelo_clasif_publicas <- cbind(NSE_2, p04, p07,p09,p109_c_01,p301,p321,p323,  
p801,p821,p822,p823,p306_r,edad,p811_r,  
p325)~1  
  
cluster_clasif_publicas <-  
poLCA(modelo_clasif_publicas,endo_2016_clasif_pub,nclass=6,nrep=10,maxiter = 5000)  
cluster_clasif_publicas  
  
# REVISAR LOS % DE CADA SEGMENTO PARA SABER EL ORDEN EN QUE SE ESTIMAN EN ESTA OCASION  
#  
# Predicted class memberships (by modal posterior prob.)  
#  
# 0.0982 Jovenes críticos  
# 0.1089 En riesgo  
# 0.2804 Nombrada optimista  
# 0.2285 Contratados aliados  
# 0.1625 Nombrado aspirante a cargos  
# 0.1215 Nombrados afectados  
  
# GENERAR BASE DE DATOS CON RESULTADOS DE LAS PREGUNTAS SIGUIENDO ESTOS CODIGOS DE  
# RESPUESTA  
# RESPETAR EL ORDEN DE LAS PREGUNTAS ESTRICTAMENTE  
# NSE_2  
# 1 NSE AB  
# 2 NSE C  
# 3 NSE D  
# 4 NSE E  
#  
# p04  
# 1 Rural  
# 2 Urbana  
#  
# p09  
# 1 Inicial  
# 2 Primaria  
# 3 Secundaria  
#  
# p109_c_01  
# 1 Hombre  
# 2 Mujer  
#  
# p301  
# 1 Carrera que ofrecia estabilidad seguridad laboral  
# 2 Economicamente no podia estudiar otra carrera  
# 3 El buen recuerdo que le dejaron algunos profesores  
# 4 Era la unica opcion de estudios disponible en su zona  
# 5 Le gustaba la idea de trabajar con ninos y adolescentes  
# 6 Tradicion familiar  
# 7 Vocacion  
#  
# p321 Satisfaccion con trabajo actual  
# 1 No  
# 2 Si
```



```

#
# p323 Volveria a ser docente
# 1 No
# 2 Si
#
# p801 Le gustaria que sus hijos sean docentes
# 1 Me daria lo mismo
# 2 No me gustaria
# 3 Si me gustaria
#
# p821 Logro aprendizaje IIEE públicas
# 1 Iguales
# 2 Mejores
# 3 Peores
#
# p822 Mejoria de IIEE públicas
# 1 Iguales
# 2 Mejores
# 3 Peores
#
# p823 Futuro de IIEE públicas
# 1 Iguales
# 2 Mejores
# 3 Peores
#
# p306_r
# 1 Contratado o sin contrato
# 2 Nombrado
#
# edad
# 1 36 a 45 anos
# 2 46 a 55 anos
# 3 55 anos a mas
# 4 Hasta 35 anos
#
# p811_r
# 1 Lo beneficia
# 2 Lo perjudica
# 3 Ni lo beneficia ni perjudica
# 4 No conoce la ley
#
# p325
# 1 Dejar la carrera docente y dedicarse a otra actividad
# 2 Ocupar cargos de direccion y gestion en una escuela
# 3 Seguir como docente de aula en esta escuela
# 4 Ser acompanante o formador de docentes
# 5 Ser director de UGEL
# 6 Ser especialista en el Ministerio de Educacion DRE o UGEL
# 7 Trabajar como docente de aula en otra escuela

#EJEMPLO DE BASE CON RESPUESTAS PARA CLASIFICAR A NUEVOS DOCENTES
ejemplo_clasificacion<-read.csv("ejemplo_clasificacion.csv")

#SE RETIRA LA PRIMERA COLUMNA DE ENCUESTA
ejemplo_clasificacion_encuesta<-ejemplo_clasificacion[1]
ejemplo_clasificacion <- ejemplo_clasificacion[-1]
ejemplo_clasificacion

#SE GENERA CLASIFICACION CON MODELO IIEE PUBLICAS
clasificacion <- polCA.posterior(lc=cluster_clasif_publicas,y=ejemplo_clasificacion)%>%
round(1)

#SE GUARDAN LAS PROBABILIDADES DE PERTENENCIA A CADA CLUSTER
write.csv(cbind(ejemplo_clasificacion_encuesta,clasificacion),"ejemplo_datos_clasificados
.csv")

#####
# CODIGO PARA CLASIFICAR NUEVOS CASOS DOCENTES IIEE PRIVADAS #
#####

#LEER BASE DE DATOS CON CASOS IMPUTADOS

```

```

#setwd("/Users/Tipologia docentes/Bases")
endo_2016_clasif<-read.csv("endo_2016_imputado.csv")
str(endo_2016_clasif)

#FILTRAR BASE SOLO PRIVADAS
endo_2016_clasif_pri <- endo_2016_clasif[ which(endo_2016_clasif$p07=='Privada'), ]
summary(endo_2016_clasif_pri)

endo_2016_clasif_pri<-droplevels(endo_2016_clasif_pri)

modelo_clasif_privadas <- cbind(NSE_2,p09,p109_c_01,p301,p321,p323,
                                p801,p821,p822,p823,
                                p325)~1

cluster_clasif_privadas <-
poLCA(modelo_clasif_privadas,endo_2016_clasif_pri,nclass=4,maxiter=5000, nrep=10)
cluster_clasif_privadas

# REVISAR LOS % DE CADA SEGMENTO PARA SABER EL ORDEN EN QUE SE ESTIMAN EN ESTA OCASION
#
# Predicted class memberships (by modal posterior prob.)
#
# 0.0881 En riesgo
# 0.2255 Críticas
# 0.242 Satisfechos
# 0.4444 Aliadas

# GENERAR BASE DE DATOS CON RESULTADOS DE LAS PREGUNTAS SIGUIENDO ESTOS CODIGOS DE
# RESPUESTA
# RESPETAR EL ORDEN DE LAS PREGUNTAS ESTRICTAMENTE
# NSE_2
# 1 NSE AB
# 2 NSE C
# 3 NSE D
# 4 NSE E
#
# p09
# 1 Inicial
# 2 Primaria
# 3 Secundaria
#
# p109_c_01
# 1 Hombre
# 2 Mujer
#
# p301
# 1 Carrera que ofrecia estabilidad seguridad laboral
# 2 Economicamente no podia estudiar otra carrera
# 3 El buen recuerdo que le dejaron algunos profesores
# 4 Era la unica opcion de estudios disponible en su zona
# 5 Le gustaba la idea de trabajar con ninos y adolescentes
# 6 Tradicion familiar
# 7 Vocacion
#
# p321 Satisfaccion con trabajo actual
# 1 No
# 2 Si
#
# p323 Volveria a ser docente
# 1 No
# 2 Si
#
# p801 Le gustaria que sus hijos sean docentes
# 1 Me daria lo mismo
# 2 No me gustaria
# 3 Si me gustaria
#
# p821 Logro aprendizaje IIEE públicas
# 1 Iguales
# 2 Mejores
# 3 Peores
#

```

```

# p822 Mejoria de IIEE públicas
# 1 Iguales
# 2 Mejores
# 3 Peores
#
# p823 Futuro de IIEE públicas
# 1 Iguales
# 2 Mejores
# 3 Peores
#
# p325
# 1 Dejar la carrera docente y dedicarse a otra actividad
# 2 Ocupar cargos de direccion y gestion en una escuela
# 3 Seguir como docente de aula en esta escuela
# 4 Ser acompanante o formador de docentes
# 5 Ser director de UGEL
# 6 Ser especialista en el Ministerio de Educacion DRE o UGEL
# 7 Trabajar como docente de aula en otra escuela

#EJEMPLO DE BASE CON RESPUESTAS PARA CLASIFICAR A NUEVOS DOCENTES
ejemplo_clasificacion_privadas<-read.csv("ejemplo_clasificacion_privadas.csv")

#SE RETIRA LA PRIMERA COLUMNA DE ENCUESTA
ejemplo_clasificacion_privadas_encuesta<-ejemplo_clasificacion_privadas[1]
ejemplo_clasificacion_privadas <- ejemplo_clasificacion_privadas[-1]
ejemplo_clasificacion_privadas

#SE GENERA CLASIFICACION CON MODELO IIEE PUBLICAS
clasificacion_privadas <-
poLCA.posterior(lc=cluster_clasif_privadas,y=ejemplo_clasificacion_privadas)%>% round(1)

#SE GUARDAN LAS PROBABILIDADES DE PERTENENCIA A CADA CLUSTER
write.csv(cbind(ejemplo_clasificacion_privadas_encuesta,clasificacion_privadas),"ejemplo_
datos_clasificados_privadas.csv")

```

**Anexo 6:**  
Presentación y análisis de resultados (formato .ppt)