

**INDICE INTEGRAL DE EVALUACIÓN DEL DESEMPEÑO DE
PROVEEDORES EXTERNOS A UNA INSTITUCIÓN BANCARIA**

APLICACIÓN DE TÉCNICAS MULTIVARIANTES



TRABAJO DE FIN DE MÁSTER

María del Rocío Mora Fallas

Tutor: Ramón Gutiérrez Sánchez

MÁSTER EN ESTADÍSTICA APLICADA

Enero 2023

RESUMEN

El análisis multivariante es una técnica estadística que permite dar solución a problemas de distintas disciplinas. En el presente estudio se aplicó las técnicas de análisis factorial exploratorio y confirmatorio para la construcción de indicadores que permitieran evaluar a proveedores externos de una entidad bancaria. También se aplicó la técnica de análisis de conglomerados para crear perfiles de dichos proveedores.

Se obtuvieron 6 variables latentes ubicadas en dos dimensiones que permitieron la creación de un score único para otorgar una calificación de desempeño global al proveedor. También se pudo crear tres perfiles de proveedores los cuales fueron asociados a una escala empresarial estandarizada internacionalmente, que permitió a la entidad bancaria poder establecer tipo de proveedores contratados en su empresa.

INDICE GENERAL

1. INDICE

| | |
|--|----|
| 1. Introducción..... | 1 |
| 2. Antecedentes..... | 2 |
| 2.1 Evaluación de proveedores..... | 2 |
| 2.1.1 Clasificación de proveedores externos..... | 2 |
| 2.1.2 Técnicas y métodos usados | 3 |
| 2.2 Situación en la institución de estudio..... | 5 |
| 2.2.1 Población de estudio | 5 |
| 3. Objetivos | 6 |
| 3.1 Objetivo General..... | 6 |
| 3.2 Objetivos específicos | 6 |
| 4. Marco teórico..... | 7 |
| 4.1 Proveedores | 7 |
| 4.2 Construcción de índices | 10 |
| 4.3 Técnicas multivariantes..... | 11 |
| 4.3.1 Análisis factorial..... | 11 |
| 4.3.2 Análisis de conglomerados..... | 18 |
| 5. Metodología..... | 20 |
| 5.1 Método científico de investigación..... | 20 |
| 5.2 Definición de variables..... | 21 |
| 6. Aplicación de escalas construidas..... | 24 |
| 6.1 Constructos | 24 |

| | | |
|-------|--|----|
| 6.1.1 | Confianza..... | 24 |
| 6.1.2 | Responsabilidad Legal | 27 |
| 6.1.3 | Responsabilidad Social | 27 |
| 6.1.4 | Seguridad Económica | 27 |
| 7. | Aplicación de las Técnicas multivariantes | 28 |
| 7.1 | Análisis factorial exploratorio | 28 |
| 7.2 | Análisis Factorial Confirmatorio (AFC) | 36 |
| 7.2.1 | Análisis iniciales | 36 |
| 7.2.2 | Análisis Factorial confirmatorio 1: Primera versión del análisis con todas las variables38 | |
| 7.2.3 | Análisis factorial confirmatorio 2: mejora del modelo obtenido en AFC 1..... | 43 |
| 7.3 | Análisis de conglomerados..... | 46 |
| 8. | Resultados..... | 51 |
| 9. | Bibliografía..... | 55 |
| 10. | Anexos | 59 |
| 10.1 | ANEXO 1: Recategorización de variables..... | 59 |
| 10.2 | ANEXO 2: Análisis factorial exploratorio realizado en R..... | 61 |
| 10.3 | ANEXO 3: Análisis factorial confirmatorio realizado en R | 72 |
| 10.4 | ANEXO 3: Análisis de conglomerados realizado en R | 86 |

INDICE DE TABLAS

| | |
|---|----|
| TABLA 1 HIPÓTESIS PROBABILÍSTICAS EN EL ANÁLISIS FACTORIAL | 12 |
| TABLA 2: DEFINICIÓN DE VARIABLES UTILIZADAS EN EL ESTUDIO | 22 |
| TABLA 3 <i>CONSTRUCTO CONFIANZA: DESCRIPCIÓN DE VARIABLES UNITARIAS UTILIZADAS</i> | 24 |
| TABLA 4 CONFORMACIÓN DE GRUPOS PARA LA CONSTRUCCIÓN DE LA VARIABLE LATENTE CONFIANZA | 26 |
| TABLA 5 <i>MEDIDAS DESCRIPTIVAS PARA VARIABLES UTILIZADAS EN ANÁLISIS FACTORIAL EXPLORATORIO</i> | 28 |
| TABLA 6 <i>ESTRUCTURA DE INDICADORES PARA EL ANÁLISIS FACTORIAL CONFIRMATORIO (AFC)</i> | 37 |
| TABLA 7 <i>MEDIDAS DE AJUSTE PARA UN AFC</i> | 39 |
| TABLA 8 <i>PESOS PARA CADA CONSTRUCTO EN LA CONFORMACIÓN DEL SCORE FINAL DEL INDICADOR DE EVALUACIÓN A PROVEEDORES</i> <i>BANCARIOS</i> | 45 |
| TABLA 9 DESCRIPCIÓN FINAL DE VARIABLES LATENTES CONSTRUIDAS CON TÉCNICAS MULTIVARIANTES..... | 51 |
| TABLA 10 ASOCIACIÓN DE CLÚSTER CON NIVELES DE ESCALA BARRET | 54 |

INDICE DE FIGURAS

| | |
|--|----|
| FIGURA 1 HIPÓTESIS TEÓRICA CONCEPTUAL DE LA CONFORMACIÓN DE VARIABLES LATENTES | 9 |
| FIGURA 2 <i>CONCEPTUALIZACIÓN Y EJEMPLO DE LA CONSTRUCCIÓN DE ÍNDICES</i> | 10 |
| FIGURA 3 ESQUEMA CONCEPTUAL DE UN ANÁLISIS FACTORIAL | 12 |
| FIGURA 4 PROPUESTA METODOLÓGICA PARA REALIZAR UN AFE..... | 16 |
| FIGURA 5 PROPUESTA METODOLÓGICA PARA REALIZAR UN AFC..... | 17 |
| FIGURA 6 IDEA INTUITIVA DE FORMACIÓN DE CLÚSTER POR K-MEDIAS | 18 |
| FIGURA 7 ADAPTACIÓN DE METODOLOGÍA DE CHURCHILL AL CONTEXTO DE LA INVESTIGACIÓN | 21 |
| FIGURA 8 CONSTRUCTO CONFIANZA: ESQUEMA TEÓRICO CONCEPTUAL..... | 25 |
| FIGURA 9 DIAGRAMAS DE CAJA PARA VARIABLES UTILIZADAS EN EL ANÁLISIS FACTORIAL EXPLORATORIO | 29 |
| FIGURA 10 PARÁMETROS DE CALIDAD PARA IDENTIFICAR VIABILIDAD DE AFE CON EL TOTAL DE VARIABLES..... | 30 |
| FIGURA 11 PARÁMETROS DE CALIDAD PARA IDENTIFICAR VIABILIDAD DE AFE EXCLUYENDO VARIABLES N_REP Y REPRES | 31 |
| FIGURA 12 <i>RESULTADOS PARA AFE 1</i> | 32 |
| FIGURA 13 GRÁFICO DE SEDIMENTACIÓN PARA AFE 1..... | 33 |
| FIGURA 14 PARÁMETROS DE CALIDAD PARA IDENTIFICAR VIABILIDAD DE AFE EXCLUYENDO VARIABLES EXP_LAP, ANT_BAN Y CONF..... | 34 |
| FIGURA 15 RESULTADOS PARA AFE 2..... | 35 |
| FIGURA 16 PUNTUACIONES SIN NORMALIZAR Y NORMALIZADAS PARA EL FACTOR OBTENIDO | 35 |
| FIGURA 17 GRÁFICO DE DISTRIBUCIÓN DEL FACTOR "NATURALEZA DE CONTRATO"..... | 36 |
| FIGURA 18 <i>ESPECIFICACIÓN DEL MODELO FACTORIAL CONFIRMATORIO CON EL PAQUETE ESTADÍSTICO "LAVAN"</i> | 38 |
| FIGURA 19 AFC 1: PARTE 1, ESPECIFICACIONES GENERALES DEL MODELO | 39 |
| FIGURA 20 AFC 1: PARTE 2, MEDIDAS DE AJUSTE DEL MODELO..... | 40 |
| FIGURA 21 AFC 1: PARTE 3, PARÁMETROS DEL MODELO | 41 |
| FIGURA 22 AFC 1: GRÁFICO ESTRUCTURAL DEL MODELO..... | 42 |
| FIGURA 23 <i>AFC 2: PARTE 1, ESPECIFICACIONES GENERALES DEL MODELO</i> | 43 |
| FIGURA 24 <i>AFC 2: PARTE 2, MEDIDAS DE AJUSTE DEL MODELO</i> | 43 |
| FIGURA 25 AFC2: VALORES ESTIMADOS Y ESTRUCTURA DE MODELO | 44 |
| FIGURA 26 DISTRIBUCIÓN DE FRECUENCIAS DEL "SCORE" FINAL OBTENIDO PARA EVALUAR DESEMPEÑO DE PROVEEDORES BANCARIOS | 45 |
| FIGURA 27 <i>HISTOGRAMAS DE CONSTRUCTOS PARA LA EVALUACIÓN DEL DESEMPEÑO DE PROVEEDORES</i> | 46 |
| FIGURA 28 <i>TIPOS DE DISTANCIAS APLICADAS EN EL ANÁLISIS DE CONGLOMERADOS</i> | 47 |
| FIGURA 29 MÉTODOS PARA SELECCIONAR TAMAÑOS DE GRUPOS O CLUSTER..... | 48 |
| FIGURA 30 <i>POSIBLES AGRUPACIONES PARA CATEGORIZAR PROVEEDORES</i> | 49 |
| FIGURA 31 VISUALIZACIÓN DE TRES GRUPOS DE PROVEEDORES | 50 |
| FIGURA 32 <i>DESCRIPCIÓN DE CLÚSTER'S</i> | 50 |
| FIGURA 33 <i>IDENTIFICACIÓN DE DIMENSIONES PARA LOS 6 CONSTRUCTOS CREADOS PARA LA EVALUACIÓN DE PROVEEDORES BANCARIOS</i> | 52 |
| FIGURA 34 <i>MODELO DE RICHART BARRET</i> | 53 |
| FIGURA 35 <i>EJEMPLO DE VISUALIZACIÓN DE INFORMACIÓN POWER BI</i> | 54 |

1. Introducción

Las instituciones bancarias en sus múltiples funciones se ven en la necesidad de requerir los servicios de personas físicas o jurídicas que les apoyen en el cumplimiento de servicios o en el abastecimiento de materiales, donde la propia institución no cuenta con el personal o la materia prima requerida que pueda solventar la necesidad que demandan. Estos entes son los llamados proveedores externos, los cuales por lo general son convenidos de forma contractual o por demanda de acuerdo con la función requerida; pueden brindar servicios de diversas índoles, como: alquileres de edificios, abastecimiento de insumos, servicios de mantenimiento de software, capacitaciones, entre otros.

Al tratarse de entidades externas a la institución, esta debe evaluar los posibles riesgos en los que un proveedor podría afectarles, los cuales van desde aspectos reputacionales, interrupción de funciones, violación de privacidad hasta multas o sanciones que impliquen pérdidas económicas (Equipo ORCA, 2021). Esta situación hace visible la necesidad de realizar una evaluación que permita poder mitigar estas posibles afectaciones.

En la actualidad las formas de evaluación a proveedores, independiente al tipo de empresa o institución a la que presten sus servicios, son diversas, como: software, evaluación de expertos, herramientas estadísticas, escalas cualitativas y a su vez estas se derivan en distintos tipos de productos finales, no obstante se pretende en cualquier caso clasificar cuales son los proveedores que podrían tener un más alto grado de incumplimiento de sus funciones hasta los que se podrían considerar una “zona segura o fiable”.

Este estudio visualiza la evaluación de los posibles riesgos de un proveedor desde la perspectiva de la valoración del desempeño, dado que, al realizar este proceso; desde varias aristas: reputacional, confianza, entre otros, se le da de forma intrínseca una visualización a los posibles riesgos que el proveedor podría estar incurriendo y afectar a la institución bancaria. Esta perspectiva, se acuña a los objetivos de la institución, donde dicha valoración, pretende únicamente dar visibilidad a su desempeño y posibles áreas de mejora para el proveedor y no puede implicar ningún aspecto que incurra en despidos o ceses de contratos en el momento actual.

2. Antecedentes

2.1 Evaluación de proveedores

2.1.1 Clasificación de proveedores externos

La contratación de entes externos (proveedores) a una institución supone riesgos de diferentes indoles, ya que “es una decisión que hace parte de la gestión de la cadena de suministro y es considerada como uno de los problemas clave que enfrentan las organizaciones en busca de garantizar el correcto desarrollo de sus operaciones y mantener la competitividad en el mercado” (López et al., 2019, p. 1), por esta razón es clave que la selección de un proveedor independiente del bien o servicio que brinde sea rigurosa y cuidada, para procurar mantener el buen funcionamiento de la institución primaria.

Múltiples investigadores han determinado distintos aspectos que se deben tomar en cuenta en el momento de evaluar el riesgo de un proveedor, ya sea en un proceso de contratación o de su desempeño durante el ciclo de trabajo de este, bajo este contexto, Arroyo y Ramos (2014), expresan que los temas relacionados a costos, son primordiales, seguidos de tópicos subyacentes a confiabilidad y calidad de la entrega, López et al., (2019) apoya estos elementos y además considera otros como: tiempo de entrega, servicio al cliente, capacidad de producción, temas administrativos, flexibilidad operativa, seguridad y reputación. Además, destacan como clave la consideración de que los posibles riesgos tienen dos fuentes, las originadas por el hombre y las causadas por eventos naturales, lo cual aumenta considerablemente el espectro de evaluación de un proveedor en el momento de crear una escala o instrumento para valorar al mismo.

Ortiz et al., (2018) además de coincidir con todos los elementos mencionados anteriormente, los clasifican en dos grandes grupos: de origen cualitativo y de origen cuantitativo, también expresan un resultado de gran importancia como lo es, la cautela respecto a los atributos considerados en la evaluación, ya que se podría enfrentar un conflicto en términos de que un aspecto que refiera satisfacción de uno puede demeritar el cumplimiento de otros aspectos considerados, por lo cual recomienda y resaltan la importancia de realizar ponderaciones en dichos elementos.

2.1.2 Técnicas y métodos usados

Las investigaciones realizadas en el tema arrojan técnicas derivadas de dos grandes metodologías. Por un lado, se tienen los instrumentos que subyacen de estudios de origen cualitativos, por lo general se refieren a estudios de caso o situaciones donde se realiza y los de origen cuantitativo. Para efectos de esta investigación se le presta preferencial atención, a los métodos utilizados bajo la vertiente cuantitativa, donde encontraron resultados desde análisis estadísticos de corte descriptivo hasta herramientas más avanzadas con estadística inferencial, también se encontraron algunos estudios con técnicas matemáticas.

Por la parte de estadística descriptiva, se evidencia el estudio elaborado por Torres et al. (2007) cuya finalidad era realizar una evaluación de los proveedores que brindaban servicio externo en el momento del estudio, a la Universidad de Holguín Cuba. La investigación se realizó bajo una metodología mixta (cualitativo-cuantitativo), donde, en el componente cuantitativo se hizo un análisis estadístico descriptivo sobre una encuesta aplicada a las partes interesadas de la Universidad y que tuviesen el conocimiento necesario de lo indagado; los investigadores plantearon hipótesis no estadísticas, basados en el conocimiento previo del contexto, estas fueron cubiertas a través de preguntas en el cuestionario, y utilizaron el software estadístico SPSS (versión 21), para construir descriptivos que permitieran evidenciar si las hipótesis planteadas al inicio conllevaban a veracidades o no, y así poder apoyar en la toma de decisiones por parte de la Institución involucrada. Lo planteado por Torres et al. deja en evidencia que para realizar estudios de este tipo que conlleven a la creación posterior de instrumentos de evaluación es necesario utilizar técnicas estadísticas inferenciales.

Bajo el contexto de estadística inferencial, se encuentran algunos estudios, donde resalta el realizado por Martín (2013), cuyo objetivo fue realizar un análisis del riesgo de los proveedores de componentes en industria automovilística, para ello empleó un modelo logístico donde utilizó como variable respuesta si la empresa proveedora actualmente se encontraba en un proceso de liquidación o quiebra (empresas fracasadas) o si por el contrario era un empresa emergente o activa de varios años sin problemas financieros (empresas sanas), dando así la respuesta dicotómica 1 y 0, respectivamente. Cabe resaltar que este estudio buscaba indagar que variables de índole económico-financiera influyen en que un proveedor tuviese su empresa en problemas financieros, lo cual es una de las vertientes más importantes de riesgo de incumplimiento hacia otra entidad.

Por otra parte para las variables predictoras, se realizó un estudio paralelo de Análisis por Componentes Principales (PCA) con el software SPSS (versión 20), donde se definieron a priori cuatro componentes: Coeficiente de rentabilidad (conformado para doce variables de origen econométrico), Coeficiente operacional (compuesto por siete variables relacionadas a la importación y exportación de materia prima), Coeficientes de estructura (conformado por cinco variables relacionadas a la liquidez y solvencia del empresa), Coeficientes por empleado (constaba de siete variables en relación a beneficios económicos obtenidos por los empleados de las empresas). Finalmente, al realizar el análisis y tomar como criterios de validez del modelo la medida KMO y la significancia de las variables, y como criterios de selección para “quitar” variables y reducir las dimensiones, se utilizó el valor de las comunalidades y los elementos de la diagonal de la matriz de correlaciones, finalmente el primer componente se redujo a cinco variables, el segundo a una, el tercero se eliminó, y el cuarto quedó con cinco variables. Finalmente, estas dimensiones reducidas se llevaron como variables predictoras al modelo logístico, el cual arrojó un 88,7% de acierto global.

El estudio anterior evidencia dos prácticas muy importantes para investigadores en el campo, las cuales son: el análisis desde la vista del proveedor y el uso de variables poco típicas en este tipo de estudios, y el proporcionar una aplicación de la regresión logística para analizar este tipo de eventos, en lugar de las típicas herramientas de corte cualitativo.

Bajo el mismo tipo de análisis estadístico inferencial realizado por Martín, se encuentra el estudio de Restrepo (2016), cuyo objetivo fue la construcción de un instrumento para evaluar y clasificar proveedores en una empresa ensambladora de motocicletas. Para este caso el investigador utilizó en una primera parte se utilizó medidas de central y dispersión y coeficientes de variación como indicadores de selección de 16 variables disponibles relacionadas al desempeño proveedor, y en una segunda parte la técnica de escalamiento multidimensional, para realizar la clasificación. Finalmente obtuvieron 12 variables indicadoras y cuatro grupos de proveedores, especificados como proveedores con desempeño superior, proveedores nicho, proveedores con buen desempeño general y proveedores con desempeño inferior, siendo este último grupo el de mayor riesgo para las entidades contratantes. Este estudio arroja resultados importantes en a fases de investigación y resultados finales, dado que da grupos de atención prioritaria en torno al riesgo.

2.2 Situación en la institución de estudio

Por temas de confidencialidad, en adelante se hablará de una institución bancaria, que fue el lugar donde realizó el estudio. Esta institución provee a sus usuarios todos los productos bancarios, que va desde sistemas de pensiones hasta tarjetas de crédito y debido. Debido a esto el contrato de proveedores externo es muy variado y depende de las necesidades de cada subgerencia de la entidad.

2.2.1 Población de estudio

Los proveedores que se tomaron en cuenta fueron aquellos que al corte del periodo de estudio mantenían un contrato vigente, prorrogado, finalizado o vencido (en los dos últimos casos con un máximo de dos años de antigüedad).

Se excluyeron notarios o abogados relacionados con procesos de crédito, aquellos inscritos como patrocinios, convenios o suscripciones, esto porque el servicio y condiciones de contratación es muy diferente al resto y en muchas ocasiones no había información suficiente.

3. Objetivos

3.1 Objetivo General

Construir un indicador de desempeño global para evaluar proveedores externos a una entidad bancaria para el año 2022, con técnicas estadísticas multivariantes

3.2 Objetivos específicos

- Describir el proceso metodológico para la construcción de un indicador global de desempeño laboral de un proveedor externo.
- Identificar variables idóneas para construir variables latentes que describan el desempeño integral de un proveedor externo, según fundamentos teóricos.
- Aplicar estadística descriptiva para la construcción de variables latentes que describan el desempeño integral de un proveedor externo, según fundamentos teóricos.
- Aplicar análisis factorial exploratorio para la identificación de variables unitarias que puedan conformar variables latentes que describan el desempeño integral de un proveedor externo, según fundamentos teóricos.
- Aplicar análisis factorial confirmatorio para validar constructos y puntuación global que describan el desempeño integral de un proveedor externo, según fundamentos teóricos.
- Aplicar métodos estadísticos no supervisados para crear perfiles de proveedores externos según su desempeño laboral.

4. Marco teórico

4.1 Proveedores

La evaluación de proveedores externos de una empresa puede tener dos líneas de aplicabilidad; ya sea, para elegir los candidatos que puedan realizar idóneamente una actividad o para medir y supervisar el desempeño realizado. El presente estudio se basó en la segunda línea de investigación; ya que, por lo general la selección de candidatos se realiza bajo estándares legales y aplicación de pruebas específicas del departamento de recursos humanos de la empresa contratante.

Dado lo anterior, se identificaron los supuestos teóricos imprescindibles para la construcción de un indicador que evalué de forma integral el desempeño de proveedores contratados, donde, indiferente de la labor que la empresa contratante se desempeñe se logran identificar aspectos relevantes a tomar en cuenta para tal evaluación. Entidades académicas especializadas en la formación de profesionales en las áreas de negocios, destacan conceptos como: (Conexión Esan, 2019):

- Reputación: brinda confianza y credibilidad, la cual puede ser medible a través de aspectos legales y logísticos.
- Calidad del bien o servicio: igual que la reputación es un aspecto intangible, no obstante, puede ser medible por tiempos de entrega, desempeño en su labor, quejas, sanciones, entre otros.
- Precios: es un aspecto tangible por sí solo, pero brinda información no solo para lo evidente: comparar la competencia en el mercado, en conjunto con otras variables sirve de referencia para evaluar calidad con respecto al servicio prestado y para evaluar riesgos del proveedor en caso de incumplimiento, ya que a mayor monto mayor riesgo para la empresa contratante.

Por otra parte, además de los conceptos anteriores Gimeno (2005), agrega:

- Eficiencia: identificable en resultados de auditoría, cumplimiento de plazos, ajuste a costos, cantidad de trabajos a su cargo, entre otros.
- Eficacia: puede ser medible a través de la satisfacción del bien o servicio, minimización de quejas o reclamos, reducción de costos.

- Aspectos legales: de acuerdo a cada empresa contratante, el apego a normas preestablecidas, este aspecto además esta ligado al aspecto reputacional.

Por último, es relevante mencionar variables o parámetros repetibles por múltiples autores, que apoyan la construcción de los conceptos antes mencionados, como lo son: la posición y situación financiera de la empresa, servicio postventa, cantidad de negocios anteriores, gestión y organización de la empresa, historial de conflictos y rendimiento (Güneri et al. 2011, como se cito en García et al., 2012).

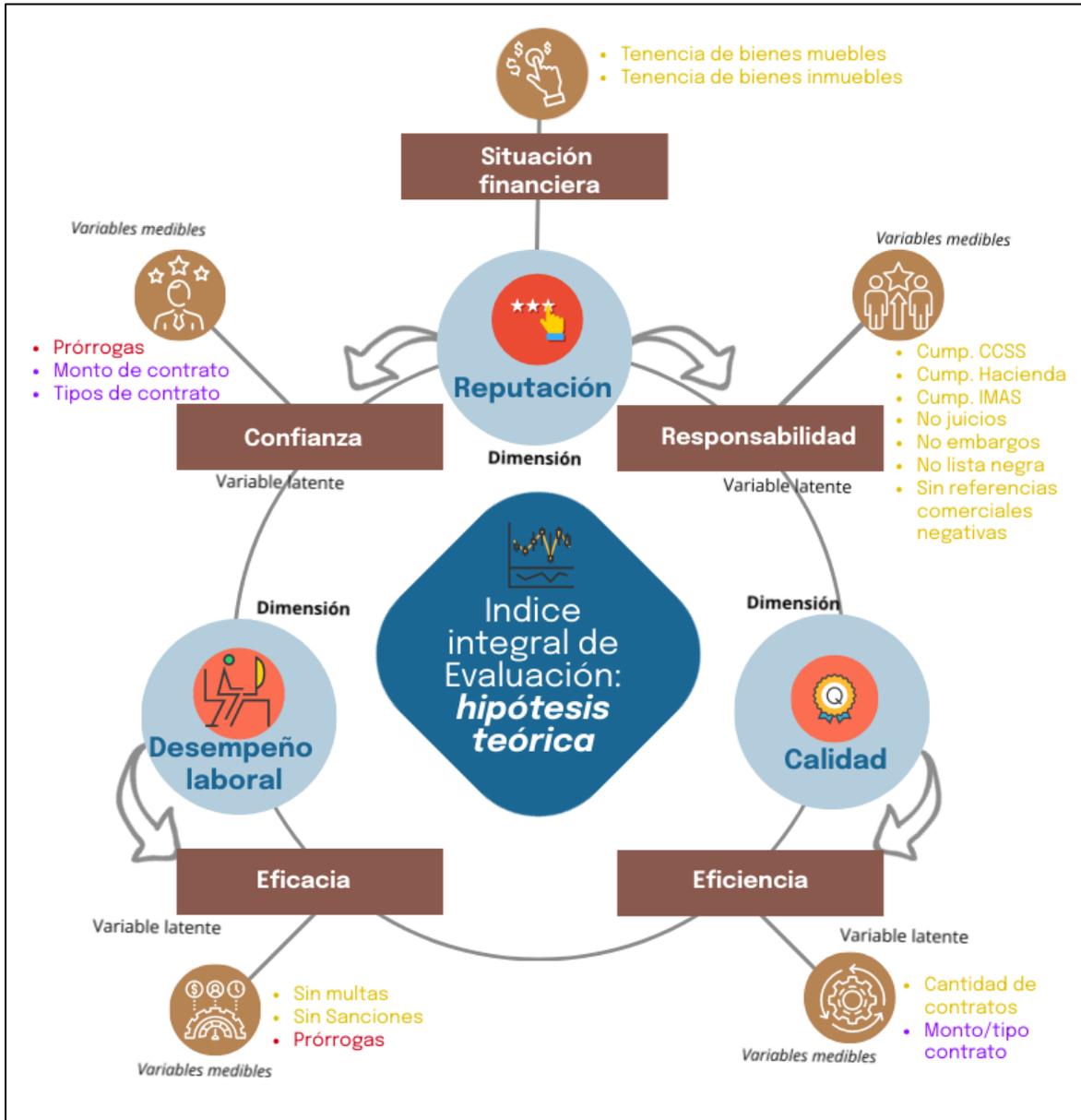
De acuerdo con los elementos teóricos identificados y estableciendo la relación al contexto de estudio, se construyó un escenario hipotético adecuando dimensiones que a su vez agrupan conceptos (constructos) medibles a través de variables disponibles para la investigación y que tuviesen relación a lo propuesto por la teoría.

En la Figura 1, se visualiza el escenario hipotético, donde se consideran tres grandes dimensiones:

- Reputación: según la REA se refiere a la “opinión o consideración que se tiene a alguien o algo” (Real Academia Española, s.f), donde, en términos del presente estudio, esta opinión puede estar derivada de la confianza, la responsabilidad y manejo de las finanzas del proveedor. Estos tres conceptos son los que se construirán a través de las variables disponibles.
- Calidad: Es la “propiedad o conjunto de propiedades inherentes a algo, que permiten juzgar su valor” (Real Academia Española, s.f.), este valor se considera puede ser juzgado por medio de la eficiencia que desempeñe el proveedor.
- Desempeño laboral: su evaluación “permite identificar de manera oportuna y eficaz fallas y limitaciones que se registran para corregirlas y garantizar el alcance de objetivos institucionales” (Pedraza et al., 2010), de acuerdo a esto es posible dar una medición del desempeño a través de la eficacia.

Figura 1

Hipótesis teórica conceptual de la conformación de variables latentes



4.2 Construcción de índices

En la construcción de índices es imprescindible la conceptualización de: constructo, indicador e índice. En cuanto al primer concepto, Montero, (2008), subraya que un constructo es un término que carece de una definición explícita y única, que requiere de un marco teórico muy sustentable y que justifique la definición según el problema, además destaca que una característica fundamental es el hecho de no ser observable o tangible por sí solo. Por otra parte, el concepto de indicador se define como la acción de operacionalizar un constructo o variable, donde el índice, viene a ser un tipo de indicador que relaciona variables o constructos de diferente naturaleza.

En términos generales, y para términos de la presente investigación se entenderá el constructo como el concepto, el indicador la categorización y selección de las variables que construyen el constructor y el índice la operacionalización final dada a los indicadores para construir un solo valor. En la figura 2, se visualiza la secuencia conceptual y un ejemplo del proceso metodológico que se siguió para realizar la construcción.

Figura 2

Conceptualización y ejemplo de la construcción de índices



4.3 Técnicas multivariantes

4.3.1 Análisis factorial

El análisis factorial, es una técnica estadística multivariante, la cual conceptualmente pretende, que, al tener una gran cantidad de variables o indicadores, se pueda realizar una reducción de estas, transformándolas en variables latentes o constructos que expliquen el fenómeno de interés, (Méndez y Rondón, 2012).

Matemáticamente el modelo factorial se expresa como un conjunto de relaciones lineales de la forma:

$$\begin{pmatrix} X_1 = P_{11}F_1 + \cdots + P_{1m}F_m + \epsilon_1 \\ \vdots \\ X_n = P_{n1}F_1 + \cdots + P_{nm}F_m + \epsilon_n \end{pmatrix}$$

Ecuación 1: Modelo de análisis factorial

De acuerdo con la Ecuación 1, los factores están representados por (F_1, \dots, F_m) , los cuales son los que explican de forma conjunta a las variables X_i , con $i = 1, \dots, n$, la parte que no se puede explicar por los factores, para cada variable X_i , esta representada por los errores $(\epsilon_i, \text{con } i = 1, \dots, n)$. Los errores son independientes entre sí, y no observables al igual que los factores, no obstante, estos no son independientes y además son aleatorios (Gutiérrez et al., Tema 1. Análisis Factorial, 1994).

El término $h_i^2 = P_{i1}^2 + \cdots + P_{im}^2$, se denomina comunalidad, donde cada P_{i1}^2 , es la variabilidad asociada a cada X_i , que resulta del factor (F_i) , donde, llamando a ψ_i , como la variabilidad que se puede explicar solo a través del factor común de cada modelo es entendible que se verifique la propiedad $h_i^2 + \psi_i = 1$.

Otro aspecto importante, subyacente de la Ecuación 1, es considerar a la matriz $P = (n \times m)$, la cual es la comúnmente llamada matriz de cargas factoriales o “loadings”, a esta matriz pertenecen los términos P_{ij} de cada X_i , respecto a su factor asociado.

Por otra parte, el modelo factorial, supone ciertas hipótesis probabilísticas básicas, las cuales se observan en la Tabla 1, específicamente para la esperanza y covariancia del modelo, donde, es importante además resaltar una definición que surge a partir de estas condiciones.

Tabla 1
Hipótesis probabilísticas en el análisis factorial

| Para la esperanza | Para la covariancia | Observaciones |
|---------------------|--|---|
| - $E[X] = \mu$ | - $Cov[X] = \Sigma$ <i>con Σ, matriz positiva de cov</i> | <i>Propiedades para la población</i> |
| - $E[F] = 0$ | - $Cov[F] = E[FF']$ | |
| - $E[\epsilon] = 0$ | - $Cov[\epsilon] = \Psi$, <i>con Ψ, diagonal de ψ_1, \dots, ψ_n</i> | |

De acuerdo con la tabla 1, se tiene $E[F] = 0$, cuando esto sucede se dice que el vector de factores comunes se encuentra centrado y si además de esta propiedad se cumple que $Cov[F] = E[FF']$, se está en presencia de un Análisis Factorial Ortogonal, ya que términos estadísticos y geométricos es la posición que toman los factores (Gutiérrez et al., Tema 1. Análisis Factorial, 1994).

Como último aspecto de importancia en lo referente al análisis factorial, en su aspecto general, son las restricciones (o no) con respecto a la matriz P, ya que esto supone una separación de dos de las formas más utilizadas de llevar a cabo un análisis de este tipo. Cuando no se da una asignación del valor 0, a priori para los factores o “loadings”, se habla de un modelo no restringido, lo cual conduce a un Análisis Factorial Exploratorio (AFE); por el contrario, cuando existe una serie de condiciones que se deben valorar de forma previa, se sugiere el uso de ceros para la identificación bajo normalidad, en este caso, se estaría frente a un modelo restringido, el cual se conoce como Análisis Factorial Confirmatorio (AFC), tal y como se cita en, Gutiérrez y González ((s.f), pág. 10)),

Cada uno de estos análisis tiene un objetivo en específico, con respecto al caso de estudio, el cual se verá con mayor detalle en las siguientes secciones. A manera de resumen, en la Figura 3, se puede observar una estructura teórica a seguir para realizar un modelo factorial.

Figura 3
Esquema conceptual de un análisis factorial



4.3.1.1 Análisis factorial exploratorio

Como se mencionó en el apartado anterior, el AFE, es un modelo no restringido, donde, a nivel técnico estas restricciones se dan únicamente para evitar indeterminaciones en la matriz de covariancias (Σ) (Gutiérrez et al., Tema 1. Análisis Factorial, 1994), sin embargo, desde la visión conceptual, lo que se busca es reducir los datos en factores o variables latentes que tengan un sentido teórico de acuerdo al problema de estudio, pero no hay nada previo esclarecido, por el contrario, se busca a través del análisis poder encontrar estos factores que puedan explicar de forma parsimoniosa el fenómeno de estudio.

El AFE, debe cumplir ciertos supuestos estadísticos, que permitan identificar que los resultados obtenidos son buenos para el estudio y que realmente se está explicando lo que se pretende. Según Pérez y Medrano (2010), los supuestos de normalidad, linealidad y

multicolinealidad, son esenciales de forma introductoria para permitir un análisis de este tipo, haciendo puntual énfasis en este último, recomendando dar atención a las correlaciones entre variables superiores a 0.90, ya que esto puede restar precisión al análisis; no obstante, en la práctica otros autores defienden que para un análisis factorial, los supuestos que se deben evaluar con mayor atención son los referentes a lo conceptual más que a lo estadístico y por esta razón es común no prestar gran atención lo referente a la normalidad y homocedasticidad tomando en cuenta que su no cumplimiento incurre en el riesgo de disminuir la correlación observada entre las variables (Chávez, 2017), que por contrario de los primeros autores, más bien sugiere fijarse en que las correlaciones no bajen de 0.30, porque podría ser un indicativo que el análisis no sea el más apropiado al conjunto de datos estudiado.

En síntesis, Tanto Pérez y Medrano (2010) como Chávez (2017), concuerdan en lo relacionado a la correlación de variables, por esta razón, para el presente estudio, se dará atención a este punto y otros indicadores estadísticos que proporcionan información de la calidad del modelo obtenido.

4.3.1.1.1 Indicadores de viabilidad de un análisis factorial exploratorio

- Matriz de correlaciones

La matriz de correlaciones o reproducción de la covariancia (*en el modelo teórico Σ*), permite la identificación en el modelo de la baja o alta relación entre las variables en estudio, de forma específica, se calcula el determinante de la matriz ($DET(\Sigma)$), dado que es un indicador del grado de multicolinealidad existente, como se explicó anteriormente, si existe una correlación muy alta (mayor a 0.90), puede ocasionar problemas para la interpretación de los factores y los resultados no sean los mejores para explicar el fenómeno de estudio. Por el contrario, valores muy bajos (menor a 0.30), podría indicar que las variables no se relacionan y realizar el análisis no sería conveniente, no obstante el determinante de la matriz funciona como indicador global de multicolinealidad, si el valor obtenido es muy cercano a cero, es cuando se podría estar en presencia de una muy alta multicolinealidad.

- Test de Bartlett:

Es una prueba para la verificación del supuesto de homocedasticidad u homogeneidad de las variancias, importante para identificar correlaciones positivas entre las variables, por lo cual si se obtienen valores pequeños de “K-squared” sería un indicativo de lo antes mencionado (Chávez, 2017), además, con la prueba se obtiene el p-valor, con el cual se plantea la hipótesis nula (H_0), de que todas las variancias de la población son iguales, frente a la hipótesis alternativa (H_1), que existen al menos dos diferentes.

En la práctica, la prueba se utiliza como un parámetro, más no, como restricción para frenar el análisis, y se busca el detrimento de los valores “K-squared” a medida que se mejore el modelo, más allá de una aceptación de H_0 .

- Medida Kaiser-Meyer-Olkin:

Es una medida de adecuación a la muestra, que indica la proporción de variancia que se puede explicar por los datos y no por otros factores ajenos al problema (Bolaños, 2020), por esta razón es un indicador de que tan adecuado es llevar a cabo el análisis con las variables involucradas. Múltiples autores han propuesto escalas de interpretación para esta medida, donde a grandes rasgos, si la medida general es inferior a 0.49, definitivamente el análisis no es conveniente, en valores de 0.50 a 0.59, es posible realizar el análisis, pero se sugiere revisar las variables, en el rango de 0.60 a 0.79, ya es aceptable, y de 0.80 en adelante es un escenario ideal.

4.3.1.1.2 Construcción del modelo factorial exploratorio.

Al determinar la viabilidad del modelo factorial exploratorio, el siguiente paso, es la extracción de factores o los llamados “factor loadings”, para esto se cuenta con dos métodos, por componentes principales y máxima verosimilitud. En este estudio se hace énfasis en el segundo método, dado que el procesamiento se lleva a cabo con la librería “Stats” versión 4.2.2 del Software R (R Core Team, 2022), que en su algoritmo interno utiliza este método.

En la Tabla 1, se definió el Modelo Factorial Ortogonal, donde su estructura esta dada por

$$X = \mu + PF + \epsilon; \text{ con } \Sigma = PP' + \Psi$$

X, y el vector de error ϵ , se suponen normalmente distribuidas, entonces los estadísticos muestrales

$$\bar{X} = \hat{\mu} = \frac{1}{N} = \sum_{\alpha=1}^N X_{\alpha} \quad \frac{s}{N} = \hat{\Sigma} = \frac{1}{N} = \sum_{\alpha=1}^N (X_{\alpha} - \bar{X})(X_{\alpha} - \bar{X})'$$

Serían los estimadores máximo-verosímiles de los parámetros μ y Σ .

A manera de resumen, en la Figura 4, se propone una estructura para realizar un AFE, incorporando los elementos estadísticos descritos y recomendaciones conceptuales sugeridas (De la fuente, 2011)

Figura 4

Propuesta metodológica para realizar un AFE



4.3.1.2 Análisis factorial confirmatorio

A diferencia de su predecesor, se utiliza cuando el investigador tiene una teoría conceptual que le respalde y desea realizar una comprobación de su teoría. Desde su concepción matemática, es la referencia a un modelo restringido ya que el investigador usa el conocimiento a priori para plantear hipótesis, las cuales conducen a restricciones para el modelo, por ejemplo; que algunos “loadings” deben tener un valor de 0 (Gutiérrez et al., Tema 1. Análisis Factorial, 1994).

En un AFC, los supuestos o indicativos de calidad del modelo son diferentes a los utilizados en el AFE, dado que aquí los factores o variables latentes ya han sido construidos previamente o ya se tiene una claridad de cuál debe ser la construcción ideal para explicar el fenómeno de estudio.

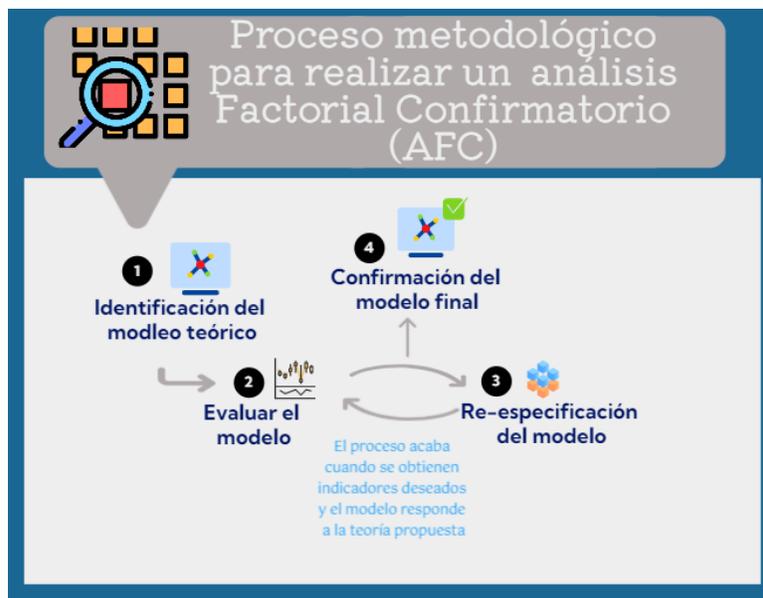
Los indicadores de calidad del modelo más utilizados son el índice de Tucker-Lewis (TLI), Índice de ajuste comparativo (CFI), Raíz cuadrada del error cuadrático medio (RMSEA) y la media de residuos cuadráticos (SRMR).

Los índices TLI y CFI, tiene como objetivo la evaluación de un modelo hipotético ante un modelo base, en ambos casos cuando la puntuación otorgada es superior a 0.95, se puede considerar que el modelo presenta un ajuste óptimo a la muestra, siendo el 1, el valor máximo (Lai, 2021).

Los índices RMSEA y SRMR sirven para realizar una evaluación del modelo en términos absolutos, donde se compara la matriz de covariancias del modelo hipotético vs. el modelo base, en ambos casos mientras menor sea el valor mejor es el ajuste, por lo general se toma de referencia ser menor a 0.05 (Jordan, 2021).

De acuerdo con lo valorado anteriormente, en la Figura 5, se muestra la metodología propuesta en el presente estudio para aplicar un AFC.

Figura 5
Propuesta metodológica para realizar un AFC



4.3.2 Análisis de conglomerados

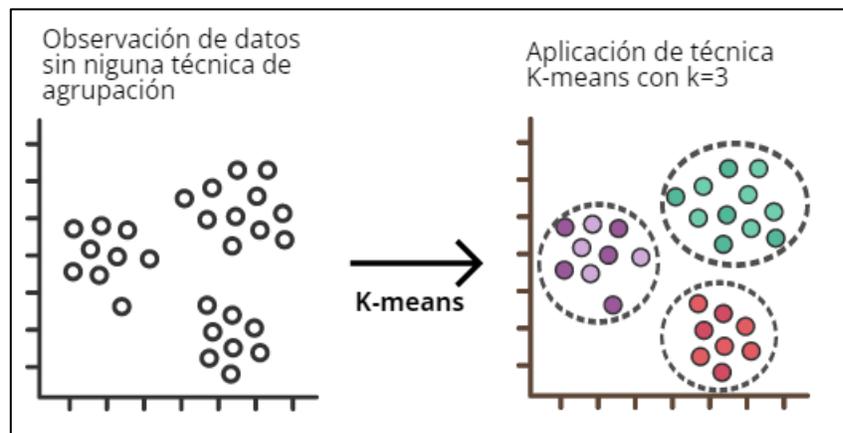
La técnica por conglomerados o más conocida como “clustering” o “cluster”, es parte de las llamadas técnicas no supervisadas y su objetivo principal es identificar grupos o en un conjunto de datos, donde cada uno de estos grupos dentro de si mismo son similares, pero no entre ellos (Amat, 2017).

En este tipo de análisis intervienen dos conceptos fundamentales, la técnica de particionamiento; entre las cuales se encuentran agrupaciones jerárquicas, agrupaciones no jerárquicas y técnicas combinadas. El otro concepto de importancia es el tipo de distancia utilizada para agrupar los elementos observados. Las distancias más comúnmente utilizadas son la distancia Euclídea, distancia de Manhattan, Distancias de Pearson, entre otras. Dado los objetivos de esta investigación, se hace énfasis en la técnica no jerárquica: K-Medias y la Distancia Euclídea.

El método de clasificación por K-Medias, tiene como particularidad y en común con todos los métodos no jerárquicos, ser una técnica de optimización; donde el investigador debe tener una idea previa de cuantos grupos debería formarse, en la Figura 6, se visualiza esta percepción de primero visualizar de forma intuitiva como se agrupan los elementos en la trama de datos, o cual es la cantidad de grupos idónea, para luego “confirmar” esta agrupación por medio de la técnica.

Figura 6

Idea intuitiva de formación de clúster por K-Medias



Como se mencionó anteriormente, la gran diferencia de los métodos jerárquicos con los no jerárquicos, es el hecho de tener preconcebido la cantidad de grupos, pero no es la única diferencia, ya que; en los métodos no jerárquicos se trabaja con una matriz de distancias, lo cual lleva a la

necesidad de especificar la técnica para calcular dicha matriz, en este caso específico se detalla el método por Distancia Euclídea.

En coordenadas cartesianas, la Distancia Euclídea, se calcula a través del teorema de Pitágoras, donde dado un par ordenado (x,y) , la distancia entre dos puntos a y b, está dada por:

$$d_{Euclídea}(a,b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} = \sum_{i=1}^n (a_i - b_i)^2$$

Donde, para el análisis de conglomerados, esta distancia tiene como objetivo minimizar la distancia entre observaciones (Amat, 2017).

La elección del método de partición de K-Medias y la Distancia Euclídea se realizó debido a que todo los índices están fundamentados en una teoría, lo cual influyó en el hecho de tener una previsión de cuantos grupos podrían tenerse al finalizar los análisis factoriales.

Por último, es imprescindible enunciar el esquema metodológico que se puede seguir para aplicar el análisis, Gutiérrez et al., (Capítulo 6. Introducción al Análisis Cluster. Consideraciones generales., 1994), sugiere cuatro etapas primordiales:

- 1- Desarrollar una clasificación previa
- 2- Indagar estructuras conceptuales
- 3- Idear a través de la exploración de los datos, hipótesis que se adapten o identifiquen la población de estudio.
- 4- Contrastar las hipótesis planteadas con estructuras conceptuales predefinidas.

5. Metodología

5.1 Método científico de investigación

El proceso de investigar conlleva el apego a una estructura o método que le respalde, existen múltiples tipificaciones y ramas, que van desde las clásicas metodologías cuantitativas, cualitativas y mixtas, en las cuales es posible encausar casi cualquier esquema investigativo (Hernández et al., 2006), no obstante el creciente auge del análisis de datos ha llevado a una diversificación y especialización de las mismas, que permite a los investigadores adaptarse a esquemas más específicos de acuerdo al problema que se aborda. Dado que el presente estudio aborda diferentes etapas y métodos, pero de forma sustancial la manipulación de técnicas estadísticas multivariantes para la creación de índices, es por esta razón que se previó la adaptación al Paradigma Metodológico de Churchill's (Churchill, 2008).

El modelo propuesto por el investigador de mercados Gilbert Churchill, es una variante al modelo de Rasch, donde una de las principales diferencias es que Churchill proporciona un esquema robusto para todo el proceso y no solo para la validez de las medidas.

En la figura 7, se visualiza el proceso que se debe seguir según la metodología propuesta, con ligeras adaptaciones del esquema clásico, dado que en este caso no se aplican medidas psicométricas clásicas, las cuales se omiten del método.

Figura 7

Adaptación de metodología de Churchill al contexto de la investigación



5.2 Definición de variables

Para realizar el modelo en principio se tenían alrededor de 45 variables, no obstante muchas tuvieron que ser eliminadas por inconsistencias, como; datos nulos, valores en blancos, variables mal definidas y en otros casos por tener poca relación al estudio.

En la Tabla 2 se muestran las variables finales con las que se realizó el estudio. Para mas detalle de la codificación de las variables ver ANEXO 1.

Tabla 2:
Definición de variables utilizadas en el estudio

| Variable original | Código de variable | Definición | Tipo de variable | Niveles |
|-----------------------|-----------------------------|--|------------------|---------|
| Prorroga | TIENE_ PRORROGAS | Define si el proveedor se le ha otorgado o no una extensión a su contrato | Binaria | 2 |
| Monto de contrato | MONTO_CAT_TI PO_CONTRATO | El monto por el cual se ha contratado al proveedor | Ordinal | 5 |
| Contrato compartido | CONTRATOS_ COMPARTIDOS | Característica de algunos proveedores, los cuales no tienen los insumos para realizar la actividad de forma independiente, por lo cual se asocian a otro proveedor y obtiene un contrato único para ambos. | Binaria | 2 |
| Cantidad de contratos | CAT_Q_CONTRATOS | Cantidad de contratos que posee el proveedor con la entidad contratante | Ordinal | 3 |
| CCSS ¹ | CCSS_INDICE | Corresponde a la situación de los proveedores con la Caja Costarricense de Seguro Social, en la cual se identifica si posee deudas | Ordinal | 4 |
| Hacienda ² | HACIENDA_ INDICE | Corresponde a la situación de los proveedores con el Ministerio de Hacienda, en la cual se identifica si posee deudas. | Ordinal | 4 |
| IMAS | IMAS ³ _INDICE | Se asume que un proveedor que esta al día en IMAS, es un proveedor responsable de sus pagos sociales | Ordinal | 4 |
| Bienes Muebles | B_MUEBLE | Si el proveedor posee bienes muebles | Binaria | 2 |

¹ Institución pública que se encarga en Costa Rica de proveer todos servicios relacionados al sector salud

² Institución gubernamental encargada de todo lo referente a comercio

³ Institución pública encarga de bienestar social y familiar

| Variable original | Código de variable | Definición | Tipo de variable | Niveles |
|-------------------------|--------------------|---|------------------|---------|
| Bienes Inmuebles | B_INMUEBLE | Si el proveedor posee bienes inmuebles | Binaria | 2 |
| Multas | MULTA | Indica si el proveedor a sido multado | Binaria | 2 |
| Sanciones | SANCION | Indica si el proveedor ha sido sancionado | Binaria | 2 |
| Juicio | JUICIO | Indica si el proveedor a sido objeto de juicios | Binaria | 2 |
| Embargos | EMBARGO | Indica si al proveedor le han realizado embargos | Binaria | 2 |
| Lista negra | L_NEGRA | Indica si el proveedor a aparecido en lista negra de lavado de dinero | Binaria | 2 |
| Referencias comerciales | R_COMER | Indica si el proveedor se encuentra en lista de malas referencias comerciales | Binaria | 2 |
| Tipo de proceso | T_PROCESO | Indica si el proveedor a terminado un contrato, ejecución de garantía o no prórroga | Ordinal | 3 |
| Tipo contrato | TIPOS_CONTRATO | Indica si el contrato es de bienes, servicios o académico. Usado para clasificar. | Categoría | 3 |

6. Aplicación de escalas construidas

6.1 Constructos

Como parte de la investigación y en conjunto con un grupo de expertos, guiados por metodología especializada sugerida para este tipo de construcciones, se llevó a cabo la creación de constructos o variables latentes, identificados en la teoría como fundamentales para la creación del índice integral de evaluación a proveedores.

La finalidad de realizar la construcción de esta forma fue obtener una creación concebida por un grupo de trabajo interdisciplinario, que utilizara las variables que mejor pudiesen estimar dichos conceptos, pero que su construcción fuese idónea a los datos con los que contaba la institución y en estudios posteriores realizar una verificación de las variables unitarias por medio de un análisis factorial confirmatorio.

A continuación, se detallan los constructos realizados:

6.1.1 Confianza

Según los aspectos teóricos investigados, el concepto confiere gran importancia para el estudio, ya que se refiere a la “esperanza firme que se tiene de alguien o algo” (Real Academia Española, s.f., definición 1). En términos del estudio pretendió estimar la seguridad que el proveedor le da al empleador (entidad bancaria) para seguir contratándolo. En este sentido se contó con tres variables que aproximaron este concepto, las cuales se muestran en la Tabla 3.

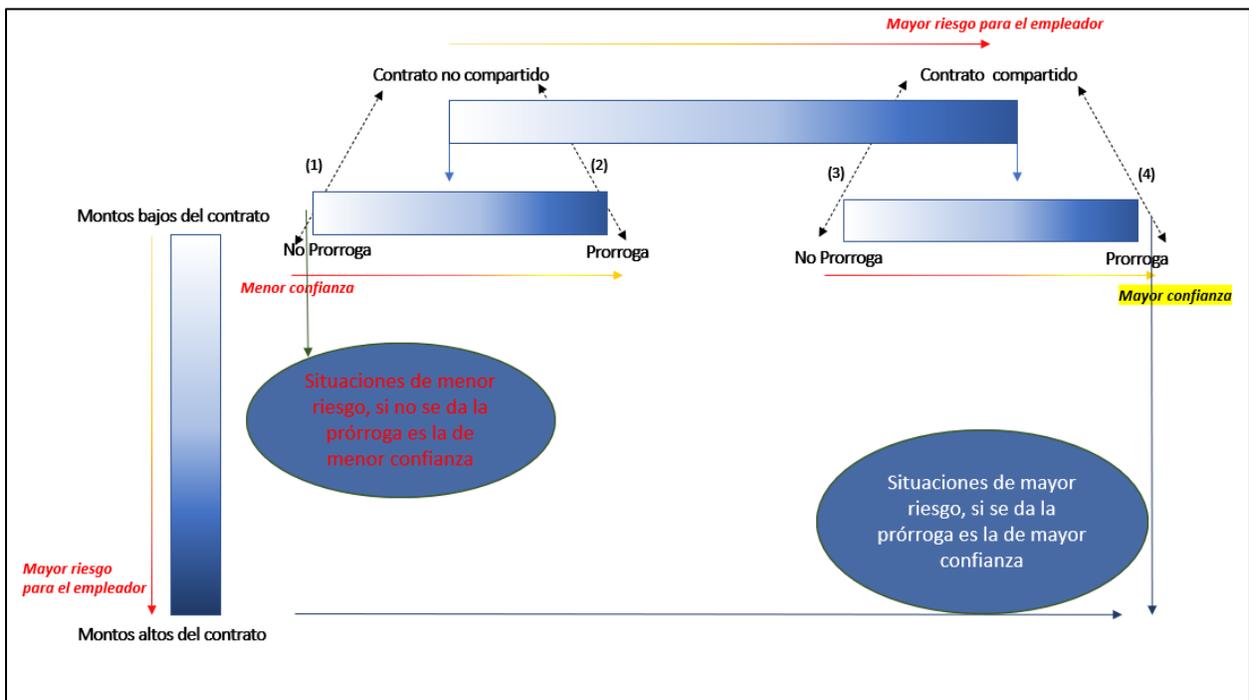
Tabla 3

Constructo Confianza: descripción de variables unitarias utilizadas

| Variable original | Tipo de variable | Niveles | Recodificación |
|---------------------------|------------------|---------|----------------------|
| Prorroga(x16) | Binaria | Si | (1) |
| | | No | (0) |
| Monto de contrato (x17) | Numérica | | Se crearon 5 niveles |
| Contrato compartido (x18) | Binaria | Si | (1) |
| | | No | (0) |

La elección de las tres variables (Tabla 3), refiere a la concepción de que a mayor monto la entidad tiene un mayor riesgo, si el proveedor le llegara a fallar, pero si se da el contrato implica confianza hacia el proveedor, si a esto, se le agrega el hecho de que el proveedor pueda tener un contrato compartido (la cual es una situación inherente del proveedor, es decir no la controla en el contrato, pero si decide si lo acepta o no), le agrega niveles a la confianza y por último, si con estas condiciones el proveedor vuelve a ser contrato, implicaría un alto grado de confianza. En la figura 8, se muestra una estructura conceptual para la variable latente.

Figura 8
Constructo Confianza: Esquema teórico conceptual



De acuerdo a la estructura conceptual mostrada en la Figura 8, se visualiza que la variable contrato compartido se jerarquiza por encima de la variable prórroga, esto debido a que el contratante no tiene control sobre este hecho, entonces de esta manera, también se pudo definir la

forma de calcular ponderadores a los diferentes niveles de cada grupo que se conformaron, ya que; al tomar la probabilidad condicional dada por:

$$P(A \setminus B) = \frac{P(A \cap B)}{P(B)} \text{ donde,}$$

A: Tener (o no) prórroga

B: Tener (o no) un contrato compartido

Se tomaron las probabilidades como ponderadores distribuidos en los diferentes niveles del monto, para los cuatro grupos que se crearon (ver Tabla 4), los cuales también representaron una jerarquía.

Tabla 4

Conformación de grupos para la construcción de la variable latente Confianza

| Prórroga | Contrato Compartido | |
|----------|------------------------------------|------------------------------------|
| | Si | No |
| Si | Grupo 1 (grupo de mayor confianza) | Grupo 2 |
| No | Grupo 3 | Grupo 4 (grupo de menor confianza) |

De acuerdo con la Tabla 4, se evidencia la jerarquía formada por los cuatro 4 grupos, ubicados de mayor a menor confianza. Posteriormente las probabilidades condicionales calculadas en cada grupo, para cada nivel de monto; otorgó los niveles totales de un índice de 20 categorías, que posteriormente de situaron en rangos para conformar el índice final. Como último paso se presentó a un grupo de expertos que validaron a nivel conceptual las categorías creadas.

6.1.2 Responsabilidad Legal

De forma análoga a la construcción realizada para la variable latente Confianza, se creó el constructo Responsabilidad Legal, el cual combinó las variables: Juicios, Embargos, Lista Negra, y Malas referencias comerciales. Se ordenó de mejor reputación, siendo esta la combinación de no tener ninguna de las 4 variables, es decir no tener juicios, embargos, malas referencias comerciales, ni encontrarse en lista negra. La peor situación sería representada por tener presencia en las 4 variables.

6.1.3 Responsabilidad Social

La creación del constructo se realizó de forma similar a las anteriormente presentadas, donde, de forma teórica se asumió, que un proveedor que está al día en Hacienda, en la CCSS e IMAS, es un proveedor responsable de sus pagos sociales, siendo este el nivel más alto de responsabilidad social y su contrario el menor nivel del índice.

6.1.4 Seguridad Económica

A nivel conceptual se definió que el tener bienes muebles e inmuebles, es un respaldo de poder responder ante posibles eventualidades económicas. Para este constructo se debió tomar en cuenta que un bien inmueble es de mayor valor un bien mueble, por tanto; se modeló una ecuación lineal simple donde:

$$X = \text{bien mueble},$$

$$2x = \text{bien inmueble},$$

De esta forma, se obtuvo $f(x) = x + 2x \Rightarrow x + 2x = 100 \Rightarrow x = 33$, utilizando esta ponderación se construyó el constructo de forma análoga a los anteriores.

7. Aplicación de las Técnicas multivariantes

7.1 Análisis factorial exploratorio

Como primer paso se realizó un análisis descriptivo de las variables analizadas en el AFE, el cual se puede observar en la Tabla 5.

Tabla 5

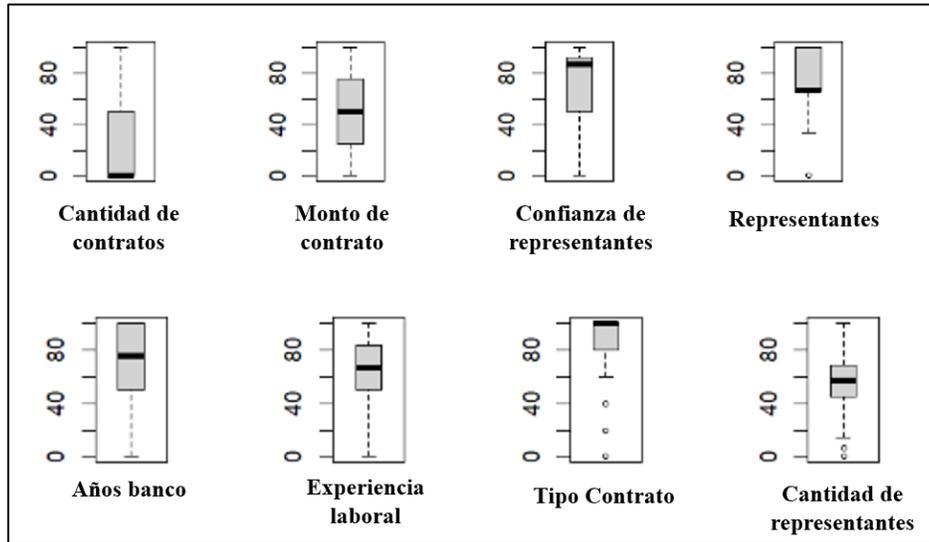
Medidas descriptivas para variables utilizadas en análisis factorial exploratorio

| Variable | Código en análisis | Media | Mediana | Moda | Desviación estándar |
|-----------------------------|--------------------|-------|---------|-------|---------------------|
| Cantidad de contratos | CQ | 31,68 | 0,00 | 0 | 39,50 |
| Monto de contrato | MC | 51,51 | 50,00 | 100 | 35,66 |
| Confianza de representantes | CONF | 72,83 | 87,00 | 87 | 26,13 |
| Representantes | REPRE | 72,56 | 66,67 | 100 | 31,99 |
| Años Banco | ANT_BAN | 67,51 | 75,00 | 100 | 28,74 |
| Experiencia laboral | EXP_LAP | 61,46 | 66,67 | 66,66 | 26,59 |
| Tipo de contrato | TC | 86,98 | 100,00 | 100 | 23,03 |
| Número de representantes | N_REP | 55,74 | 56,67 | 60 | 15,90 |

De acuerdo con la Tabla 5, se puede observar que la variable Cantidad de Contratos y Número de representantes tienen un promedio más bajo que el resto de las variables analizadas, sin embargo el dato más relevante se puede observar en la desviación estándar, ya que; tomando en cuenta que todas las variables son indicadores en escala de 1 a 100, tienen una dispersión de 20 hasta los 40 puntos porcentuales aproximadamente, a excepción de la última variable en la tabla que es la de menor dispersión con 15. Por otra parte, para las tres medidas analizadas en cada variable es notable que en ningún caso coinciden exactamente, lo cual habla de la simetría, hecho que se puede observar con más claridad en la Figura 9.

Figura 9

Diagramas de caja para variables utilizadas en el análisis factorial exploratorio



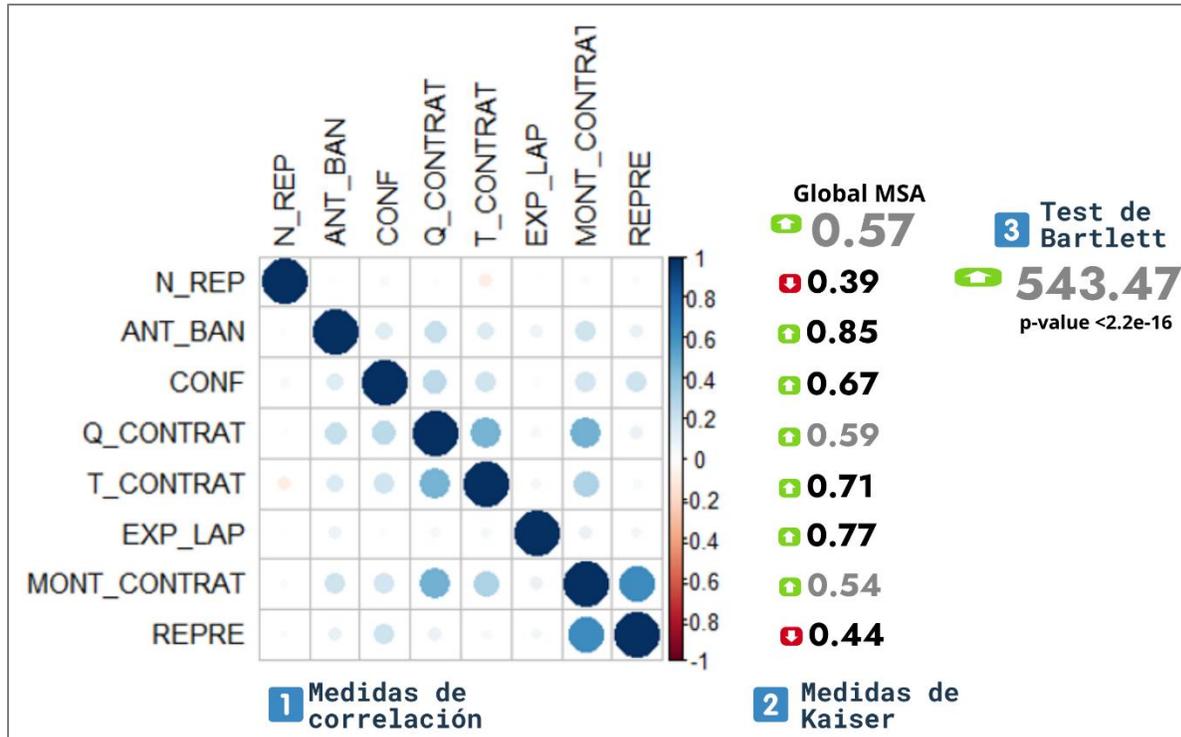
De la figura 9, es observable que las variables: Cantidad de representantes y Monto de contrato son las que tienen una distribución más simétrica, no obstante, esta última es la de mayor dispersión. Por otra parte, Confianza de representantes, Representantes, Años banco, Experiencia laboral y Tipo de contrato, presentan asimetría negativa, siendo esta la predominante y solo la variable Cantidad de contratos presenta asimetría positiva. Por último, es destacable que las variables; Representantes, Tipo de contrato y Cantidad de Representantes, evidencian valores extremos que deben analizarse a fin de interpretar si estos son “puntos aislados” o realmente tienen ese valor, en cuyo caso debe determinarse si los valores pueden afectar el análisis y sea mejor eliminarlos.

Posteriormente, se comenzó a indagar los primeros indicadores sugeridos por la teoría para iniciar un AFE. En primer lugar, se realizó la matriz de correlaciones, se calculó la medida de Kaiser y el test de Bartlett con todas las variables disponibles para esta etapa. En la figura 4 se pueden visualizar que las variables, N_REP, EXP_LAP y REPRE son las que mostraron menor correlación con el resto de las variables (1), sin embargo, la variable EXP_LAP tuvo un MSA deseable en la prueba de Kaiser (2), además de ser una variable de interés para el estudio. A manera general del conjunto de datos, se presentaron como aspectos de mejora, aumentar el MSA global, dado que según la escala teórica, un valor de 0,57, es un indicador con el cual se puede seguir adelante con el análisis, pero sería deseable mejorarlo, y el valor del “Bartlett's K-squared” es

recomendable que disminuya, dado que según los expertos a menores valores de este indicador se indica una mejor homogeneidad en las variancias (Chávez, 2017).

Figura 10

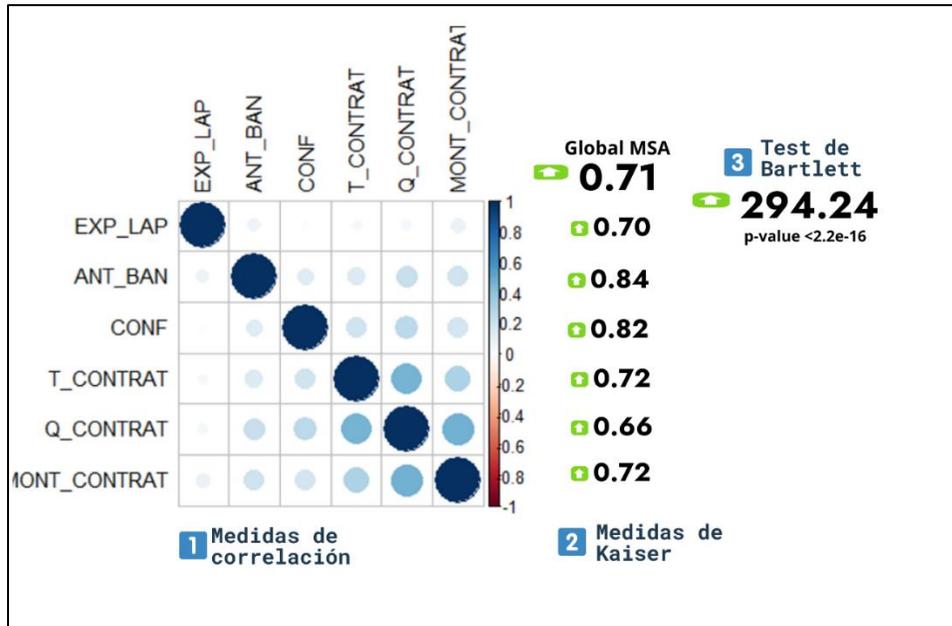
Parámetros de calidad para identificar viabilidad de AFE con el total de variables



Dado los resultados obtenidos en la Figura 10, se procedió a mejorar estos indicadores, por lo que se excluyeron las variables N_REP y REPRE y se repitió el proceso anteriormente descrito, el cual se visualiza en la Figura 11, donde se puede apreciar, que si bien a nivel correlacional (1) la variable EXP_LAP es la de menor desempeño, en cuanto a las medidas de Kaiser (2) se incrementan gradualmente, llegando a valores teóricos muy satisfactorios, y en el test de Bartlett, se logra disminuir el indicador en un 47% con respecto al obtenido en la Figura 10.

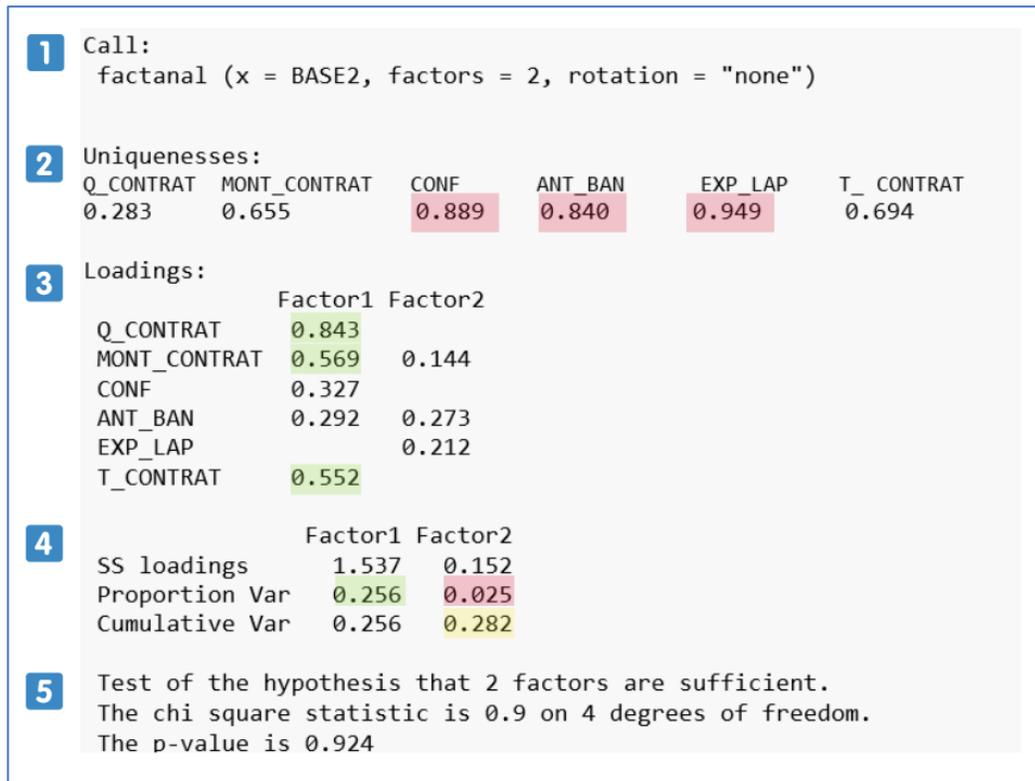
Figura 11

Parámetros de calidad para identificar viabilidad de AFE excluyendo variables N_REP y REPRE



Con los resultados de este segundo análisis, se procedió al cálculo de los factores. En la Figura 12, se visualiza los resultados obtenidos, donde se realizó un primer estudio sin aplicar una rotación específica e inicialmente con dos factores (1), dado que se utilizó el paquete estadístico de R, 'stats' versión 4.2.2 (R Core Team, 2022), el cual utiliza por defecto la prueba chi cuadrado, con $h_0 = dos\ factores\ son\ suficientes$. En cuanto a la proporción de variancia que no pudo ser explicada por los factores, resultados que se ven reflejados por las unicidades (2), resalta las variables CONF, ANT_BAN Y EXP_LAP, que tienen valores superiores al 70%, lo cual refleja poco aporte en los factores. Por otra parte, se obtienen las comunalidades (3) donde se evidencia que las variables Q_CONTRAT, MONT_CONTRAT y T_CONTRAT, se encuentran definidas en el primer factor, y el resto no muestran una ubicación adecuada a ninguno de los dos factores, además que el valor de variancia explicada que proporcionan es muy bajo. Por último, se puede observar que el porcentaje de variancia explicada por los dos factores es de 28%, donde el 25.6 proviene del factor 1(4) y la prueba chi cuadrado, arroja que dos factores no son suficientes, no obstante esto puede deberse a la cantidad de variables en análisis.

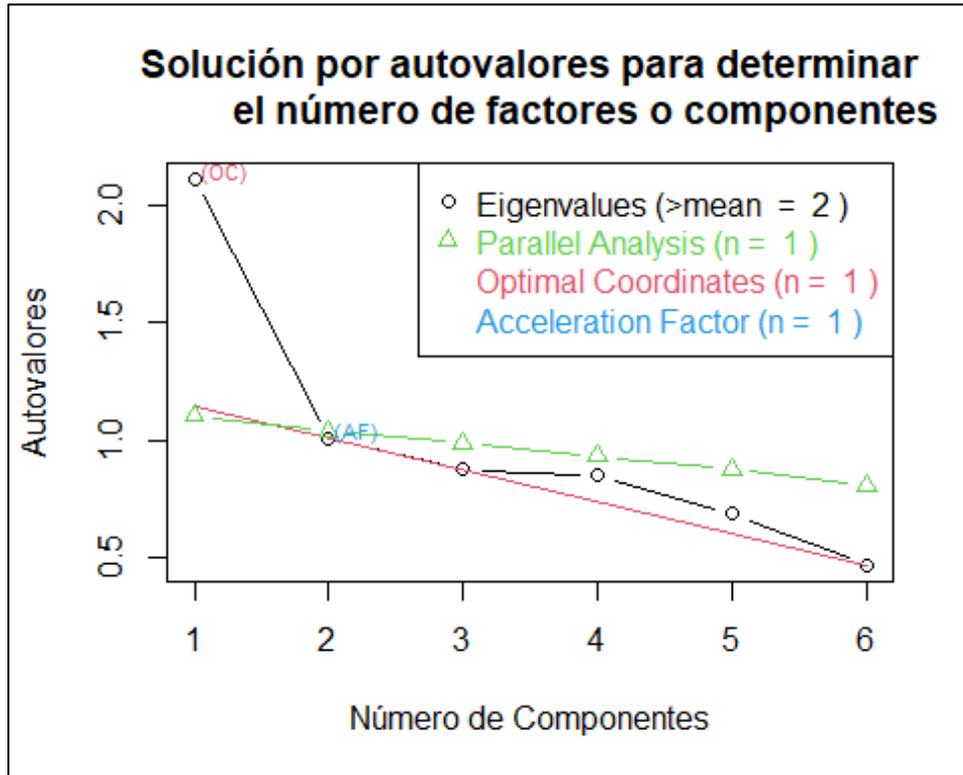
Figura 12
Resultados para AFE 1



Es importante resaltar, que en este punto a nivel conceptual, las variables ubicadas en el factor 1, sí corresponden a un fundamento teórico, el cual se puede conceptualizar como “Naturaleza de contrato”, dado que los tres elementos hacen referencia a un control administrativo, concepto de suma importancia en la evaluación de un proveedor (Matín, 2022), mientras que las tres variables restantes no se visualizan conceptualmente como un constructo en ninguna de las formas que pudiesen agruparse, no obstante con la finalidad de saber si estos resultados se podían mejorar, se realizó una rotación con “varimax”, la cual no arrojó resultados muy diferentes a los obtenidos en la figura 12, (ver ANEXO 2), por último se obtuvo el gráfico de sedimentación el cual mayormente sugiere un factor (Figura 13)

Figura 13

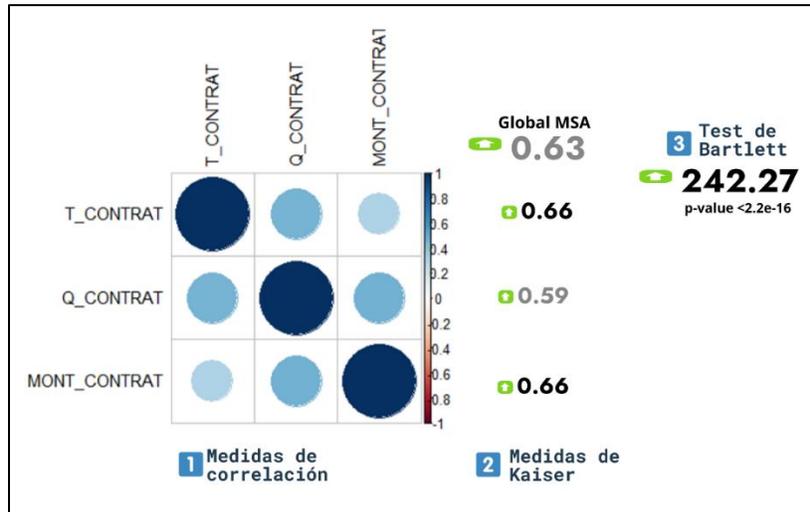
Gráfico de sedimentación para AFE 1



Dados los resultados obtenidos tras el primer AFE, se procedió a realizar de nuevo el proceso ejemplificado a partir de la Figura 10, solamente con las variables T_CONTRAT, Q_CONTRAT y MONT_CONTRAT, donde se obtuvo una matriz de correlación con valores muy acorde a los sugeridos por la teoría (Figura 14, (1)), en cuanto al MSA en la prueba Kaiser; si bien, es menor al obtenido en el anterior proceso, sigue siendo un indicador adecuado (2), además al ser solo tres variables este valor tiende a bajar, por lo cual es bastante satisfactorio para solo tres variables, por último el test de Bartlet (3) fue inferior al obtenido en el último proceso, lo cual era lo que se requería.

Figura 14

Parámetros de calidad para identificar viabilidad de AFE excluyendo variables EXP_LAP, ANT_BAN y CONF



Con los indicadores obtenidos en la Figura 14, se procedió al cálculo de los factores, el cual se visualiza en la Figura 15. El proceso al igual que el anterior se realizó sin ninguna rotación especial al inicio, no obstante, solo se indicó un factor (1), por ser solo tres variables y ya estar “apuntando” a un concepto teórico. Es observable que la proporción de variancia que no pudo ser explicada por los factores, es inferior al 70% para las tres variables (2), en cuanto a las comunialidades (3) se definen bien en el único factor y con valores por encima del 50%. Por último, se puede observar que el porcentaje de variancia explicado del factor es de 44% (4), lo cual representa casi el doble de lo que se había obtenido la primera vez. Adicionalmente se obtuvo el gráfico de sedimentación el cual sugirió un factor, lo cual confirmó los resultados obtenidos, además se realizaron las rotaciones “promax” y “varimax”, las cuales no arrojaron ningún resultado diferente al obtenido en la Figura 9 (ver ANEXO 1), no obstante, se procedió a calcular las puntuaciones con “varimax”, ya que “al ser una rotación ortogonal, tiende a asimilar cada variable con un eje” (Estamática, 2017, p. 6).

Figura 15
Resultados para AFE 2

```

1 Call:
factanal (x = BASE3, factors = 1, rotation = "none")

2 Uniquenesses:
  Q_CONTRAT MONT_CONTRAT  T_CONTRAT
    0.282      0.692      0.696

3 Loadings:
          Factor1
Q_CONTRAT  0.848
MONT_CONTRAT 0.555
T_CONTRAT  0.551

4          Factor1
SS loadings  1.330
Proportion Var 0.443

The degrees of freedom for the model is 0 and the fit was 0

```

En la figura 16, se muestran las puntuaciones factoriales obtenidas directamente del factor, con la rotación “varimax” (1), sin embargo, fue necesario realizar una normalización para poder utilizarlo como indicador (2).

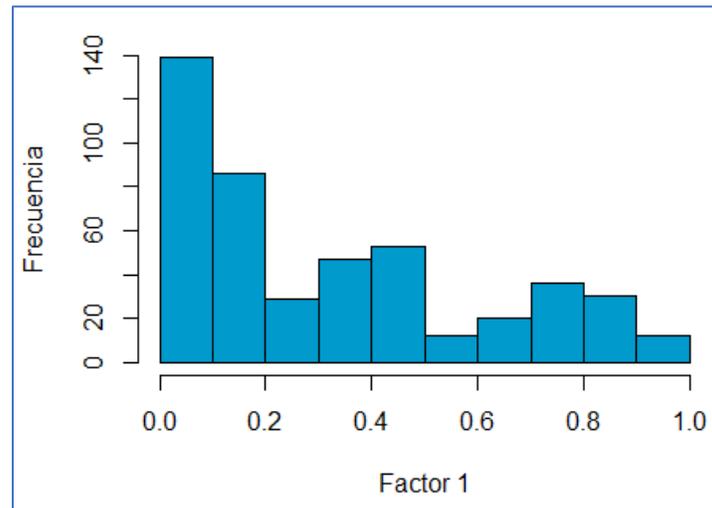
Figura 16
Puntuaciones sin normalizar y normalizadas para el factor obtenido

| 1 | Primeros 10 valores del factor | 2 | Primeros 6 valores del factor |
|---|---|--|-------------------------------|
| | Sin normalizar | | Normalizado |
| | Rotación Varimax | | Rotación Varimax |
| | <pre> factanal (BASE3, factors = 1, rotation = "varimax", scores = "regression") \$scores Factor1 [1,] 0.10066603 [2,] -0.88129866 [3,] -0.37540354 [4,] -0.78468930 [5,] 0.48008738 [6,] -0.37540354 [7,] 0.48008738 [8,] -0.50187732 [9,] -0.62835110 [10,] -0.18214594 </pre> | <pre> puntuaciones <- factanal(BASE3, factors = 1, rotation = "varimax", scores = "regression")\$scores BASE3 <- cbind(BASE3, puntuaciones) BASE3\$Factor1 <- round(((BASE3\$Factor1 - min(BASE3\$Factor1))/(max(BASE3\$Factor1) - min(BASE3\$Factor1))), 2) head(BASE3) Q_CONTRAT MONT_CONTRAT T_CONTRAT Factor1 1 50 25 16.66667 0.36 2 0 0 16.66667 0.05 3 0 100 16.66667 0.21 4 0 50 0.00000 0.08 5 50 100 16.66667 0.48 6 0 100 16.66667 0.21 </pre> | |

Por último, en la Figura 17, se visualiza la distribución del factor con las puntuaciones normalizadas, finalmente nombrado “Naturaleza de contrato”.

Figura 17

Gráfico de distribución del factor "Naturaleza de contrato"



7.2 Análisis Factorial Confirmatorio (AFC)

7.2.1 Análisis iniciales

En esta sección se procede a detallar los resultados de aplicar el Análisis Factorial Confirmatorio (AFC), para la construcción de un indicador que evaluara de forma integral el desempeño de un proveedor.

Como se explicó en las secciones anteriores, primero se realizó una construcción de indicadores de forma teórica con insumos matemáticos y estadísticos básicos, en apoyo de un grupo de expertos, se agregó además un indicador creado con la técnica de AFE, finalmente en esta sección se aplica un análisis confirmatorio. Como primer paso se recomienda tener una estructura teórica específica de lo que se desea confirmar (Peña, 2018), por lo cual en la tabla 6 se presenta dicho insumo.

Tabla 6*Estructura de indicadores para el Análisis Factorial Confirmatorio (AFC)*

| Factor | Nombre del factor | Indicadores |
|----------|--------------------------|---|
| Factor 1 | Responsabilidad legal | Juicio (x1), Embargo (x2), Lista_negra (x3), Referencias (x4) |
| Factor 2 | Responsabilidad social | Hacienda (x5), CCSS (x6), Imas (x7) |
| Factor 3 | Seguridad económica | Bienes muebles (x8), Bienes inmuebles (x9) |
| Factor 4 | Cumplimiento de servicio | Sanciones(x10), Multas (x11), Tipo_proceso (x12) |
| Factor 5 | Naturaleza de contrato | Q_contrat (x13), Mont_contrat_N (x14), T_contrat (x15) |
| Factor 6 | Confianza | Prorroga (x16), Mont_contrat _Cat(x17), Contrat_compartidos (x18) |

Nota: Los códigos señalados en color ● corresponde a la simbología utilizada en el análisis para facilitar el análisis

De acuerdo con la Tabla 6, se presenta el fundamento teórico construido y analizado de forma previa, donde cada factor o constructo corresponde a conformación de variables latentes, idóneas a lo que se desea medir (desempeño de proveedores) de acuerdo al estudio teórico realizado previamente y adecuado a los datos disponibles en la entidad bancaria. En la sección 6 se puede recordar la conceptualización de cada constructo (Factor 1, Factor 2, Factor 3, Factor 4, Factor 6), y en la sección 8.1 la construcción para el indicador Naturaleza de contrato (Factor 5).

Dado que los constructos ya fueron previamente explorados, en este caso no es necesario la especificación descriptiva de las variables unitarias, no obstante, se puede observar en el ANEXO 3.

7.2.2 Análisis Factorial confirmatorio 1: Primera versión del análisis con todas las variables

El análisis fue realizado en el software estadístico R versión 4.2.2 (R Core Team, 2022), con los paquetes “Lavaan” versión 0.6-12 (Rosseel, 2012) y “Pysch” versión 2.2.9 (Revelle, 2022). Al utilizar estas herramientas, el primer paso luego de la concepción teórica es la especificación del modelo de acuerdo con la Tabla 6, el cual se puede visualizar en la Figura 18

Figura 18

Especificación del modelo factorial confirmatorio con el paquete estadístico "Lavaan"

```
model <- " resp_legal =~ x1 + x2 + x3+ x4
          resp_social =~ x5 + x6 + x7
          seg_econ =~ x8 + x9
          cump_servicio =~ x10 + x11 +x12
          nat_contrato =~ x13 + x14 + x15
          confianza =~ x16 + x17 + x18"
```

En los Análisis Factoriales Confirmatorios, existen múltiples medidas de ajuste utilizadas para evaluar si la estructura teórica propuesta es la adecuada, no obstante, en el presente estudio se utilizaron los de uso más frecuente, los cuales se visualizan en la Tabla 7.

El primer indicador de ajuste considerado es el CFI, (índice de ajuste comparativo), este valor sugiere si el modelo presenta un buen ajuste a valores próximos a 1. El TLI, es el índice de Tucker Lewis o índice de ajuste no normalizado, este valor realiza una comparación entre el ajuste de los grados de libertad del modelo que se propone, contra un modelo nulo, es decir una ausencia de relación entre las variables involucradas. El RMSEA es la raíz cuadrada del error cuadrático medio, el cual es una medición promedio de los errores al cuadrado, por último, el SMMR, es la raíz cuadrada media residual estandarizada, e indica que cuanto más tenga una cercanía a valores de 0, mejor es el ajuste de los datos (Montalvo, 2014).

Tabla 7*Medidas de ajuste para un AFC*

| Medida de ajuste | Valor teórico esperado |
|------------------|------------------------|
| CFI | $\geq .95$ |
| TLI | $\geq .95$ |
| RMSEA | $\leq .05$ |
| SRMR | $\leq .06$ |

Una vez estructurado el modelo y definido los valores de ajuste a considerar, se procedió a realizar el análisis⁴. En la Figura 19, se puede ver la información general del modelo donde se visualiza que se trabajaron 551 datos (number of observation), y el p-valor de la prueba Chi cuadrado para bondad de ajuste, rechaza la hipótesis nula, es decir que inicialmente los datos no muestra un buen ajuste, sin embargo es importante añadir que en muchos casos, esta hipótesis se rechaza por la cantidad de datos y no por que necesariamente no haya un ajuste adecuado, por esta razón es que se utilizan otro tipo de indicadores para evaluar el modelo.

Figura 19*AFC 1: parte 1, especificaciones generales del modelo*

```
lavaan 0.6-12 ended normally after 313 iterations

Estimator                      ML
Optimization method             NLMINB
Number of model parameters      51

Number of observations          551

Model Test User Model:

Test statistic                   154.194
Degrees of freedom               120
P-value (Chi-square)            0.019

Model Test Baseline Model:

Test statistic                   1224.562
Degrees of freedom               153
P-value                          0.000
```

⁴ Para las figuras utilizadas en la sección de AFC, la colorimetría se interpreta según la escala:



En la Figura 20, se pueden visualizar los valores de ajuste especificados en la Tabla 4, donde todos cumplen los parámetros esperados, sin embargo, el valor TLI, se encuentra en límite de valor deseado.

Figura 20

AFC 1: parte 2, medidas de ajuste del modelo

| User Model versus Baseline Model: | |
|--|------------|
| Comparative Fit Index (CFI) | 0.968 |
| Tucker-Lewis Index (TLI) | 0.959 |
| Loglikelihood and Information Criteria: | |
| Loglikelihood user model (H0) | 6359.272 |
| Loglikelihood unrestricted model (H1) | 6436.369 |
| Akaike (AIC) | -12616.544 |
| Bayesian (BIC) | -12396.646 |
| Sample-size adjusted Bayesian (BIC) | -12558.542 |
| Root Mean Square Error of Approximation: | |
| RMSEA | 0.023 |
| 90 Percent confidence interval - lower | 0.010 |
| 90 Percent confidence interval - upper | 0.033 |
| P-value RMSEA <= 0.05 | 1.000 |
| Standardized Root Mean Square Residual: | |
| SRMR | 0.037 |

Una vez evaluados los parámetros de ajuste del modelo, es necesario observar los valores estimados de cada variable en su correspondiente factor, así como su p-valor. En la figura 21, se observa que las variables x3 y x12, no son significativas ⁵ para su constructo (se rechaza h0), además de que sus valores estimados estandarizados para el modelo son muy bajos, situación que ocurre con x7 y x18.

⁵ Bajo un nivel de significancia del 0.05

Figura 21

AFC 1: parte 3, parámetros del modelo

| Parameter Estimates: | | | | | | |
|----------------------|----------------------------------|---------|---------|------------|--------|---------|
| | Standard errors | | | Standard | | |
| | Information | | | Expected | | |
| | Information saturated (h1) model | | | Structured | | |
| Latent Variables: | | | | | | |
| | Estimate | Std.Err | z-value | P(> z) | Std.lv | Std.all |
| resp_legal =~ | | | | | | |
| x1 | 1.000 | | | | 0.127 | 0.600 |
| x2 | 0.240 | 0.032 | 7.387 | 0.000 | 0.030 | 0.725 |
| x3 | 0.002 | 0.007 | 0.312 | 0.755 | 0.000 | 0.016 |
| x4 | 0.207 | 0.034 | 6.015 | 0.000 | 0.026 | 0.348 |
| resp_social =~ | | | | | | |
| x5 | 1.000 | | | | 0.072 | 0.750 |
| x6 | 1.201 | 0.136 | 8.844 | 0.000 | 0.087 | 0.932 |
| x7 | 0.028 | 0.012 | 2.340 | 0.019 | 0.002 | 0.106 |
| seg_econ =~ | | | | | | |
| x8 | 1.000 | | | | 0.112 | 0.682 |
| x9 | 1.158 | 0.217 | 5.342 | 0.000 | 0.130 | 0.390 |
| cump_servicio =~ | | | | | | |
| x10 | 1.000 | | | | 0.066 | 0.427 |
| x11 | 0.476 | 0.172 | 2.770 | 0.006 | 0.031 | 0.466 |
| x12 | -0.002 | 0.111 | -0.015 | 0.988 | -0.000 | -0.001 |
| nat_contrato =~ | | | | | | |
| x13 | 1.000 | | | | 0.326 | 0.847 |
| x14 | 0.580 | 0.077 | 7.499 | 0.000 | 0.189 | 0.525 |
| x15 | 0.308 | 0.041 | 7.584 | 0.000 | 0.100 | 0.552 |
| confianza =~ | | | | | | |
| x16 | 1.000 | | | | 0.235 | 0.483 |
| x17 | 1.027 | 0.333 | 3.088 | 0.002 | 0.242 | 0.671 |
| x18 | 0.377 | 0.109 | 3.443 | 0.001 | 0.089 | 0.225 |

Después de los resultados obtenidos en la Figura 21, se realizó un Gráfico estructural del modelo para tener una mejor visualización, e identificar la estructura teórica del modelo, el gráfico se hizo utilizando el paquete estadístico “semTools” versión 0.5-6 (Terrence et al., 2022).

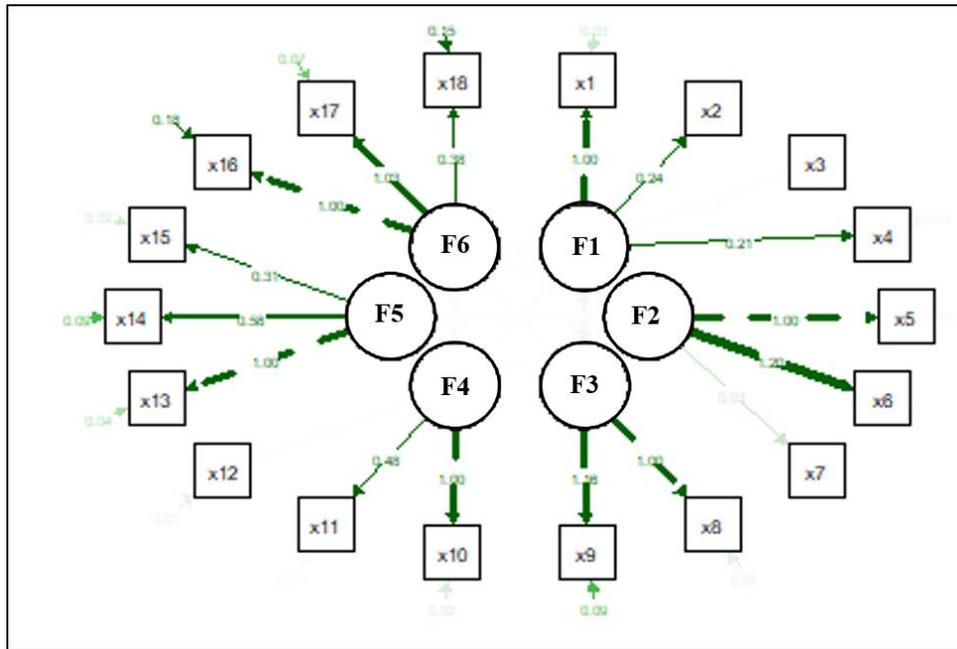
En la Figura 22 se puede observar la estructura que sugiere el análisis, donde teóricamente, tiene una composición bastante definida y entendible en términos del problema, ya que se ve claramente dos grandes grupos, los cuales se pueden interpretar como dos dimensiones, dado que por una lado se tiene a los factores 1, 2 y 3, los cuales corresponden a responsabilidad legal, social y seguridad económica, respectivamente y según la teoría estos aspectos hacen relación a una dimensión reputacional (García et al., 2012). Por otra parte, las factores 3, 4 y 5, que se refieren a naturaleza de contrato, cumplimiento de servicio y confianza, son aspectos totalmente laborales, por lo cual se puede interpretar como una dimensión de desempeño laboral, entonces se puede

concluir que los 6 constructos definen dos grandes dimensiones: Reputacional y Laboral, lo cual constituyen conceptualmente un score bastante Robusto en cuanto a una calificación integral de un proveedor.

En cuanto a los aspectos estadísticos de la estructura mostrada, se puede observar que la variable x3, perteneciente al factor 1(F1), no tiene línea de relación, lo cual indica su poco o nulo aporte al constructo, revisando la definición con los expertos, se determinó que la variable, la cual representa los individuos que se encuentran en lista negra judicial, tiene muy pocos valores, lo cual puede ocasionar este resultado, además de que si bien, tiene importancia, es una variable que cuesta realizar actualizaciones. Una situación similar presenta la variable x7, en el factor 2. Además, la variable x18 del factor 6(F6), si bien su p-valor es significativo, su aporte al constructo es bajo con respecto a los otros dos; por lo cual se decidió realizar de nuevo el análisis sin estas variables en el factor 1, factor 2 y factor 6 respectivamente.

Figura 22

AFC 1: Gráfico estructural del modelo



7.2.3 Análisis factorial confirmatorio 2: mejora del modelo obtenido en AFC 1

De la misma forma que se procedió en la sección anterior, primero se definió el nuevo modelo (VER ANEXO 3) y se obtuvo las especificaciones generales del modelo, donde se puede visualizar, que en esta ocasión la prueba Chi cuadrado de bondad de ajuste no rechaza H0, por lo que en primera instancia esta nueva estructura presentó un buen ajuste de los datos.

Figura 23

AFC 2: parte 1, especificaciones generales del modelo

```
lavaan 0.6-12 ended normally after 209 iterations
```

| | |
|----------------------------|----------|
| Estimator | ML |
| Optimization method | NLMINB |
| Number of model parameters | 43 |
| Number of observations | 551 |
| Model Test User Model: | |
| Test statistic | 76.667 |
| Degrees of freedom | 62 |
| P-value (Chi-square) | 0.100 |
| Model Test Baseline Model: | |
| Test statistic | 1126.984 |
| Degrees of freedom | 91 |
| P-value | 0.000 |

Con respecto a los indicadores de ajuste evaluados, en la Figura 24, se puede observar que todos los parámetros, además de cumplir con los valores deseados, mejoraron con respecto a los obtenidos en el AFC 1.

Figura 24

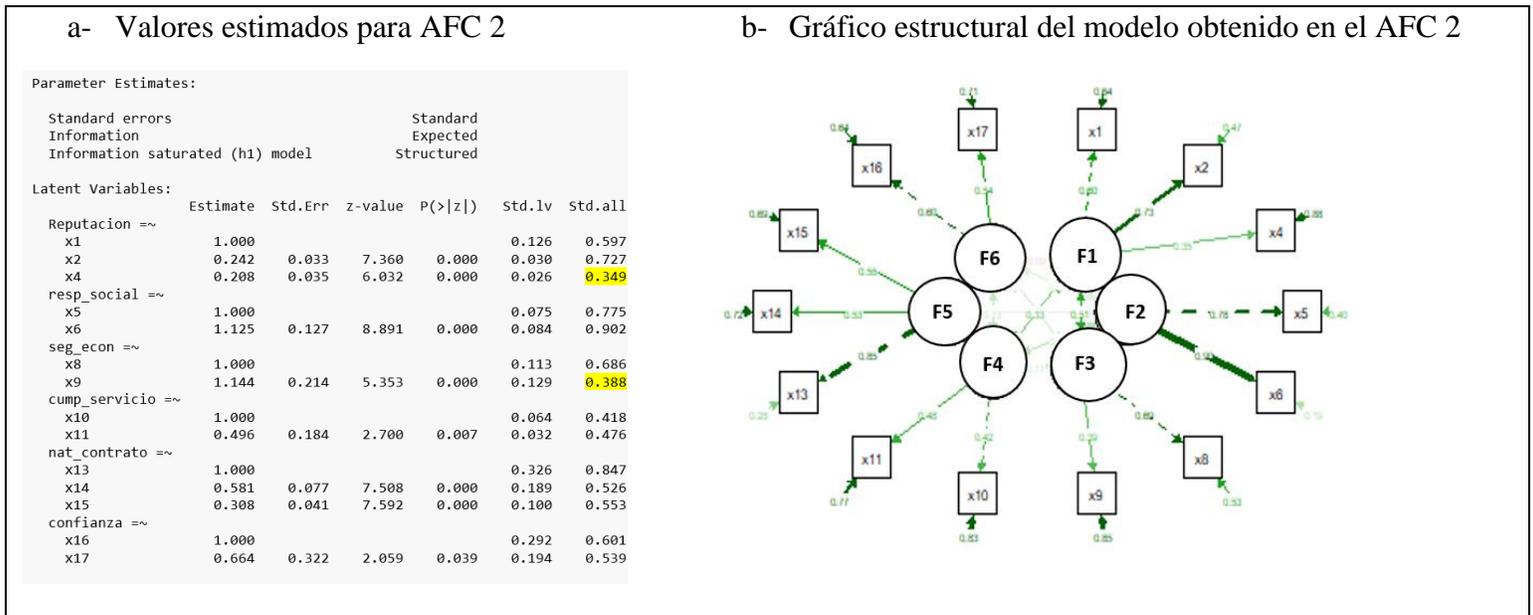
AFC 2: parte 2, medidas de ajuste del modelo

```
User Model versus Baseline Model:
```

| | |
|--|-----------|
| Comparative Fit Index (CFI) | 0.986 |
| Tucker-Lewis Index (TLI) | 0.979 |
| Loglikelihood and Information Criteria: | |
| Loglikelihood user model (H0) | 3314.931 |
| Loglikelihood unrestricted model (H1) | 3353.265 |
| Akaike (AIC) | -6543.862 |
| Bayesian (BIC) | -6358.458 |
| Sample-size adjusted Bayesian (BIC) | -6494.958 |
| Root Mean Square Error of Approximation: | |
| RMSEA | 0.021 |
| 90 Percent confidence interval - lower | 0.000 |
| 90 Percent confidence interval - upper | 0.035 |
| P-value RMSEA <= 0.05 | 1.000 |
| Standardized Root Mean Square Residual: | |
| SRMR | 0.029 |

Por último, en la Figura 25, se puede apreciar, que todas las variables es sus constructos tiene cargas deseables, y todas son significativas (a), y el modelo conceptualmente mantiene su estructura y presenta visualmente una mejor estructura (b). Por lo que se definió este último proceso como el análisis factorial confirmatorio final.

Figura 25
AFC2: Valores estimados y estructura de modelo



Teniendo una estructura teórica definida y los seis constructos validados, se procedió a calcular una puntuación general, basado en los valores de cada constructo, no obstante, fue necesario evaluar el “peso” que debía tener cada constructo, en el score final, para esto se utilizaron las cargas factoriales finales como ponderadores, En la tabla 8, se puede visualizar que el constructo de Responsabilidad legal, es el de mayor ponderación y Responsabilidad social el menor.

Tabla 8

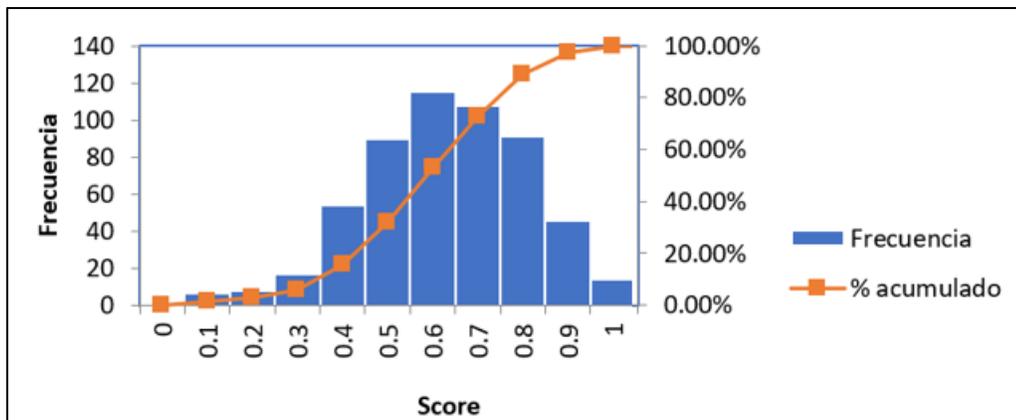
Pesos para cada constructo en la conformación del score final del indicador de evaluación a proveedores bancarios

| CONSTRUCTO | PESO V1 | PESO V2 | PESO V3 | TOTAL | RESCALAMIENTO DE VALORES |
|--------------------------|---------|---------|---------|-------|--------------------------|
| RESPONSABILIDAD LEGAL | 0.88 | 0.47 | 0.64 | 1.99 | 0.23 |
| CUMPLIMIENTO DE SERVICIO | 0.83 | 0.77 | | 1.6 | 0.19 |
| NATU_CONTRATO | 0.26 | 0.72 | 0.69 | 1.67 | 0.19 |
| CONFIANZA | 0.71 | 0.64 | | 1.35 | 0.16 |
| SEGURIDAD_ECONOMICA | 0.53 | 0.85 | | 1.38 | 0.16 |
| RESPONSABILIDAD SOCIAL | 0.4 | 0.19 | | 0.59 | 0.07 |

Por último, se puede visualizar en la figura 26, la distribución del “score final”, el cual oscila entre valores de 0 a 1, y tiene una forma asimétrica positiva, lo que indica mayor concentración en los puntajes altos.

Figura 26

Distribución de frecuencias del "score" final obtenido para evaluar desempeño de proveedores bancarios

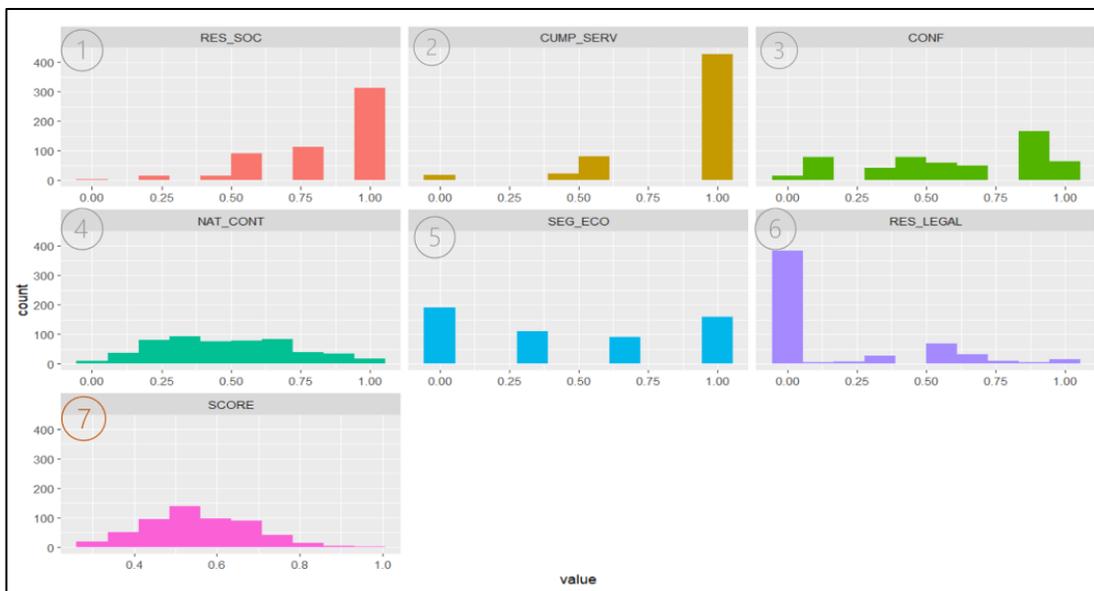


7.3 Análisis de conglomerados

Con la finalidad de categorizar a los proveedores, luego de la construcción de índices y realizar el AFC, se realizó un análisis de conglomerados o como comúnmente se le conoce “clusterización”. Se utilizaron los seis constructos realizados y el Score final obtenido. En la figura 27 se puede observar la distribución de cada variable, donde es observable que el constructo cumplimiento de servicio (2), es el que muestra una distribución más discreta a lo largo de las posibles puntuaciones (0 a 1)

Figura 27

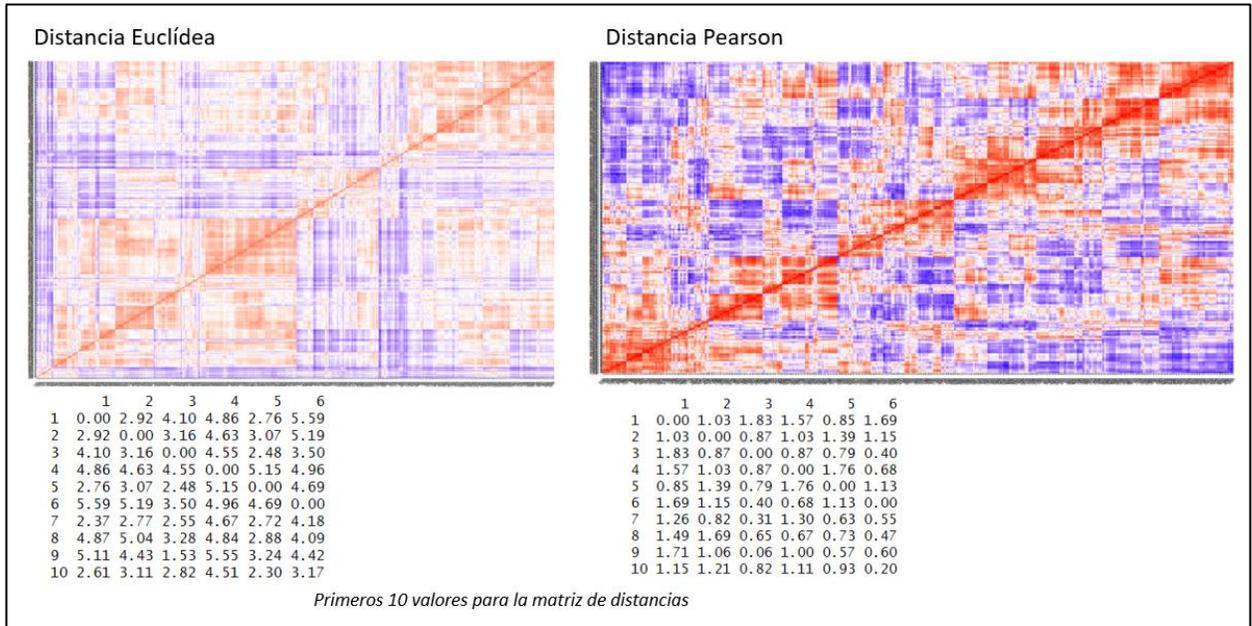
Histogramas de constructos para la evaluación del desempeño de proveedores



Para realizar el análisis se utilizó los paquetes del software estadístico R “cluster” (Maechler et al., 2022) y “Ggplot2” (Wickham, 2016). El primer paso, fue indagar el tipo de distancia a utilizar, donde se visualiza que la distancia Euclídea es la que muestra una mejor distribución para los datos (figura 28), por lo que en adelante se utilizó esta forma de obtener las distancias.

Figura 28

Tipos de distancias aplicadas en el Análisis de conglomerados



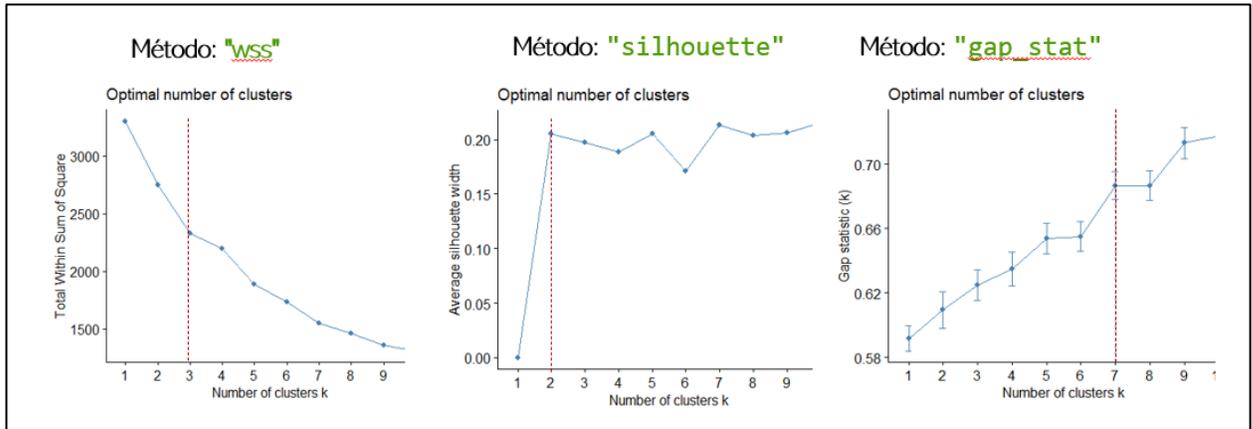
El siguiente paso, fue determinar el número idóneo de los grupos o cluster a obtener. De forma orientativa se utilizaron tres métodos el “wss” o suma de cuadrados internos; es el método más utilizado y busca seleccionar la cantidad ideal de grupos optimizando la suma de cuadrados (Ayala, 2020), con este método se sugieren 3 grupos (figura 28)

Otro método indagado fue el “silhouette” o método silueta, cuyo valor indica que similitud tiene un elemento a su propio cúmulo, en comparación a la separación que presenta a otros objetos (Rousseeuw, 1987). Esta medida indicó que el número idóneo es de dos grupos.

Por último, se estudió la medida “Gap statistic”, el cual es una estandarización del gráfico wss (Kassambara. A y Mundt, 2020), con este método se dio una sugerencia de 7 grupos.

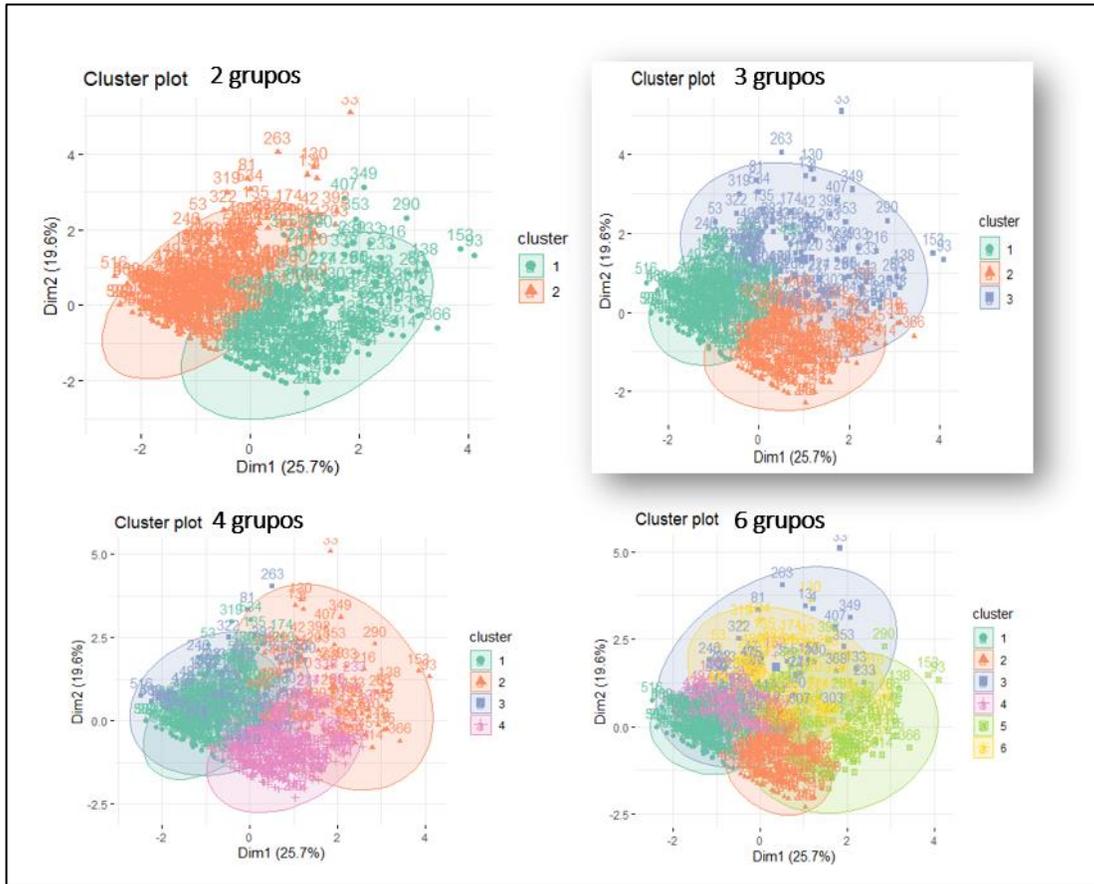
Figura 29

Métodos para seleccionar tamaños de grupos o cluster



Con los resultados obtenidos en la Figura 29, se procedió a explorar cual era la mejor agrupación (mayor detalle en ANEXO 4), donde para cuatro y seis grupos es visible que no se muestra una clara definición y dificultaba establecer perfiles de los proveedores, por el contrario con dos grupos se visualizó una definición más definida, no obstante, teóricamente fue difícil identificar perfiles, situación que no ocurrió con tres grupos, donde sí se pudo identificar una agrupación más precisa. En la Figura 30, se muestra una exploración más a fondo de la conformación de tres cluster.

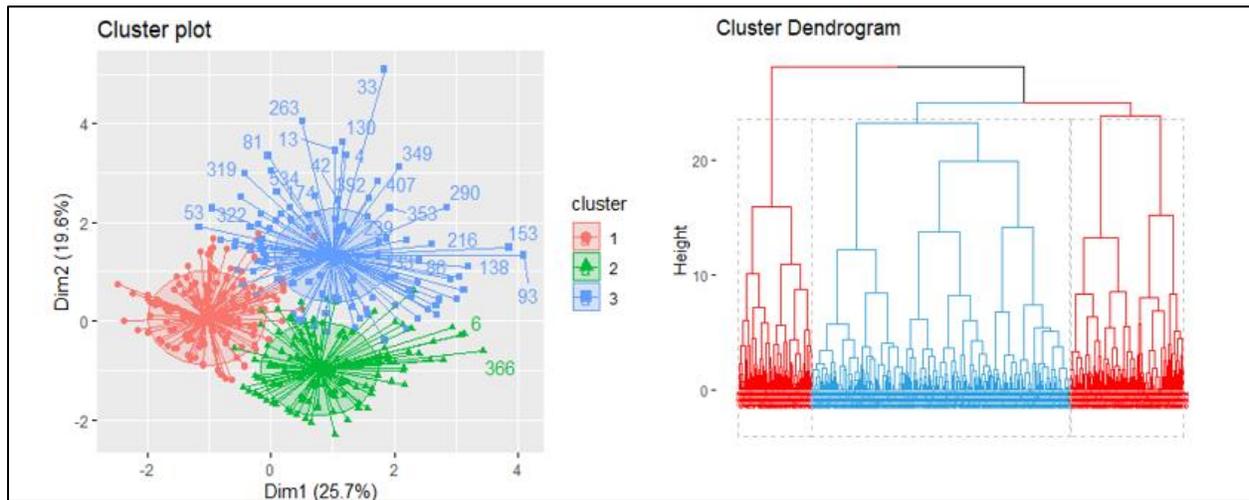
Figura 30
Posibles agrupaciones para categorizar proveedores



Con tres grupos es identificable perfiles claros (Figura 30), a nivel estadístico se visualiza como las distancias calculadas para los individuos de la población realizan una definición de grupos donde no hay una traslape ni confusión en la identificación de los grupos.

Figura 31

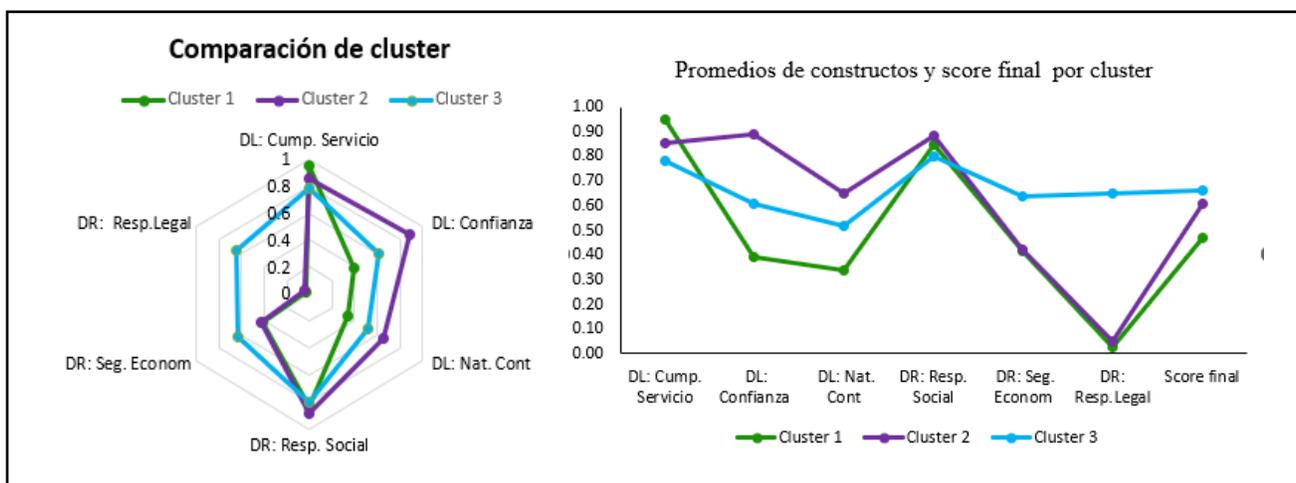
Visualización de tres grupos de proveedores



A nivel conceptual, se puede visualizar en la Figura 25 que el clúster 1, no presenta una definición clara en alguna dimensión en específico, ya que sus valores más altos se encuentran en los constructos de Cumplimiento de servicio y Responsabilidad social, el resto de los constructos tiene valores bajos y el promedio del Score final es el más bajo de los tres. Para el clúster 2, presenta valores promedios más altos en la Dimensión laboral que en la dimensión reputacional, mientras que el clúster 3 es el grupo más ecuánime en todos los constructos, ya que no son valores ni muy altos ni muy bajos, pero mantiene un promedio similar en todos los constructos y es el posee el score más alto de los tres clústeres.

Figura 32

Descripción de clúster's



8. Resultados

Con la aplicación de técnicas estadísticas multivariadas, como lo fueron análisis factorial exploratorio, análisis factorial confirmatorio y análisis de conglomerados, se logró establecer, en primera instancia 6 constructos o variables latentes, las cuales funcionan como indicadores por si solos ya que todas fueron escaladas en valores de 0 a 1 que funcionan como puntuaciones, donde el valor menor indica el peor desempeño en el constructo indicado y el mayor valor el mejor desempeño. En la tabla 9 se muestra la descripción y construcción final de cada indicador.

Tabla 9

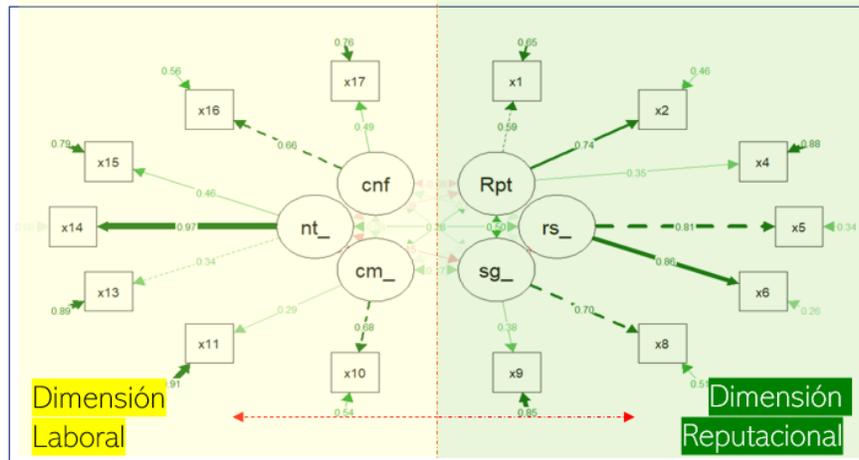
Descripción final de variables latentes construidas con técnicas multivariantes

| Constructo | VARIABLES UNITARIAS FINALES UTILIZADAS PARA ESTIMACIÓN | Objetivo de medición |
|--------------------------|--|--|
| Responsabilidad Legal | Embargos Juicios Referencias comerciales | Mide el encontrarse libre de acciones legales que puedan causar una mancha legal en el expediente |
| Responsabilidad Social | Deudas a la CCSS Deudas en Hacienda | Mide el cumplimiento con los empleados a través de obligaciones patronales y estatutos de empresa |
| Seguridad Económica | Tenencia de bienes muebles Tenencia de bienes inmuebles | Mide la solvencia de la empresa a través de tenencia de bienes materiales |
| Cumplimiento de servicio | Cantidad de sanciones Cantidad de multas | Mide el desenvolvimiento del proveedor al no presentar sanciones o multas en su desempeño laboral |
| Naturaleza de contrato | Monto de contrato Tipo de contrato Cantidad de contratos | Mide aspectos administrativos de contratación |
| Confianza | Monto de contrato Prorroga | Mide la confianza de la entidad bancaria hacia el proveedor en la medida de obtener una prórroga de trabajo en relación al monto contratado. |

Otro resultado de gran importancia fue, la tipificación de aspectos o dimensiones sociales que evidencian una evaluación integral de los proveedores contratados, en la figura 33 se muestra la identificación de una dimensión laboral y una dimensión reputacional, esta relación se pudo establecer gracias a los aspectos conceptuales identificados en la teoría.

Figura 33

Identificación de dimensiones para los 6 constructos creados para la evaluación de proveedores bancarios.



De acuerdo con la figura 33, se visualiza como el AFC además de estructurar la mejor manera de componer las variables latentes identificadas, propone una estructura que teóricamente corresponde a las dimensiones mencionadas.

Finalmente con la técnica de conglomerados fue posible la creación de tres perfiles de proveedores, que no solo permiten realizar una identificación de desempeño de los mismo, sino que fue posible realizar una relación a escalas estandarizadas internacionalmente, ampliamente utilizadas en ámbitos empresariales como lo es la escala de Barret o más conocida como niveles de conciencia, la cual a nivel empresarial permite identificar grupos que ofrecen distintos niveles de desempeño (Robles, 2015), en la Figura 34, se puede visualizar los 7 niveles propuestos Richard Barret, donde propone una estructura distinta a la típica pirámide, por el contrario, en esta teoría se asume que se puede empezar una zona de confort (niveles 1,2 y 3), llegar a un punto de equilibrio, donde empieza una transformación (nivel 4) para llegar a una zona de servicio (niveles 5, 6 y 7).

Figura 34
Modelo de Richart Barret



Nota: Tomado de Robles 2015. URL: <https://domo.es/negocios-solidos>

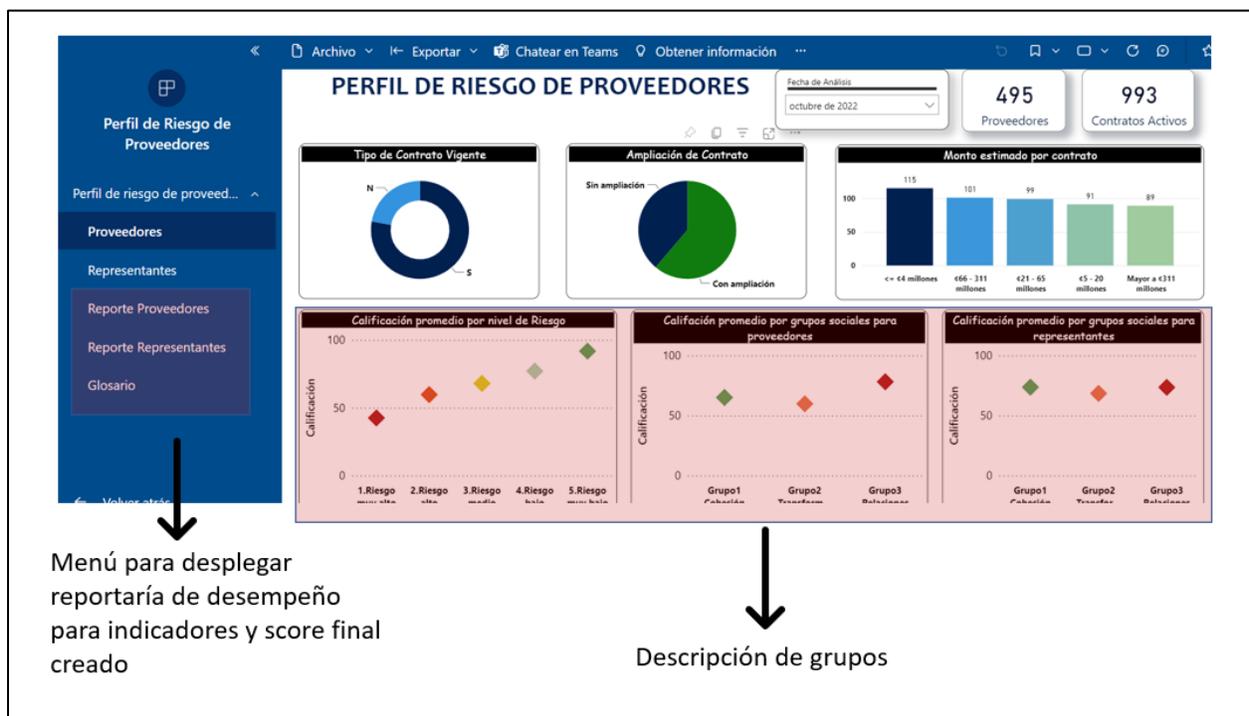
De acuerdo a la figura 34, y con los resultados obtenidos para el indicador y los constructos creados, es posible identificar los tres grupos o clúster en distintos niveles de la escala de Barret, donde el clúster 1, es un grupo con un alto cumplimiento de servicio y responsabilidad social lo cual se asocia a la satisfacción del cliente, no obstante hay una oportunidad de mejora en la confianza, estas características hacen que el grupo se asocie al nivel 2 llamado “Relaciones”.

El clúster 2, en comparación al 1, mejora el indicador de confianza y naturaleza de contrato, lo cual es un indicativo de responsabilidad y adaptabilidad, esto hace al grupo nivel 4 de la escala de Barret, llamado “transformación”. Por último, el clúster 3 es un grupo que tiene valores medios en la mayoría de los constructos sin llegar a la excelencia en ninguno, no obstante recordando que las variables latentes construidas se asocian con aspectos reputacionales y laborales se puede decir que tienen un desempeño medio en ambas dimensiones y por lo tanto, se asocia la construcción de la confianza, cooperación y compromiso con la entidad contratante, esto hace que el grupo se ubique en el nivel 5, llamado “Cohesión interna”. Esta asociación a la escala de Barret le permite a la entidad bancaria identificar fortalezas y áreas de mejora de sus proveedores y poder ver crecimiento o no, en el transcurso del tiempo.

Tabla 10*Asociación de Clúster con Niveles de Escala Barret*

| Clúster | DL: Cump. Servicio | DL: Confianza | DL: Nat. Cont | DR: Resp. Social | DR: Seg. Econom | DR: Resp. Legal | Score final | Asociación escala de Barret |
|------------------|--------------------|---------------|---------------|------------------|-----------------|-----------------|-------------|-----------------------------|
| 1 | 0.95 | 0.39 | 0.34 | 0.85 | 0.41 | 0.03 | 0.47 | Nivel 2 |
| 2 | 0.85 | 0.89 | 0.65 | 0.89 | 0.42 | 0.05 | 0.61 | Nivel 4 |
| 3 | 0.78 | 0.61 | 0.52 | 0.80 | 0.64 | 0.65 | 0.66 | Nivel 5 |
| Promedio general | 0.88 | 0.60 | 0.48 | 0.85 | 0.47 | 0.17 | 0.55 | |

Por último, con la finalidad de poder dar una comunicación efectiva de los resultados a los usuarios finales, se creó una Dashboard con el software de Microsoft PowerBi, el cual se visualiza en la figura 35.

Figura 35*Ejemplo de visualización de información Power BI*

9. Bibliografía

- Amat, J. (09 de 2017). *Clustering y heatmaps: aprendizaje no supervisado*. Retrieved 01 de 09 de 2022, from rpubs: https://rpubs.com/Joaquin_AR/310338
- Arroyo, M. d., & Ramos, J. A. (2014). Proceso para el diseño de un sistema de evaluación y selección de proveedores. *XIX Congreso Internacional de Contaduría, Administración e Informática*. México, D.F.
- Austria, M. A. (2007). Construcción de índices de desempeño financiero mediante el Análisis de Componentes Principales, 2004-2005. *Análisis Económico*, XXII(51). <https://www.redalyc.org/articulo.oa?id=41311486011>
- Ayala, J. (2020). *Minería de datos*. Métodos de particionamiento: <https://rpubs.com/JairoAyala>
- Bolaños, L. (10 de marzo de 2020). *Rpubs*. Análisis Factorial: <https://rpubs.com/>
- Chávez, M. (13 de Febrero de 2017). *Introducción a los Métodos Multivariantes*. rpubs: https://rpubs.com/marcelo-chavez/multivariado_1
- Churchill, G. (2008). Churchill's Paradigm for Better Measures. *Rasch Measurement Transactions*, 22(2).
- Conexión Esan. (05 de marzo de 2019). *ESAN Graduate School of Business*. ¿Con qué criterios debes evaluar a tus proveedores?: <https://www.esan.edu.pe/conexion-esan/con-que-criterios-debes-evaluar-a-tus-proveedores>
- De la fuente, S. (2011). *Análisis Factorial*. Universidad Autónoma de Madrid.
- Equipo ORCA. (15 de abril de 2021). *Riesgo de proveedores, ¿por qué deberías evaluarlos?* Equipo ORCA : <https://blog.orcagrc.com/>
- Escobar, R. M., Blanco, S. A., & Cruz, F. S. (2011). *Metodología para la identificación y evaluación del riesgo que amenaza el negocio en marcha, en empresas donde exista concentración de proveedores*. Retrieved 24 de 8 de 2022, from <http://ri.ues.edu.sv/id/eprint/11400>
- Estamática. (26 de enero de 2017). *Rotación Varimax*. Retrieved 23 de octubre de 2022, from Análisis Factorial con rotación Varimax: <https://estamatica.net/analisis-factorial-malaga/>
- García, J. L., Alvarado, A., & Maldonado, A. (01 de 02 de 2012). *Selección de proveedores basada en análisis dimensional*. Scielo Org.: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0186-10422013000300011

- Gimeno, F. (2005). *Indicadores de rendimiento y evaluación al proveedor*. AENOR.
- Gutiérrez, R., González, A., Torres, F., & Gallardo, J. (1994). Tema 1. Análisis Factorial. En *Técnicas de Análisis de datos multivariable. Tratamiento computacional* (págs. 1-79). UGR.
- Gutiérrez, R., González, A., Torres, F., & Gallardo, J. (1994). Capítulo 6. Introducción al Análisis Cluster. Consideraciones generales. En *Técnicas de Análisis de datos multivariable. Tratamiento computacional*.
- Hernández, S., Fernández, S., & Baptista, P. (2006). *Metodología de la investigación, Cuarta Edición*. McGraw Hill.
- Jordan, F. (2021). Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *Revista de investigación en Psicología Social*, 7(1).
- Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of CRAN.R-project.org/package=factoextra*
- Lai, K. (2021). Fit Difference Between Nonnested Models Given Categorical Data: Measures and Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, 28, 99-120.
- López, J. C., Calderón, C. A., & Gómez, J. C. (2019). *Metodología multicriterio para la selección de proveedores bajo consideraciones de riesgo*. Retrieved 24 de 8 de 2022, from <http://revistas.utp.edu.co/index.php/revistaciencia/article/view/19681>
- Maechler, M., Rousseeuw, P., Struyf, A., & Hubert, M. (2022). *Cluster: Cluster Analysis Basics*. <https://CRAN.R-project.org/package=cluster>
- Martín, J. L. (2013). *Análisis Económico-financiero en el sector de componentes del automóvil [Tesis de Maestría]*. Universidad de Valladolid, Repositorio Documental: <https://uvadoc.uva.es/handle/10324/4507>
- Matín, E. (01 de 09 de 2022). *Como puedo evaluar eficazmente el cumplimiento de un proveedor*. Cibernos Grupo: <https://www.grupocibernos.com/>
- Méndez, C., & Rondón, M. (2012). Introducción al análisis factorial exploratorio. *Revista Colombiana de Psiquiatría*, 41(1), 197-207.
- Montalvo, M. (mayo de 2014). *Análisis Estadístico Multivariante*. Retrieved 01 de 07 de 2022, from Modelo de desarrollo económico para la diversificación de la estructura productiva y la articulación

del tejido empresarial en Baja California del Sur. Tesis Doctoral: <https://www.eumed.net/tesis-doctorales>

Montero, E. (2008). ESCALAS O ÍNDICES PARA LA MEDICIÓN DE CONSTRUCTOS: EL DILEMA DEL ANALISTA DE DATOS. *Avances en Medición*, 6, 15-24.

Ortiz, M., Márquez, F., Oramas, O., & Marrero, Y. (2018). Metodología para la evaluación de proveedores. Caso de estudio: Empresa Especializada Importadora, Exportadora y Distribuidora para la Ciencia y la Técnica (EMIDICT). *Revista Espacios*, 39(27), 24. <https://doi.org/ISSN 0798 1015>

Parra, C., Osorio, J. C., & Escandón, J. C. (2019). Metodología multicriterio para la selección de proveedores bajo consideraciones de riesgo. *Scientia et Technica*, 24(2). <https://doi.org/ISSN 0122-701>

Pedraza, E., Amaya, G., & Conde, M. (2010). Desempeño laboral y estabilidad del personal administrativo contratado de la Facultad de Medicina de la Universidad del Zulia. *Revista de Ciencias Sociales*, 16(3). Retrieved 01 de 09 de 2022, from http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1315-95182010000300010

Peña, B. (22 de 11 de 2018). *Una guía amigable para el uso de Lavaan*. Retrieved 2022 de 09 de 15, from Potencia de R para el análisis factorial confirmatorio: <https://brianmsm.github.io/manual-lavaan/index.html>

Pérez, E., & Medrán, L. (2010). Análisis Factorial Exploratorio: Bases Conceptuales y Metodológicas. *Revista Argentina de Ciencias del Comportamiento*, 2(1), 58-66.

R Core Team. (2022). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>

Real Academia Española. (s.f.). *Calidad*, versión 23.6 en línea. Retrieved 01 de 09 de 2022, from En Diccionario de la lengua española: <https://dle.rae.es/calidad>

Real Academia Española. (s.f.). *Reputación*, versión 23.6 en línea. Retrieved 1 de 10 de 2022, from En Diccionario de la lengua española: <https://dle.rae.es/reputaci%C3%B3n>

Restrepo, R. (2016). *Evaluación y clasificación de proveedores: metodología y caso de estudio en una ensambladora de motocicletas*. [Tesis de Maestría], Universidad EAFIT, Departamento de Ingeniería de Producción, Medellín. <https://repository.eafit.edu.co/handle/10784/11456>

- Revelle, W. (2022). *Psych: Procedures for Personality and Psychological Research*. Evanston, Illinois, USA: Northwestern University. <https://CRAN.R-project.org/package=psych> Version = 2.2.9.
- Robles, H. (09 de noviembre de 2015). *Domo blog*. Si quieres negocios Sólidos satisface necesidades profundas.: <https://domo.es/negocios-solidos>
- Rosell, Y. (2012). *lavaan: An R Package for Structural Equation Modeling*. (Vol. 48). Journal of Statistical Software. <https://doi.org/10.18637/jss.v048.i02>
- Rousseeuw, P. (1987). *Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis*.
- Terrence, D., Pornprasertmanit, S., Schoemann, A., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling*. <https://CRAN.R-project.org/package=semTools>
- Torres, V., Gallardo, M. R., Martínez, H., & Leyva, L. (Enero-Abril de 2007). Evaluación de la gestión de proveedores en la Universidad de Holguín. *Revista Electrónica Cooperación Universidad Sociedad*, 6(1), 56-63. <https://doi.org/ISSN 25298-8075>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>

10. Anexos

10.1 ANEXO 1: Recategorización de variables

1- Monto de contrato

| Etiqueta | Escala |
|---|---------------|
| si es menor igual a ¢ 4.053.136.9518 | 1 |
| si es menor igual a ¢ 20.083.145.7328 | 2 |
| si es menor igual a ¢ 65.554.487.46384 | 3 |
| si es menor igual a ¢ 311.523.809.30436 | 4 |
| si es mayor a ¢ 311.523.809.30436 | 5 |

2- Prórrogas

| Etiqueta | Escala |
|-------------------|---------------|
| no tiene prorroga | 0 |
| tienen prorroga | 1 |

3- Cantidad de contratos

| Etiqueta | Escala |
|-----------------|---------------|
| un contrato | 1 |
| 2 y 3 contratos | 2 |
| de 4 adelante | 3 |

4- CCSS_INDICE

| Etiqueta | Escala |
|---------------------|---|
| A-AD_1 | 1 activo al día |
| I-AD_NR-NR_blanco_2 | 3 activo cobro administrativo |
| A-CA_I-CA_3 | 2 inactivo al día |
| I-CJ_A-CJ_4 | 3 inactivo cobro administrativo |
| | 4 inactivo cobro judicial - activo cobro judicial |
| | 2 no registrado |
| | 2 blancos |

5- HACIENDA_INDICE

Composición:

Se recodifica la variable para asignar niveles de riesgo de menor a mayor en torno a sus deudas con la entidad. Agrupación de los niveles de la variable para formar una variable ordinal.

Constructo de las variables: se recodificó la variable original de menos riesgo a más riesgo

Se recodificó la variable original de menos riesgo a más riesgo :

B-ON-MN 1
B-OS-MS 3
D-ON-MN 1
D-ON-MS 4
D-OS-MN 4
D-OS-MS 4
I-ON-MN 1
I-ON-MS 2
I-OS-MN 2
I-OS-MS 3
N-ON-MN 1
(en blanco) 2

| Etiqueta | Escala |
|--|--------|
| Desincritos morosos y/o omisos y no inscrito | 4 |
| moroso y omisos | 3 |
| morosos o omisos | 2 |
| Al día y desincritos | 1 |

6- IMAS_INDICE

Composición:

se recodifica para asignar puntuaciones a quien está o no al día.

| | |
|---------------------------|---|
| moroso al día o en blanco | 2 |
| | 1 |

10.2 ANEXO 2: Análisis factorial exploratorio realizado en R

1.b Análisis exploratorio

Con grupo de variables que no se le veía en principio un sentido teórico se realizó un AFE

María del Rocío

8/9/2022

LIBRERIAS

```
library(scales)
library(dplyr)

## ## Attaching package: 'dplyr'

library(psych)

##
## Attaching package: 'psych'

library(polycor)

##
## Attaching package: 'polycor'

library(ggcorrplot)

##
## Attaching package: 'ggplot2'

library(openxlsx) #Librería que interactúa con MSEXcel
library(corrplot) #Librería para el gráfico de correlaciones

## corrplot 0.92 loaded

library(corr)
library(stats)
```

LECTURA DE DATOS

```
library(readxl)
base_explo_INT3 <- read_excel("base_explo_INT3v2.xlsx")
#attach(BASE)
BASE=base_explo_INT3
BASE=select(BASE, Q_CONTRAT:N_REP)
```

QUITAR VALORES FALTANTES

```
BASE = BASE[complete.cases(BASE),]
```

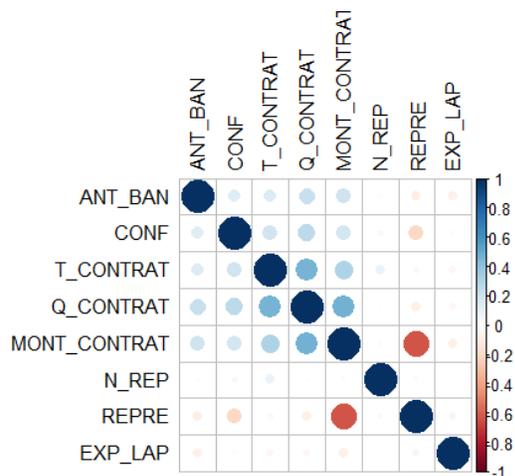
PRIMER ANALISIS MATRIZ DE CORRELACIONES

```
matriz_correlaciones <- cor(BASE, use = "pairwise.complete.obs")
matriz_correlaciones

##           Q_CONTRAT MONT_CONTRAT      CONF      REPRES      ANT_BAN
## Q_CONTRAT  1.0000000  0.47018127  0.26537324 -0.08865188  0.22397701
## MONT_CONTRAT 0.47018127  1.00000000  0.18952586 -0.62262822  0.20113176
## CONF        0.26537324  0.18952586  1.00000000 -0.20428551  0.13178320
```

```
## REPRE      -0.08865188  -0.62262822 -0.20428551  1.00000000 -0.09263432
## ANT_BAN    0.22397701   0.20113176  0.13178320 -0.09263432  1.00000000
## EXP_LAP   -0.04970022  -0.08689970 -0.01887758  0.05152926 -0.07921227
## T_CONTRAT 0.46708922   0.30567973  0.19567515 -0.03295765  0.14957660
## N_REP      0.01031587  -0.02763865 -0.03526763  0.02088214 -0.01231878
##           EXP_LAP  T_CONTRAT      N_REP
## Q_CONTRAT -0.0497002247  0.46708922  0.0103158675
## MONT_CONTRAT -0.0868996992  0.30567973 -0.0276386537
## CONF       -0.0188775795  0.19567515 -0.0352676329
## REPRE      0.0515292563 -0.03295765  0.0208821368
## ANT_BAN   -0.0792122677  0.14957660 -0.0123187806
## EXP_LAP    1.0000000000 -0.04254242  0.0005160852
## T_CONTRAT -0.0425424249  1.00000000  0.0807569175
## N_REP      0.0005160852  0.08075692  1.0000000000
```

```
corrplot(cor(BASE), order = "hclust", tl.col='black', tl.cex=1) #Gráfico de Las correlaciones
```



INVERTIR VARIABLES SE PUEDE VISUALIZAR QUE HAY VARIABLES EN UNA DIRECCIÓN CONTRARIA, POR LO CUAL SE DEBEN INVERTIR

```
BASE$EXP_LAP=with(BASE, max(EXP_LAP)-EXP_LAP)
BASE$REPRE=with(BASE, max(REPRE)-REPRE)
BASE$N_REP=with(BASE, max(N_REP)-N_REP)
```

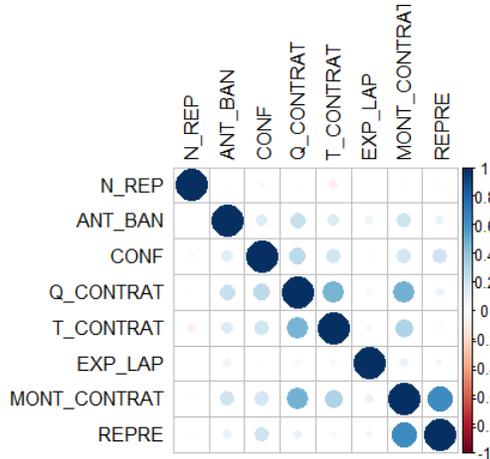
MATRIZ DE CORRELACIONES

```
matriz_correlaciones <- cor(BASE, use = "pairwise.complete.obs")
matriz_correlaciones
```

```
##           Q_CONTRAT MONT_CONTRAT      CONF      REPRE      ANT_BAN
## Q_CONTRAT  1.00000000  0.47018127  0.26537324  0.08865188  0.22397701
## MONT_CONTRAT 0.47018127  1.00000000  0.18952586  0.62262822  0.20113176
## CONF       0.26537324  0.18952586  1.00000000  0.20428551  0.13178320
## REPRE      0.08865188  0.62262822  0.20428551  1.00000000  0.09263432
## ANT_BAN    0.22397701  0.20113176  0.13178320  0.09263432  1.00000000
## EXP_LAP    0.04970022  0.08689970  0.01887758  0.05152926  0.07921227
## T_CONTRAT  0.46708922  0.30567973  0.19567515  0.03295765  0.14957660
## N_REP      -0.01031587  0.02763865  0.03526763  0.02088214  0.01231878
##           EXP_LAP  T_CONTRAT      N_REP
## Q_CONTRAT  0.0497002247  0.46708922 -0.0103158675
## MONT_CONTRAT 0.0868996992  0.30567973  0.0276386537
## CONF       0.0188775795  0.19567515  0.0352676329
## REPRE      0.0515292563  0.03295765  0.0208821368
```

```
## ANT_BAN      0.0792122677  0.14957660  0.0123187806
## EXP_LAP      1.0000000000  0.04254242  0.0005160852
## T_CONTRAT   0.0425424249  1.00000000  -0.0807569175
## N_REP       0.0005160852 -0.08075692  1.0000000000
```

```
corrplot(cor(BASE), order = "hclust", tl.col='black', tl.cex=1) #Gráfico de Las correlaciones
```



OTRA FORMA DE CORRELACIÓN

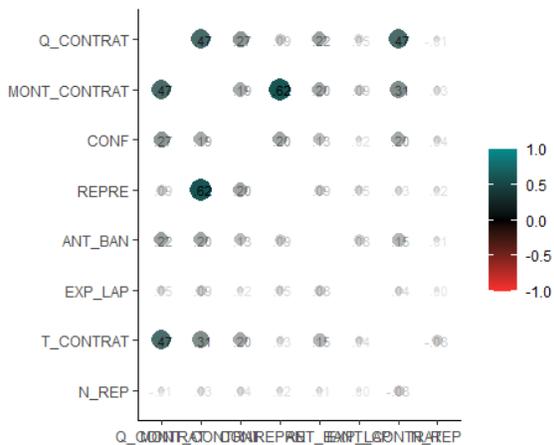
```
BASE_correlaciones <- correlate(BASE) #Cálculo de un objeto de correlaciones
```

```
##
## Correlation method: 'pearson'
## Missing treated using: 'pairwise.complete.obs'
```

```
rplot(BASE_correlaciones, legend = TRUE, colours = c("firebrick1", "black",
  "darkcyan"), print_cor = TRUE) #Opción gráfica de Las correlaciones
```

```
## Don't know how to automatically pick scale for object of type noquote. Defaulting to contin
uous.
```

```
## Don't know how to automatically pick scale for object of type noquote. Defaulting to contin
uous.
```



se puede ver muchas variables sin correlación

PRUEBAS QUE INDICAN SI ES CONVENIENTE REALIZAR EL ANÁLISIS

1- DETERMINANTE DE MATRIZ DE CORRELACIONES

```
det(matriz_correlaciones)
## [1] 0.2672193
```

El determinante es mayor a 0 lo cual es bueno, ya que implica baja multicolinealidad

2- TEST DE BARTLETT

```
bartlett.test(BASE)
##
## Bartlett test of homogeneity of variances
##
## data:  BASE
## Bartlett's K-squared = 543.47, df = 7, p-value < 2.2e-16
```

como se rechaza el p value, tomando H_0 : como variancias homogeneas, está hipótesis se rechaza

3- MEDIDA DE KAYSER Como referencia, Kaiser puso los siguientes valores en los resultados:

0.00 a 0.49 inaceptable. 0.50 a 0.59 miserable. 0,60 a 0,69 mediocre. 0.70 a 0.79 medio. 0,80 a 0,89 meritorio. 0.90 a 1.00 maravilloso.

```
KMO(BASE)
## Kaiser-Meyer-Olkin factor adequacy
## Call:  KMO(r = BASE)
## Overall MSA = 0.57
## MSA for each item =
##      Q_CONTRAT MONT_CONTRAT      CONF      REPRES      ANT_BAN      EXP_LAP
##      0.59      0.54      0.67      0.44      0.85      0.77
##      T_CONTRAT      N_REP
##      0.71      0.39
```

Se esperan valores mayores a 0.5, lo cual no sucede en todas las variables, por lo cual se repite el análisis eliminando la variable "REPRES" y "N_REP", que es la que da un valor que según la teoría se da como inaceptable Y ADEMÁS PRESENTAN MUY BAJAS CORRELACIONES

SEGUNDO ANALISIS Se crea BASE2, sin las variables anteriores

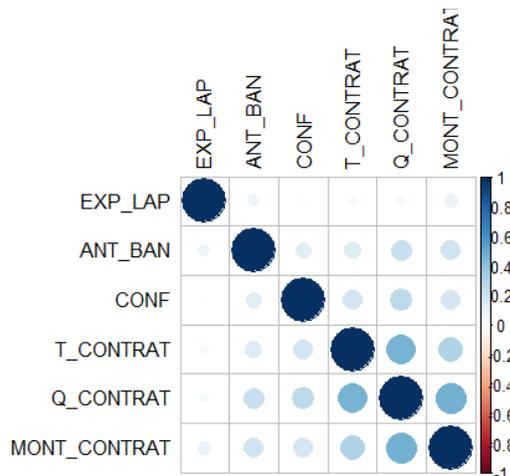
```
BASE2=select(BASE, -REPRES)
BASE2=select(BASE2, -N_REP)
```

MATRIZ DE CORRELACIONES

```
matriz_correlaciones2 <- cor(BASE2, use = "pairwise.complete.obs")
matriz_correlaciones2
##           Q_CONTRAT MONT_CONTRAT      CONF      ANT_BAN      EXP_LAP
## Q_CONTRAT  1.00000000  0.4701813  0.26537324  0.22397701  0.04970022
## MONT_CONTRAT 0.47018127  1.00000000  0.18952586  0.20113176  0.08689970
## CONF        0.26537324  0.1895259  1.00000000  0.13178320  0.01887758
## ANT_BAN     0.22397701  0.2011318  0.13178320  1.00000000  0.07921227
## EXP_LAP     0.04970022  0.0868997  0.01887758  0.07921227  1.00000000
## T_CONTRAT   0.46708922  0.3056797  0.19567515  0.14957660  0.04254242
##           T_CONTRAT
## Q_CONTRAT  0.46708922
```

```
## MONT_CONTRAT 0.30567973
## CONF          0.19567515
## ANT_BAN       0.14957660
## EXP_LAP       0.04254242
## T_CONTRAT     1.00000000
```

```
corrplot(cor(BASE2), order = "hclust", tl.col='black', tl.cex=1) #Gráfico de Las correlaciones
```



OTRA FORMA DE CORRELACIÓN

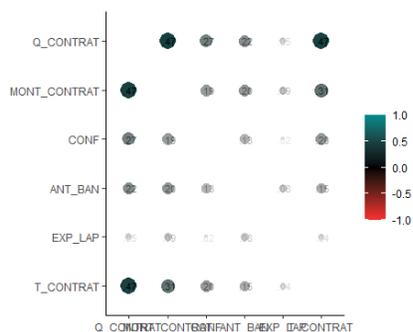
```
BASE_correlaciones2 <- correlate(BASE2) #Cálculo de un objeto de correlaciones
```

```
##
## Correlation method: 'pearson'
## Missing treated using: 'pairwise.complete.obs'
```

```
rplot(BASE_correlaciones2, legend = TRUE, colours = c("firebrick1", "black",
"darkcyan"), print_cor = TRUE) #Opción gráfica de Las correlaciones
```

```
## Don't know how to automatically pick scale for object of type noquote. Defaulting to contin
uous.
```

```
## Don't know how to automatically pick scale for object of type noquote. Defaulting to contin
uous.
```



Esta matriz representa una mejor correlación que las expuestas anteriormente

PRUEBAS QUE INDICAN SI ES CONVENIENTE REALIZAR EL ANÁLISIS

1- DETERMINANTE DE MATRIZ DE CORRELACIONES

```
det(matriz_correlaciones2)
```

```
## [1] 0.5093474
```

El valor aumenta, sin llegar a multicolinealidad perfecta, lo cual si permite el análisis

2- TEST DE BARTLET

```
bartlett.test(BASE2)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: BASE2  
## Bartlett's K-squared = 294.24, df = 5, p-value < 2.2e-16
```

como se rechaza el p value, y el valor de Bartlett's K-squared disminuye, lo que representa una mejora al modelo

3- MEDIDA DE KAYSER

```
library(psych)  
KMO(BASE2)
```

```
## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = BASE2)  
## Overall MSA = 0.71  
## MSA for each item =  
## Q_CONTRAT MONT_CONTRAT CONF ANT_BAN EXP_LAP T_CONTRAT  
## 0.66 0.72 0.82 0.81 0.70 0.72
```

TODAS LAS MEDIDAS MEJORAN, por lo que se procede al análisis factorial

ANALISIS FACTORIAL EXPLORATORIO Se inicia con dos factores, sin rotación

```
factanal(BASE2, factors = 2, rotation = "none")
```

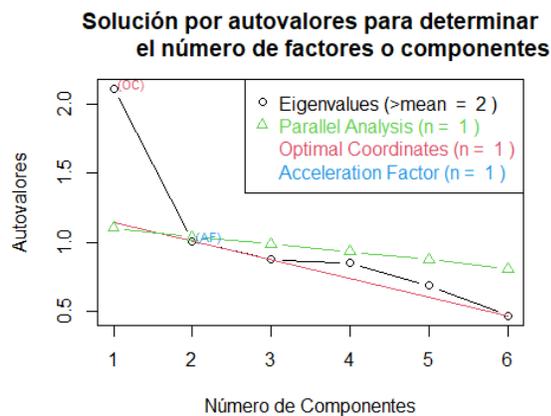
```
##  
## Call:  
## factanal(x = BASE2, factors = 2, rotation = "none")  
##  
## Uniquenesses:  
## Q_CONTRAT MONT_CONTRAT CONF ANT_BAN EXP_LAP T_CONTRAT  
## 0.283 0.655 0.889 0.840 0.949 0.694  
##  
## Loadings:  
## Factor1 Factor2  
## Q_CONTRAT 0.843  
## MONT_CONTRAT 0.569 0.144  
## CONF 0.327  
## ANT_BAN 0.292 0.273  
## EXP_LAP 0.212  
## T_CONTRAT 0.552  
##  
## Factor1 Factor2  
## SS loadings 1.537 0.152  
## Proportion Var 0.256 0.025  
## Cumulative Var 0.256 0.282  
##  
## Test of the hypothesis that 2 factors are sufficient.
```

```
## The chi square statistic is 0.9 on 4 degrees of freedom.
## The p-value is 0.924
```

no se rechaza la HN, lo cual indica que el número de factores es idóneo, no obstante hay variables que no presentan una buena definición y la variancia acumulada es poca

```
library(nFactors)

ev <- eigen(cor(BASE2)) # Obtención de Los autovalores
ap <- parallel(subject=nrow(BASE2), var=ncol(BASE2), rep=100, cent=.05)
nS <- nScreen(x=ev$values, aparallel=ap$eigen$evpea)
plotnScreen(nS, xlab = "Número de Componentes", ylab = "Autovalores",
  main = "Solución por autovalores para determinar
  el número de factores o componentes")
```



2 factores son los idóneos

se hace con rotación varimax con la finalidad de ver si mejora la definición de los factores

```
factanal(BASE2, factors = 2, rotation = "varimax")

##
## Call:
## factanal(x = BASE2, factors = 2, rotation = "varimax")
##
## Uniquenesses:
##      Q_CONTRAT MONT_CONTRAT      CONF      ANT_BAN      EXP_LAP      T_CONTRAT
##      0.283      0.655      0.889      0.840      0.949      0.694
##
## Loadings:
##              Factor1 Factor2
## Q_CONTRAT    0.838    0.124
## MONT_CONTRAT 0.518    0.276
## CONF         0.301    0.145
## ANT_BAN      0.218    0.335
## EXP_LAP      0.000    0.225
## T_CONTRAT    0.543    0.106
##
##              Factor1 Factor2
## SS loadings  1.403    0.287
## Proportion Var 0.234    0.048
## Cumulative Var 0.234    0.282
##
## Test of the hypothesis that 2 factors are sufficient.
```

```
## The chi square statistic is 0.9 on 4 degrees of freedom.  
## The p-value is 0.924
```

Si se presenta una mejor definición, pero sigue con los problemas de la variancia por lo cual se repetir+a el análisis quitando la variable “EXP_LAP”, “CONF” Y “EXP_LAP” y se procede a crear un único factor con estas tres variables, la cuales representan mucho sentido teórico ya que se refieren a la naturaleza del contrato de los proveedores

ANALISIS FACTORIAL EXPLORATORIO 2

TERCER ANALISIS Se crea BASE5, sin las variables

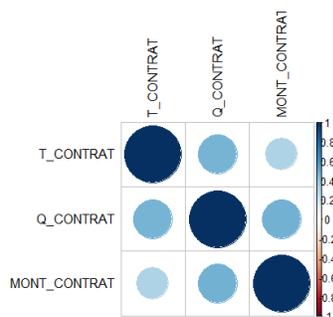
```
BASE3=select(BASE2, -EXP_LAP)  
BASE3=select(BASE3, -CONF)  
BASE3=select(BASE3, -ANT_BAN)
```

MATRIZ DE CORRELACIONES

```
matriz_correlaciones3 <- cor(BASE3, use = "pairwise.complete.obs")  
matriz_correlaciones3
```

```
##           Q_CONTRAT MONT_CONTRAT T_CONTRAT  
## Q_CONTRAT  1.0000000  0.4701813  0.4670892  
## MONT_CONTRAT 0.4701813  1.0000000  0.3056797  
## T_CONTRAT   0.4670892  0.3056797  1.0000000
```

```
corrplot(cor(BASE3), order = "hclust", tl.col='black', tl.cex=1) #Gráfico de Las correlaciones
```



OTRA FORMA DE CORRELACIÓN

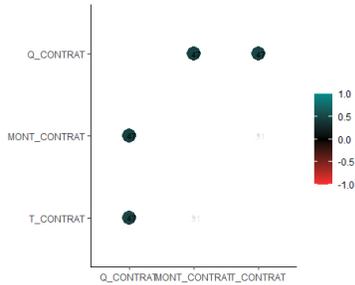
```
BASE_correlaciones3 <- correlate(BASE3) #Cálculo de un objeto de correlaciones
```

```
##  
## Correlation method: 'pearson'  
## Missing treated using: 'pairwise.complete.obs'
```

```
rplot(BASE_correlaciones3, legend = TRUE, colours = c("firebrick1", "black",  
  "darkcyan"), print_cor = TRUE) #Opción gráfica de Las correlaciones
```

```
## Don't know how to automatically pick scale for object of type noquote. Defaulting to continuous.
```

```
## Don't know how to automatically pick scale for object of type noquote. Defaulting to continuous.
```



se evidencia mayor correlación

PRUEBAS QUE INDICAN SI ES CONVENIENTE REALIZAR EL ANÁLISIS

1- DETERMINANTE DE MATRIZ DE CORRELACIONES

```
det(matriz_correlaciones3)
```

```
## [1] 0.6015818
```

el valor del determinante aumenta, pero no llega a ser demasiado alto

2- TEST DE BARTLET

```
bartlett.test(BASE3)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: BASE3
## Bartlett's K-squared = 242.27, df = 2, p-value < 2.2e-16
```

El test se rechaza, y el valor del Bartlett's K-squared disminuye un poco con respecto al anterior

3- MEDIDA DE KAYSER

```
library(psych)
KMO(BASE3)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = BASE3)
## Overall MSA = 0.63
## MSA for each item =
##   Q_CONTRAT MONT_CONTRAT   T_CONTRAT
##         0.59         0.66         0.66
```

los valores obtenidos son bajos, pero aceptables

Se procede de nuevo al analisis

Sin rotación

```
factanal(BASE3, factors = 1, rotation = "none")
```

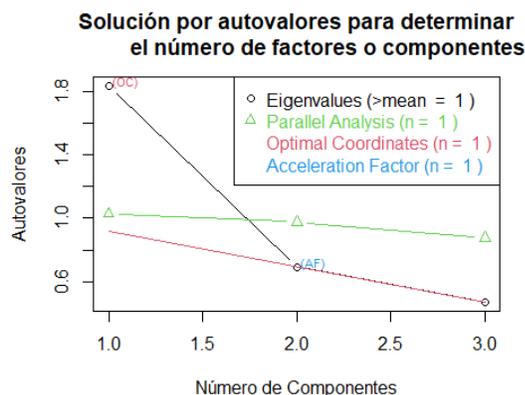
```
##
## Call:
## factanal(x = BASE3, factors = 1, rotation = "none")
##
## Uniquenesses:
##   Q_CONTRAT MONT_CONTRAT   T_CONTRAT
```

```
##          0.282          0.692          0.696
##
## Loadings:
##          Factor1
## Q_CONTRAT  0.848
## MONT_CONTRAT 0.555
## T_CONTRAT   0.551
##
##          Factor1
## SS loadings  1.330
## Proportion Var  0.443
##
## The degrees of freedom for the model is 0 and the fit was 0
```

se mejora notablemente la variancia explicada, y las puntuaciones, por otra parte esta combinación de variables toman un sentido teorico, que se refiere a la naturaleza del contrato

Gráfico de factores

```
library(nFactors)
ev <- eigen(cor(BASE3)) # Obtención de Los autovalores
ap <- parallel(subject=nrow(BASE3),var=ncol(BASE3),rep=100,cent=.05)
nS <- nScree(x=ev$values, aparallel=ap$eigen$gevpea)
plotnScree(nS,xlab = "Número de Componentes",ylab = "Autovalores",
           main = "Solución por autovalores para determinar
                 el número de factores o componentes")
```



se sugieren dos factores, no obstante no es ideal ni por el número de variables, ni por el sentido teórico

se hace con rotación varimax

```
factanal(BASE3, factors = 1, rotation = "varimax",scores = "Bartlett")
##
## Call:
## factanal(x = BASE3, factors = 1, scores = "Bartlett", rotation = "varimax")
##
## Uniquenesses:
##      Q_CONTRAT MONT_CONTRAT      T_CONTRAT
##          0.282          0.692          0.696
##
## Loadings:
##          Factor1
## Q_CONTRAT  0.848
## MONT_CONTRAT 0.555
```

```
## T_CONTRAT    0.551
##
##                Factor1
## SS loadings    1.330
## Proportion Var 0.443
##
## The degrees of freedom for the model is 0 and the fit was 0
```

se obtiene los mismo resultados que sin rotación

se hace con rotación Promax

```
factanal(BASE3, factors = 1, rotation = "Promax")

##
## Call:
## factanal(x = BASE3, factors = 1, rotation = "Promax")
##
## Uniquenesses:
##      Q_CONTRAT MONT_CONTRAT    T_CONTRAT
##      0.282      0.692      0.696
##
## Loadings:
##                Factor1
## Q_CONTRAT    0.848
## MONT_CONTRAT 0.555
## T_CONTRAT    0.551
##
##                Factor1
## SS loadings    1.330
## Proportion Var 0.443
##
## The degrees of freedom for the model is 0 and the fit was 0
```

se obtiene los mismo resultados

se procede a calcular las puntuaciones con varimax

```
factanal(BASE3, factors = 1, rotation = "varimax", scores = "regression")$scores

##                Factor1
## [1,] 0.10066603
## [2,] -0.88129866
## [3,] -0.37540354
## [4,] -0.78468930
## [5,] 0.48008738
## [6,] -0.37540354
## [7,] 0.48008738
## [8,] -0.50187732
## [9,] -0.62835110
## [10,] -0.18214594
```

Factor 1= Para poder utilizar los valores como indices se deben normalizar

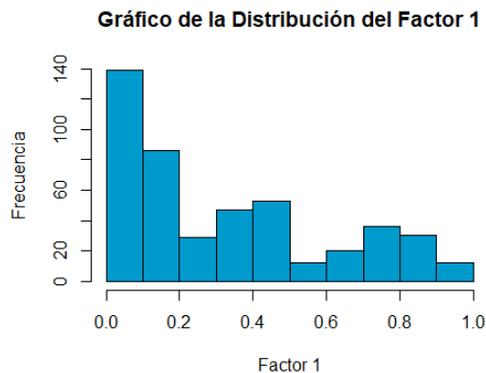
```

puntuaciones <- factanal(BASE3, factors = 1, rotation = "varimax", scores = "regression")$scores
BASE3 <- cbind(BASE3, puntuaciones)
BASE3$Factor1 <- round(((BASE3$Factor1 - min(BASE3$Factor1))/(max(BASE3$Factor1) - min(BASE3$Factor1))), 2)
head(BASE3)

##   Q_CONTRAT MONT_CONTRAT T_CONTRAT Factor1
## 1         50           25  16.66667    0.36
## 2          0            0  16.66667    0.05
## 3          0          100  16.66667    0.21
## 4          0           50   0.00000    0.08
## 5         50          100  16.66667    0.48
## 6          0          100  16.66667    0.21

hist(BASE3$Factor1, freq = TRUE, main = "Gráfico de la Distribución del Factor 1",
      xlab = "Factor 1", ylab = "Frecuencia", col = "#009ACD")

```



se crea la variable naturaleza de contrato la cual se integra al análisis factorial

de esta forma se agrego el constructo de naturaleza de contrato calculando las puntuaciones con los scores del AFE, y se integró al AFC

10.3 ANEXO 3: Análisis factorial confirmatorio realizado en R

María del Rocío

9/9/2022

LIBRERIAS

```
library(PASWR2)
```

```
library(lattice)
library(ggplot2)
library(dplyr)

library(psych)

library(openxlsx)
library(corrplot)

## corrplot 0.92 loaded

library(corr)
library(scales)

library(psych)
library(lavaan)

## This is lavaan 0.6-12
## lavaan is FREE software! Please report any bugs.

##
## Attaching package: 'lavaan'

## The following object is masked from 'package:psych':
##
##   cor2cov

library(semTools)

## Attaching package: 'semTools'

library(kableExtra)

##
## Attaching package: 'kableExtra'

library(scales)
library(semPlot)
```

LECTURA DE DATOS

```
library(readxl)
latentes <- read_excel("latentes2.xlsx")
```

comprobar valores faltantes VALORES FALTANTES

```
latentes = latentes[complete.cases(latentes),]
```

Explicación de los constructos



Selección de datos para análisis

```
BASE=latentes
BASE=select(latentes, -NUM_PROVEEDOR)
```

TEST DE BARTLET

```
cortest.bartlett(correlacion)->p_esf
## Warning in cortest.bartlett(correlacion): n not specified, 100 used
p_esf$p
## [1] 0.000114175
```

como se rechaza el p value, tomandop h0:como variancias homogeneas, está hipótesis se rechaza

3- MEDIDA DE KAYSER

```
KMO(BASE)
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = BASE)
## Overall MSA = 0.59
## MSA for each item =
## x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 x15 x16
## 0.62 0.64 0.43 0.65 0.54 0.54 0.62 0.73 0.69 0.57 0.58 0.43 0.57 0.60 0.64 0.48
## x17 x18
## 0.47 0.64
```

ANÁLISIS FACTORIAL CONFIRMATORIO 1

Se incluye la base ponderada para cada variable en cada constructo

#ordenar la base y almacenar en un objeto

```
tibble::as_tibble(BASE)

## # A tibble: 551 × 18
##   x1    x2    x3    x4    x5    x6    x7    x8    x9    x10   x11   x12   x13
##   <dbl> <dbl>
## 1  0     0     0     0     0.1  0.2  0.1  0     0     0     0     0     0.5
## 2  0     0     0     0     0.1  0.1  0.1  0.33  0.67  0     0     0     0
## 3  0     0.15  0     0     0.1  0.1  0.1  0.33  0.67  0.4   0     0     0
## 4  0.5   0.15  0     0     0.2  0.3  0.1  0.33  0.67  0     0     0     0
## 5  0     0.15  0     0     0.1  0.1  0.1  0.33  0     0     0     0     0

datos <- tibble::as_tibble(
  BASE)
#View(datos)
```

#creacion de modelo SE CREA EL MODELO DE ACUERDO A LA IMAGEN 1

```
model <- " Reputacion =~ x1 + x2 + x3+ x4
  resp_social =~ x5 + x6 + x7
  seg_econ =~ x8 + x9
  cump_servicio =~ x10 + x11 +x12
  nat_contrato =~ x13 + x14 + x15
  confianza =~ x16 + x17 + x18"
```

#analisis

```
library(lavaan)

fit <- cfa(model = model,
  data = datos)
#summary(datos)

#summary(fit)
summary(fit, fit.measures = TRUE, standardized = TRUE)

## lavaan 0.6-12 ended normally after 313 iterations
##
## Estimator ML
## Optimization method NLMINB
## Number of model parameters 51
##
## Number of observations 551
##
## Model Test User Model:
##
## Test statistic 154.194
## Degrees of freedom 120
## P-value (Chi-square) 0.019
##
## Model Test Baseline Model:
##
## Test statistic 1224.562
## Degrees of freedom 153
## P-value 0.000
##
## User Model versus Baseline Model:
##
## Comparative Fit Index (CFI) 0.968
## Tucker-Lewis Index (TLI) 0.959
```

```

##
## Loglikelihood and Information Criteria:
##
## Loglikelihood user model (H0)          6359.272
## Loglikelihood unrestricted model (H1)  6436.369
##
## Akaike (AIC)                          -12616.544
## Bayesian (BIC)                         -12396.646
## Sample-size adjusted Bayesian (BIC)    -12558.542
##
## Root Mean Square Error of Approximation:
##
## RMSEA                                  0.023
## 90 Percent confidence interval - lower  0.010
## 90 Percent confidence interval - upper  0.033
## P-value RMSEA <= 0.05                  1.000
##
## Standardized Root Mean Square Residual:
##
## SRMR                                    0.037
##
## Parameter Estimates:
##
## Standard errors                        Standard
## Information                            Expected
## Information saturated (h1) model       Structured
##
## Latent Variables:
##
##      Estimate Std.Err  z-value  P(>|z|)  Std.lv  Std.all
## Reputacion =~
##   x1          1.000
##   x2          0.240    0.032    7.387    0.000    0.030    0.725
##   x3          0.002    0.007    0.312    0.755    0.000    0.016
##   x4          0.207    0.034    6.015    0.000    0.026    0.348
## resp_social =~
##   x5          1.000
##   x6          1.201    0.136    8.844    0.000    0.087    0.932
##   x7          0.028    0.012    2.340    0.019    0.002    0.106
## seg_econ =~
##   x8          1.000
##   x9          1.158    0.217    5.342    0.000    0.130    0.390
## cump_servicio =~
##   x10         1.000
##   x11         0.476    0.172    2.770    0.006    0.031    0.466
##   x12        -0.002    0.111   -0.015    0.988   -0.000   -0.001
## nat_contrato =~
##   x13         1.000
##   x14         0.580    0.077    7.499    0.000    0.189    0.525
##   x15         0.308    0.041    7.584    0.000    0.100    0.552
## confianza =~
##   x16         1.000
##   x17         1.027    0.333    3.088    0.002    0.242    0.671
##   x18         0.377    0.109    3.443    0.001    0.089    0.225
##
## Covariances:
##
##      Estimate Std.Err  z-value  P(>|z|)  Std.lv  Std.all
## Reputacion ~~
##   resp_social  -0.000    0.001   -0.483    0.629   -0.027   -0.027
##   seg_econ      0.007    0.001    5.590    0.000    0.508    0.508
##   cump_servicio 0.003    0.001    2.822    0.005    0.334    0.334
##   nat_contrato  0.000    0.002    0.116    0.907    0.007    0.007

```

```

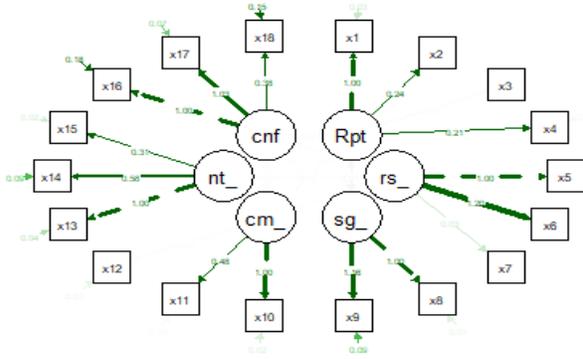
##   confianza      -0.003    0.002   -1.268    0.205   -0.097   -0.097
## resp_social ~~
##   seg_econ      -0.004    0.001   -5.634    0.000   -0.450   -0.450
##   cump_servicio  0.001    0.000    1.678    0.093    0.142    0.142
##   nat_contrato  -0.001    0.001   -0.863    0.388   -0.045   -0.045
##   confianza      0.001    0.001    1.227    0.220    0.079    0.079
##   seg_econ ~~
##   cump_servicio  0.001    0.001    1.079    0.281    0.114    0.114
##   nat_contrato  0.000    0.002    0.091    0.928    0.006    0.006
##   confianza      0.001    0.002    0.662    0.508    0.054    0.054
##   cump_servicio ~~
##   nat_contrato  -0.002    0.002   -0.952    0.341   -0.082   -0.082
##   confianza      0.002    0.002    1.417    0.156    0.158    0.158
##   nat_contrato ~~
##   confianza      -0.001    0.005   -0.143    0.886   -0.009   -0.009
##
## Variances:
##           Estimate Std.Err z-value P(>|z|) Std.lv Std.all
##   .x1          0.029  0.003  10.611  0.000  0.029  0.640
##   .x2          0.001  0.000   6.569  0.000  0.001  0.474
##   .x3          0.000  0.000  16.596  0.000  0.000  1.000
##   .x4          0.005  0.000  15.406  0.000  0.005  0.879
##   .x5          0.004  0.001   6.681  0.000  0.004  0.437
##   .x6          0.001  0.001   1.414  0.157  0.001  0.131
##   .x7          0.000  0.000  16.574  0.000  0.000  0.989
##   .x8          0.015  0.003   5.811  0.000  0.015  0.535
##   .x9          0.094  0.006  14.517  0.000  0.094  0.848
##   .x10         0.019  0.002   9.611  0.000  0.019  0.818
##   .x11         0.004  0.000   8.228  0.000  0.004  0.783
##   .x12         0.011  0.001  16.598  0.000  0.011  1.000
##   .x13         0.042  0.013   3.237  0.001  0.042  0.282
##   .x14         0.094  0.007  13.262  0.000  0.094  0.724
##   .x15         0.023  0.002  12.534  0.000  0.023  0.695
##   .x16         0.181  0.021   8.661  0.000  0.181  0.766
##   .x17         0.071  0.019   3.703  0.000  0.071  0.550
##   .x18         0.148  0.009  15.742  0.000  0.148  0.949
##   Reputacion   0.016  0.003   5.493  0.000  1.000  1.000
##   resp_social  0.005  0.001   6.970  0.000  1.000  1.000
##   seg_econ     0.013  0.003   4.638  0.000  1.000  1.000
##   cump_servicio 0.004  0.002   2.349  0.019  1.000  1.000
##   nat_contrato 0.106  0.015   6.964  0.000  1.000  1.000
##   confianza    0.055  0.020   2.754  0.006  1.000  1.000

```

SE AANALIZANLOS SIGUIENTES INDICADORES CFI ($\geq .95$), TLI ($\geq .95$), RMSEA ($\leq .05$) y SRMR ($\leq .06$).

para el modelo sepuede ver CFI= 0.968 TLI) = 0.959 RMSEA=0.023 SRMR=0.037 si BIEN, LAS MEDIDAS CUMPLEN, ES POSIBLE UNA MEJORA DEL MODELO, DADO QUE LOS VALORES ESTIMADOS, NO EN TODOS LOS CONSTRUCTOS PROPORCIONAN DEFINICIÓN.

```
semPaths(fit, what = "est", layout = "circle", title=TRUE, style = "LISREL")
```



DEL GRÁFICO ANTERIOR, SE PUESE VER QUE SE AGRUPAN BIEN LOS CONSTRUCTOS, NO OBSTANTE HAY VARIABLES QUE NO APORTAN, por lo cual se vuelve a realizar el análisis, quitando x3 a reputación y x7 a responsabilidad social

SEGUNDO ANALISIS MODIFICANDO algunos inddices

LECTURA DE DATOS

```
library(readxl)
latentes2 <- read_excel("latentes2v3.xlsx")
```

comprobar valores faltantes VALORES FALTANTES

```
latentes2 = latentes2[complete.cases(latentes2),]
```

seleccion de datos para analisis

```
BASE2=latentes2
BASE2=select(latentes2, -NUM_PROVEEDOR)
```

TEST DE BARTLETT

```
bartlett.test(BASE2)

##
## Bartlett test of homogeneity of variances
##
## data: BASE2
## Bartlett's K-squared = 6647.7, df = 14, p-value < 2.2e-16
```

MEDIDA DE KAYSER

```
KMO(BASE2)

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = BASE2)
## Overall MSA = 0.59
## MSA for each item =
## x1 x2 x4 x5 x6 x8 x9 x10 x11 x13 x14 x15 x16 x17 x18
## 0.61 0.64 0.65 0.54 0.54 0.73 0.71 0.57 0.61 0.57 0.61 0.64 0.48 0.48 0.66
```

se mantiene le mismo valor

ANALISIS FACTORIAL CONFIRMATORIO 2

#ordenar la base y almacenar en un objeto

```
datos2 <- tibble::as_tibble(
  BASE2)
#View(datos)
```

```
#creacion de modelo
```

```
model2 <- " Reputacion =~ x1 + x2 + x4
  resp_social =~ x5 + x6
  seg_econ =~ x8 + x9
  cump_servicio =~ x10 + x11
  nat_contrato =~ x13 + x14 + x15
  confianza =~ x16 + x17 + x18"
```

```
#analisis
```

```
library(lavaan)

fit <- cfa(model = model2,
  data = datos2)
#summary(datos)

#summary(fit)
summary(fit, fit.measures = TRUE, standardized = TRUE)

## lavaan 0.6-12 ended normally after 204 iterations
##
## Estimator ML
## Optimization method NLMINB
## Number of model parameters 45
##
## Number of observations 551
##
## Model Test User Model:
##
## Test statistic 91.052
## Degrees of freedom 75
## P-value (Chi-square) 0.100
##
## Model Test Baseline Model:
##
## Test statistic 1156.011
## Degrees of freedom 105
## P-value 0.000
##
## User Model versus Baseline Model:
##
## Comparative Fit Index (CFI) 0.985
## Tucker-Lewis Index (TLI) 0.979
##
## Loglikelihood and Information Criteria:
##
## Loglikelihood user model (H0) 3053.386
## Loglikelihood unrestricted model (H1) 3098.912
##
## Akaike (AIC) -6016.772
## Bayesian (BIC) -5822.744
## Sample-size adjusted Bayesian (BIC) -5965.593
##
## Root Mean Square Error of Approximation:
##
## RMSEA 0.020
```

```

## 90 Percent confidence interval - lower      0.000
## 90 Percent confidence interval - upper     0.033
## P-value RMSEA <= 0.05                    1.000
##
## Standardized Root Mean Square Residual:
##
## SRMR                                       0.032
##
## Parameter Estimates:
##
## Standard errors                          Standard
## Information                              Expected
## Information saturated (h1) model         Structured
##
## Latent Variables:
##      Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
## Reputacion =~
##   x1           1.000
##   x2           0.241   0.033   7.364   0.000   0.030   0.727
##   x4           0.207   0.034   6.012   0.000   0.026   0.348
## resp_social =~
##   x5           1.000
##   x6           1.143   0.130   8.814   0.000   0.085   0.910
## seg_econ =~
##   x8           1.000
##   x9           1.153   0.216   5.332   0.000   0.130   0.390
## cump_servicio =~
##   x10          1.000
##   x11          0.483   0.174   2.770   0.006   0.065   0.424
## nat_contrato =~
##   x13          1.000
##   x14          0.581   0.077   7.504   0.000   0.189   0.525
##   x15          0.308   0.041   7.588   0.000   0.100   0.553
## confianza =~
##   x16          1.000
##   x17          1.007   0.325   3.100   0.002   0.239   0.664
##   x18          0.375   0.109   3.432   0.001   0.089   0.226
##
## Covariances:
##      Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
## Reputacion ~~
##   resp_social  -0.000   0.001  -0.627   0.531  -0.036  -0.036
##   seg_econ      0.007   0.001   5.583   0.000   0.507   0.507
##   cump_servicio 0.003   0.001   2.802   0.005   0.332   0.332
##   nat_contrato  0.000   0.002   0.117   0.907   0.007   0.007
##   confianza    -0.003   0.002  -1.262   0.207  -0.096  -0.096
## resp_social ~~
##   seg_econ     -0.004   0.001  -5.789   0.000  -0.460  -0.460
##   cump_servicio 0.001   0.000   1.695   0.090   0.146   0.146
##   nat_contrato -0.001   0.001  -0.828   0.408  -0.043  -0.043
##   confianza     0.001   0.001   1.195   0.232   0.078   0.078
## seg_econ ~~
##   cump_servicio 0.001   0.001   1.064   0.287   0.112   0.112
##   nat_contrato  0.000   0.002   0.088   0.930   0.006   0.006
##   confianza     0.001   0.002   0.680   0.497   0.056   0.056
## cump_servicio ~~
##   nat_contrato -0.002   0.002  -0.944   0.345  -0.081  -0.081
##   confianza     0.002   0.002   1.411   0.158   0.158   0.158
## nat_contrato ~~
##   confianza    -0.001   0.005  -0.127   0.899  -0.008  -0.008
##

```

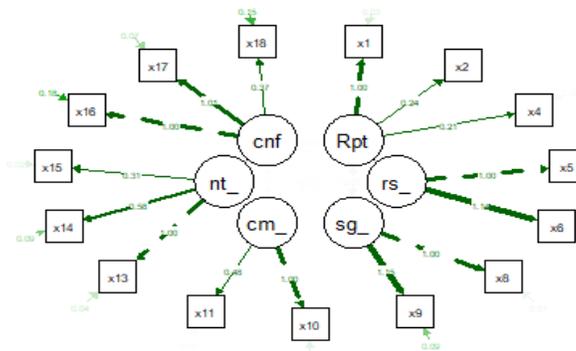
```

## Variances:
##           Estimate Std.Err z-value P(>|z|) Std.lv Std.all
##    .x1           0.029   0.003  10.636  0.000   0.029   0.642
##    .x2           0.001   0.000   6.502  0.000   0.001   0.472
##    .x4           0.005   0.000  15.408  0.000   0.005   0.879
##    .x5           0.004   0.001   6.011  0.000   0.004   0.409
##    .x6           0.001   0.001   1.934  0.053   0.001   0.173
##    .x8           0.015   0.003   5.764  0.000   0.015   0.533
##    .x9           0.094   0.006  14.533  0.000   0.094   0.848
##    .x10          0.019   0.002   9.724  0.000   0.019   0.820
##    .x11          0.004   0.000   8.114  0.000   0.004   0.780
##    .x13          0.042   0.013   3.251  0.001   0.042   0.283
##    .x14          0.094   0.007  13.257  0.000   0.094   0.724
##    .x15          0.023   0.002  12.529  0.000   0.023   0.695
##    .x16          0.180   0.021   8.518  0.000   0.180   0.762
##    .x17          0.072   0.019   3.843  0.000   0.072   0.559
##    .x18          0.147   0.009  15.734  0.000   0.147   0.949
##    Reputacion    0.016   0.003   5.477  0.000   1.000   1.000
##    resp_social   0.005   0.001   7.044  0.000   1.000   1.000
##    seg_econ       0.013   0.003   4.639  0.000   1.000   1.000
##    cump_servicio 0.004   0.002   2.346  0.019   1.000   1.000
##    nat_contrato  0.106   0.015   6.964  0.000   1.000   1.000
##    confianza     0.056   0.020   2.767  0.006   1.000   1.000

```

con respecto al modelo anterior, para el modelo sepuede ver CFI= 0.98 TLI) = 0.97 RMSEA=0.020 SRMR=0.032, se puede ver que mejora los indicadores, no obstante el p-value es significativo

```
semPaths(fit, what = "est", layout = "circle", title=TRUE,style = "LISREL")
```



para este caso, vemos que la variable x18 da poco al constructo de confianza

TERCER ANALISIS MODIFICANDO algunos indices

comprobar valores faltantes VALORES FALTANTES

```
latentes2 = latentes2[complete.cases(latentes2),]
```

seleccion de datos para analisis

```

BASE3=latentes2
BASE3=select(BASE2, -x18)

```

TEST DE BARTLET

```

bartlett.test(BASE3)

##
## Bartlett test of homogeneity of variances
##
## data: BASE3
## Bartlett's K-squared = 6398.2, df = 13, p-value < 2.2e-16

```

MEDIDA DE KAYSER

```

KMO(BASE3)

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = BASE3)
## Overall MSA = 0.59
## MSA for each item =
## x1 x2 x4 x5 x6 x8 x9 x10 x11 x13 x14 x15 x16 x17
## 0.61 0.63 0.66 0.54 0.54 0.73 0.71 0.57 0.61 0.57 0.61 0.63 0.45 0.45

```

Se mantiene el valor global

ANALISIS FACTORIAL CONFIRMATORIO

#ordenar la base y almacenar en un objeto

```

tibble::as_tibble(BASE3)

datos3 <- tibble::as_tibble(
  BASE3)
#View(datos)

```

#creacion de modelo

```

model3 <- ' Reputacion =~ x1 + x2 + x4
resp_social =~ x5 + x6
seg_econ =~ x8 + x9
cump_servicio =~ x10 + x11
nat_contrato =~ x13 + x14 + x15
confianza =~ x16 + x17'

```

analisis

```

library(lavaan)

fit <- cfa(model = model3,
  data = datos3)
#summary(datos)

#summary(fit)
summary(fit, fit.measures = TRUE, standardized = TRUE)

## lavaan 0.6-12 ended normally after 209 iterations
##
## Estimator ML
## Optimization method NLMINB
## Number of model parameters 43
##
## Number of observations 551
##
## Model Test User Model:
##

```

```

## Test statistic 76.667
## Degrees of freedom 62
## P-value (Chi-square) 0.100
##
## Model Test Baseline Model:
##
## Test statistic 1126.984
## Degrees of freedom 91
## P-value 0.000
##
## User Model versus Baseline Model:
##
## Comparative Fit Index (CFI) 0.986
## Tucker-Lewis Index (TLI) 0.979
##
## Loglikelihood and Information Criteria:
##
## Loglikelihood user model (H0) 3314.931
## Loglikelihood unrestricted model (H1) 3353.265
##
## Akaike (AIC) -6543.862
## Bayesian (BIC) -6358.458
## Sample-size adjusted Bayesian (BIC) -6494.958
##
## Root Mean Square Error of Approximation:
##
## RMSEA 0.021
## 90 Percent confidence interval - lower 0.000
## 90 Percent confidence interval - upper 0.035
## P-value RMSEA <= 0.05 1.000
##
## Standardized Root Mean Square Residual:
##
## SRMR 0.029
##
## Parameter Estimates:
##
## Standard errors Standard
## Information Expected
## Information saturated (h1) model Structured
##
## Latent Variables:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## Reputacion =~
## x1 1.000 0.126 0.597
## x2 0.242 0.033 7.360 0.000 0.030 0.727
## x4 0.208 0.035 6.032 0.000 0.026 0.349
## resp_social =~
## x5 1.000 0.075 0.775
## x6 1.125 0.127 8.891 0.000 0.084 0.902
## seg_econ =~
## x8 1.000 0.113 0.686
## x9 1.144 0.214 5.353 0.000 0.129 0.388
## cump_servicio =~
## x10 1.000 0.064 0.418
## x11 0.496 0.184 2.700 0.007 0.032 0.476
## nat_contrato =~
## x13 1.000 0.326 0.847
## x14 0.581 0.077 7.508 0.000 0.189 0.526
## x15 0.308 0.041 7.592 0.000 0.100 0.553
## confianza =~

```

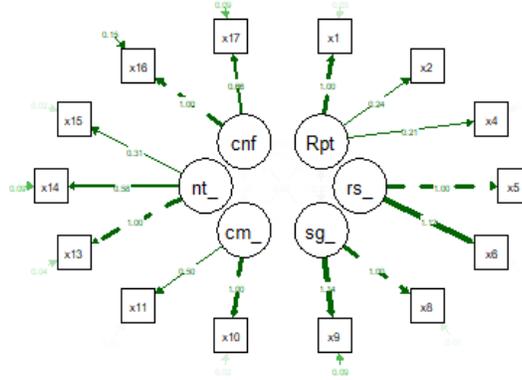
```

##      x16          1.000
##      x17          0.664    0.322    2.059    0.039    0.194    0.539
##
## Covariances:
##      Estimate Std.Err z-value P(>|z|) Std.lv Std.all
##      Reputacion ~~
##      resp_social -0.000    0.001   -0.660    0.509   -0.038   -0.038
##      seg_econ     0.007    0.001   5.583    0.000    0.506    0.506
##      cump_servicio 0.003    0.001   2.745    0.006    0.331    0.331
##      nat_contrato 0.000    0.002    0.116    0.908    0.007    0.007
##      confianza    -0.003    0.003   -1.163    0.245   -0.091   -0.091
##      resp_social ~~
##      seg_econ    -0.004    0.001   -5.856    0.000   -0.461   -0.461
##      cump_servicio 0.001    0.000    1.713    0.087    0.149    0.149
##      nat_contrato -0.001    0.001   -0.815    0.415   -0.043   -0.043
##      confianza     0.001    0.001    0.988    0.323    0.066    0.066
##      seg_econ ~~
##      cump_servicio 0.001    0.001    1.034    0.301    0.109    0.109
##      nat_contrato 0.000    0.002    0.084    0.933    0.006    0.006
##      confianza     0.004    0.003    1.254    0.210    0.110    0.110
##      cump_servicio ~~
##      nat_contrato -0.002    0.002   -0.929    0.353   -0.080   -0.080
##      confianza     0.003    0.002    1.196    0.232    0.134    0.134
##      nat_contrato ~~
##      confianza    -0.001    0.007   -0.193    0.847   -0.013   -0.013
##
## Variances:
##      Estimate Std.Err z-value P(>|z|) Std.lv Std.all
##      .x1         0.029    0.003   10.682    0.000    0.029    0.644
##      .x2         0.001    0.000    6.482    0.000    0.001    0.471
##      .x4         0.005    0.000   15.395    0.000    0.005    0.878
##      .x5         0.004    0.001    5.848    0.000    0.004    0.399
##      .x6         0.002    0.001    2.131    0.033    0.002    0.186
##      .x8         0.014    0.003    5.723    0.000    0.014    0.530
##      .x9         0.095    0.006   14.580    0.000    0.095    0.849
##      .x10        0.020    0.002    9.786    0.000    0.020    0.825
##      .x11        0.003    0.000    7.766    0.000    0.003    0.774
##      .x13        0.042    0.013    3.261    0.001    0.042    0.283
##      .x14        0.094    0.007   13.253    0.000    0.094    0.724
##      .x15        0.023    0.002   12.527    0.000    0.023    0.694
##      .x16        0.151    0.043    3.556    0.000    0.151    0.639
##      .x17        0.092    0.019    4.808    0.000    0.092    0.709
##      Reputacion  0.016    0.003    5.467    0.000    1.000    1.000
##      resp_social 0.006    0.001    7.114    0.000    1.000    1.000
##      seg_econ    0.013    0.003    4.664    0.000    1.000    1.000
##      cump_servicio 0.004    0.002    2.297    0.022    1.000    1.000
##      nat_contrato 0.106    0.015    6.964    0.000    1.000    1.000
##      confianza    0.085    0.043    1.990    0.047    1.000    1.000

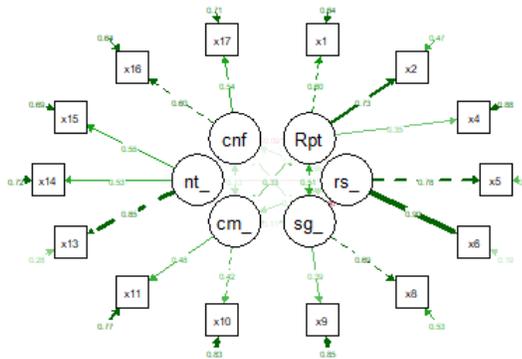
```

se puede ver que los indicadores son similares al modelo anterior, los cuales cumplen, no obstante mejoran los valores estandarizados (Std.all)

```
semPaths(fit, what = "est", layout = "circle", title=TRUE, style = "LISREL")
```



```
semPaths(fit, what = "stand", layout = "circle", title=TRUE, style = "LISREL")
```



se mantiene este modelo, utilizando los valores estandarizados como pesos para la creación de índice final

| CONSTRUCTO | PESO V1 | PESO V2 | PESO V3 | TOTAL | RESCALAMIENTO DE VALORES |
|------------------------------|---------|---------|---------|-------|--------------------------|
| CUMPLIMIENTO_SERVICIO | 0.83 | 0.77 | | 1.6 | 0.19 |
| CONFIANZA_1 | 0.71 | 0.64 | | 1.35 | 0.16 |
| SEGURIDAD_ECONOMICA_ESCALA_2 | 0.53 | 0.85 | | 1.38 | 0.16 |
| NATU_CONTRATO | 0.26 | 0.72 | 0.69 | 1.67 | 0.19 |
| RES_SOCIAL2 | 0.4 | 0.19 | | 0.59 | 0.07 |
| REPUTACION_3 | 0.88 | 0.47 | 0.64 | 1.99 | 0.23 |

Los valores en la columna final son los pesos por constructo, los cuales tiene un buen sentido teórico

En el siguiente histograma se muestra la distribución por notas para os proveedores

10.4 ANEXO 3: Análisis de conglomerados realizado en R

Cluster tamaño

María del Rocío

14/10/2022

```
library(dplyr)

library(tidyverse)

library(cluster)
library(factoextra)

library(NbClust)
library(tidyr)

library(readxl)
library(readxl)
df <- read_excel("D:/mmorafa/Documents/documentos Rocío/Proveedores/Documents/analisis final/CONSTRUCTOS FINALESv5.xlsx")
View(df)

library(readxl)
df_T <- read_excel("D:/mmorafa/Documents/documentos Rocío/Proveedores/Documents/analisis final/CONSTRUCTOS FINALESv3.xlsx")

#normalizar las puntuaciones

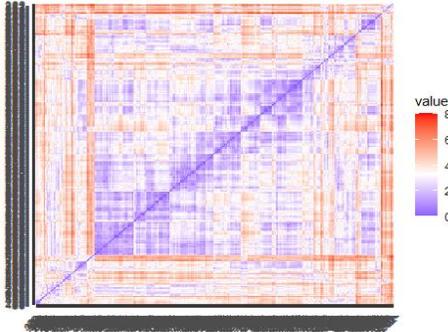
#si uso la base con puntaje no reescalo

df <- scale(df)
head(df)

##      RESPONBILIDAD_SOCIAL_REESCALADA CUMPLIMIENTO_SERVICIO_REESCALADA
## [1,]                -0.2295340                   0.4865972
## [2,]                0.7142961                   0.4865972
## [3,]                0.7142961                   -1.1582804
## [4,]               -2.1171942                   -1.9807192
## [5,]                0.7142961                   0.4865972
## [6,]                0.7142961                   -3.6255969
##      CONFIANZA_1_REESCALADA NATURALEZA_CONTRATO_REESCALADA
## [1,]          -0.1024680                -0.6357522
## [2,]          -0.5769208                -1.3697907
## [3,]          -1.0513736                 1.0770042
## [4,]           0.3719848                -1.3697907
## [5,]          -0.1024680                 0.9954444
## [6,]           1.3208903                 1.0770042
##      SEGURIDAD_ECONOMICA_2_REESCALADA REPUTACION_REESCALADA
## [1,]          -1.1455860                -0.5951634
## [2,]           1.3105148                -0.5951634
## [3,]           1.3105148                 0.1087109
## [4,]           1.3105148                 1.9387840
## [5,]          -0.2859508                 0.1087109
## [6,]           1.3105148                -0.5951634
```

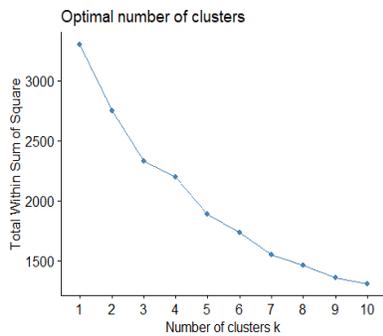
#calcular la matriz de distancias

```
m.distancia <- get_dist(df, method = "euclidean") #el método aceptado también puede ser: "maximum", "manhattan", "canberra", "binary", "minkowski", "pearson", "spearman" o "kendall"  
fviz_dist(m.distancia, gradient = list(low = "blue", mid = "white", high = "red"))
```

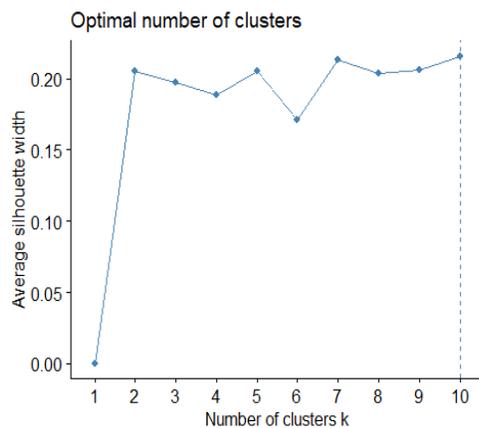


#estimar el número de clústers #Elbow, silhouette o gap_stat method

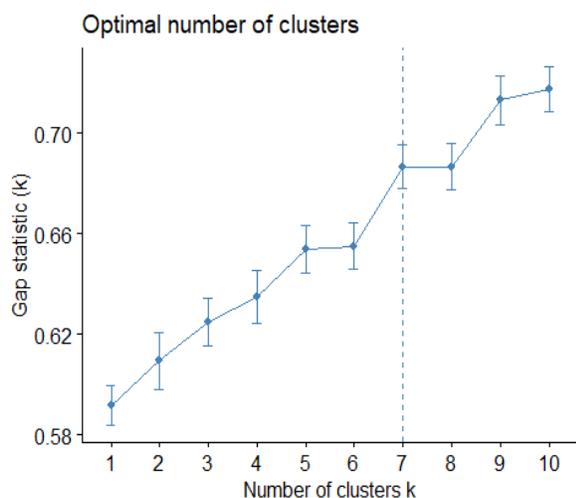
```
fviz_nbclust(df, kmeans, method = "wss")
```



```
fviz_nbclust(df, kmeans, method = "silhouette")
```



```
fviz_nbclust(df, kmeans, method = "gap_stat")
```



#con esta función se pueden calcular: #the index to be calculated. This should be one of : "kl", "ch", "hartigan", "ccc", "scott", #"marriot", "trcovw", "tracew", "friedman", "rubin", "cindex", "db", "silhouette", "duda", #"pseudot2", "beale", "ratkowsky", "ball", "ptbiserial", "gap", "frey", "mcclain", "gamma", #"gplus", "tau", "dunn", "hubert", "sdindex", "dindex", "sdbw", "all" (all indices except GAP, #Gamma, Gplus and Tau), "alllong" (all indices with Gap, Gamma, Gplus and Tau included).

```
#resnumclust<-NbClust(df, distance = "euclidean", min.nc=2, max.nc=7, method = "kmeans", index = "alllong")
#fviz_nbclust(resnumclust)
```

#calculamos los dos clústers

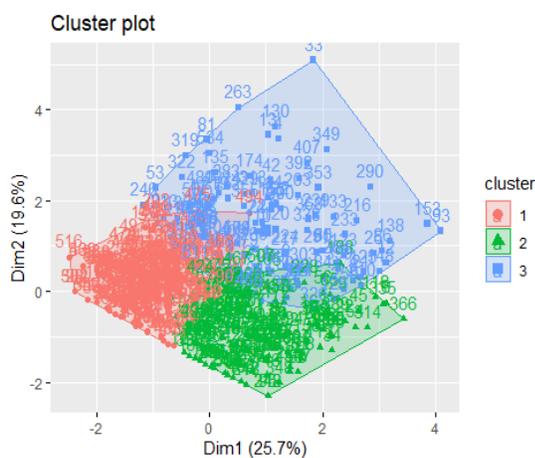
```
k3 <- kmeans(df, centers = 3, nstart = 25)
k3

## K-means clustering with 3 clusters of sizes 252, 180, 119
##
## Cluster means:
##  RESPONSABILIDAD_SOCIAL_REESCALADA CUMPLIMIENTO_SERVICIO_REESCALADA
## 1 -0.01604862 0.2875148
## 2 0.17421554 -0.1256628
## 3 -0.22953400 -0.4187766
##  CONFIANZA_1_REESCALADA NATURALEZA_CONTRATO_REESCALADA
## 1 -0.69553400 -0.5982088
## 2 0.96241488 0.7267501
## 3 0.01714192 0.1675093
##  SEGURIDAD_ECONOMICA_2_REESCALADA REPUTACION_REESCALADA
## 1 -0.1265966 -0.4940513
## 2 -0.1044722 -0.4242783
## 3 0.4261121 1.6879918
##
## Clustering vector:
## [1] 1 1 3 3 2 2 1 3 3 2 1 3 3 1 3 1 2 1 1 2 3 2 3 1 3 1 3 1 2 1 1 1 3 2 2 2 1
## [38] 2 2 2 1 3 1 1 2 1 3 3 1 3 2 2 3 2 3 2 1 1 3 2 2 3 1 2 3 1 2 1 3 1 3 1 2 3
## [75] 3 1 2 1 1 3 3 1 1 3 2 3 2 2 1 1 2 1 3 2 2 1 1 1 2 2 3 2 1 1 2 2 1 1 2 3 1
## [112] 2 3 3 1 2 3 2 1 3 1 3 3 2 3 3 2 1 1 3 3 1 3 2 3 3 2 3 1 3 1 2 1 1 1 2 1 3
```

```
## [149] 1 1 1 3 3 1 1 2 3 2 2 3 1 1 2 2 1 3 2 2 2 2 2 2 2 3 1 1 1 1 3 2 1 1 2 2 2
##
## Within cluster sum of squares by cluster:
## [1] 840.7127 685.4940 779.6052
## (between_SS / total_SS = 30.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
str(k3)
## List of 9
## $ cluster      : int [1:551] 1 1 3 3 2 2 1 3 3 2 ...
## $ centers      : num [1:3, 1:6] -0.016 0.174 -0.23 0.288 -0.126 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:3] "1" "2" "3"
## .. ..$ : chr [1:6] "RESPONBILIDAD_SOCIAL_REESCALADA" "CUMPLIMIENTO_SERVICIO_REES
CALADA" "CONFIANZA_1_REESCALADA" "NATURALEZA_CONTRATO_REESCALADA" ...
## $ totss       : num 3300
## $ withinss    : num [1:3] 841 685 780
## $ tot.withinss: num 2306
## $ betweenss   : num 994
## $ size        : int [1:3] 252 180 119
## $ iter        : int 3
## $ ifault      : int 0
## - attr(*, "class")= chr "kmeans"
```

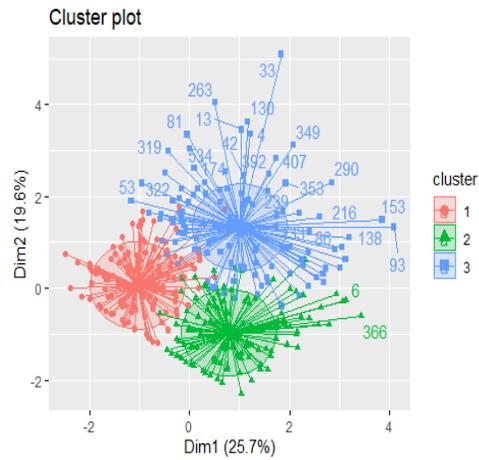
#plotear los cluster

```
fviz_cluster(k3, data = df)
```

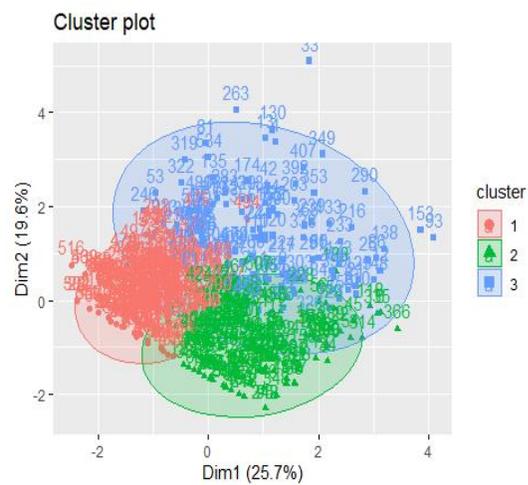


```
fviz_cluster(k3, data = df, ellipse.type = "euclid", repel = TRUE, star.plot = TRUE) #e
llipse.type= "t", "norm", "euclid"
```

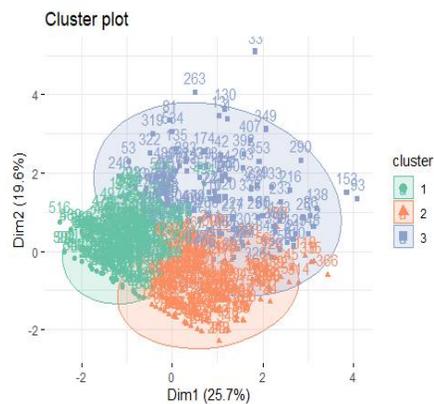
```
## Warning: ggrepel: 525 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
fviz_cluster(k3, data = df, ellipse.type = "norm")
```



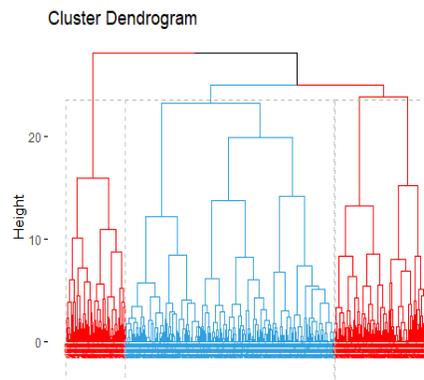
```
fviz_cluster(k3, data = df, ellipse.type = "norm", palette = "Set2", ggtheme = theme_minimal())
```



```
res3 <- hcut(df, k = 3, stand = TRUE)
fviz_dend(res3, rect = TRUE, cex = 0.5,
           k_colors = c("red", "#2E9FDF"))
```

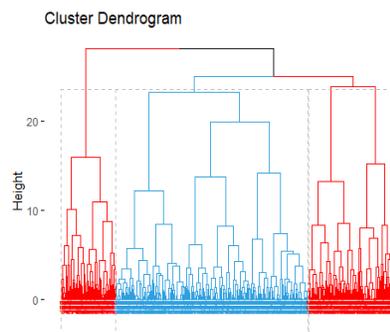
```
## Warning in get_col(col, k): Length of color vector was shorter than the number
## of clusters - color vector was recycled

## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as
## of ggplot2 3.3.4.
## i The deprecated feature was likely used in the factoextra package.
## Please report the issue at <]8;https://github.com/kassambara/factoextra/issues]8;https://github.com/kassambara/factoextra/issues]8;>.
```



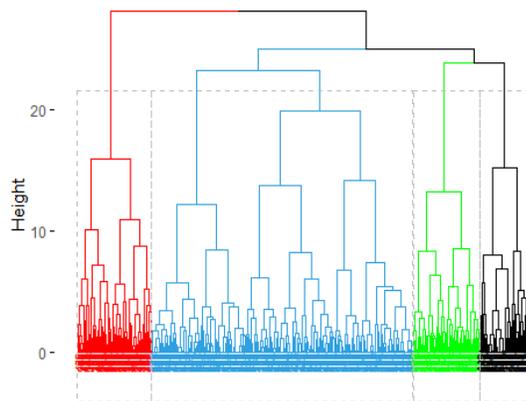
```
res2 <- hcut(df, k = 2, stand = TRUE)
fviz_dend(res3, rect = TRUE, cex = 0.5,
           k_colors = c("red", "#2E9FDF"))
```

```
## Warning in get_col(col, k): Length of color vector was shorter than the number
## of clusters - color vector was recycled
```



```
res4 <- hcut(df, k = 4, stand = TRUE)
fviz_dend(res4, rect = TRUE, cex = 0.5,
           k_colors = c("red", "#2E9FDF", "green", "black"))
```

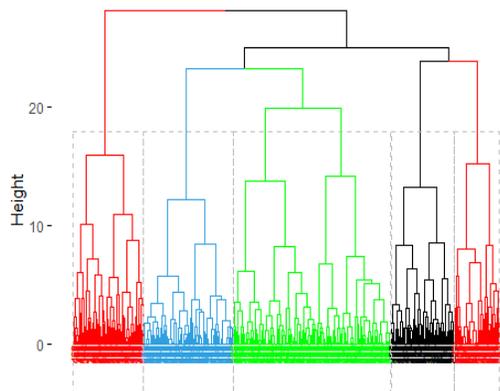
Cluster Dendrogram



```
res5 <- hcut(df, k = 5, stand = TRUE)
fviz_dend(res5, rect = TRUE, cex = 0.5,
           k_colors = c("red", "#2E9FDF", "green", "black"))
```

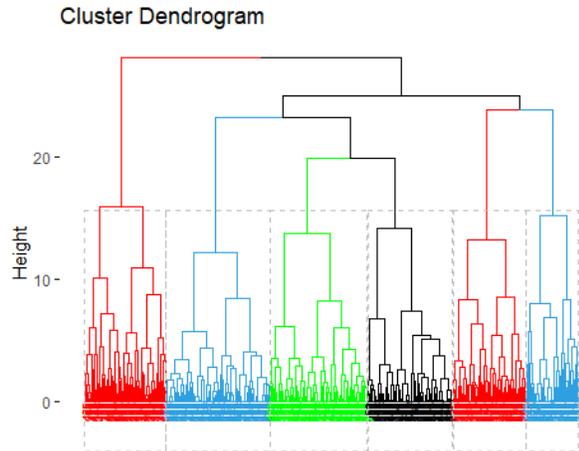
```
## Warning in get_col(col, k): Length of color vector was shorter than the number
## of clusters - color vector was recycled
```

Cluster Dendrogram



```
res6 <- hcut(df, k = 6, stand = TRUE)
fviz_dend(res6, rect = TRUE, cex = 0.5,
           k_colors = c("red", "#2E9FDF", "green", "black"))
```

```
## Warning in get_col(col, k): Length of color vector was shorter than the number
## of clusters - color vector was recycled
```



#analisis para seis #calculamos los dos clústers

```
k6 <- kmeans(df, centers = 6, nstart = 25)
k6

## K-means clustering with 6 clusters of sizes 140, 131, 39, 98, 75, 68
##
## Cluster means:
##   RESPONSABILIDAD_SOCIAL_REESCALADA CUMPLIMIENTO_SERVICIO_REESCALADA
## 1                0.06035667                0.3221094
## 2                0.20995940                0.4865972
## 3               -2.43180421               -0.3147534
## 4                0.02086990                0.3187525
## 5                0.31159526               -1.8381632
## 6                0.49221843                0.1479459
##   CONFIANZA_1_REESCALADA NATURALEZA_CONTRATO_REESCALADA
## 1                -0.7904246                -0.60662371
## 2                 0.9876868                 0.57581595
## 3                 0.1530066                -0.14639324
## 4                -0.5769208                -0.57749520
## 5                 0.5997221                 0.99000707
## 6                -0.1931722                -0.03604758
##   SEGURIDAD_ECONOMICA_2_REESCALADA REPUTACION_REESCALADA
## 1                -0.88155519                -0.4900850
## 2                -0.17908225                -0.4200008
## 3                -0.22297381                 1.1356454
## 4                 0.98595858                -0.5254942
## 5                 0.09720097                 0.1340504
## 6                 0.75969803                 1.7762718
##
## Clustering vector:
## [1] 1 4 5 3 2 5 1 5 5 5 4 5 3 6 6 1 5 1 4 2 5 5 6 4 6 1 6 1 5 4 1 1 3 5 5 5 1
## [38] 5 2 2 1 6 4 4 5 4 3 6 1 6 2 2 6 5 6 5 1 4 6 5 5 6 4 2 6 4 2 2 2 1 6 1 2 3

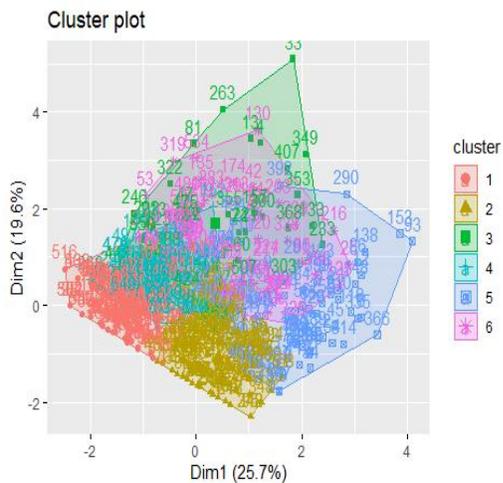
## Within cluster sum of squares by cluster:
## [1] 313.6475 293.5969 230.0803 213.8019 358.6780 234.6843
## (between_SS / total_SS = 50.2 %)
##
```

```

## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"   "size"         "iter"         "ifault"
str(k6)
## List of 9
## $ cluster      : int [1:551] 1 4 5 3 2 5 1 5 5 ...
## $ centers      : num [1:6, 1:6] 0.0604 0.21 -2.4318 0.0209 0.3116 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:6] "1" "2" "3" "4" ...
## .. ..$ : chr [1:6] "RESPONSABILIDAD_SOCIAL_REESCALADA" "CUMPLIMIENTO_SERVICIO_REES
CALADA" "CONFIANZA_1_REESCALADA" "NATURALEZA_CONTRATO_REESCALADA" ...
## $ totss       : num 3300
## $ withinss    : num [1:6] 314 294 230 214 359 ...
## $ tot.withinss: num 1644
## $ betweenss   : num 1656
## $ size        : int [1:6] 140 131 39 98 75 68
## $ iter        : int 5
## $ ifault      : int 0
## - attr(*, "class")= chr "kmeans"

fviz_cluster(k6, data = df)

```



```

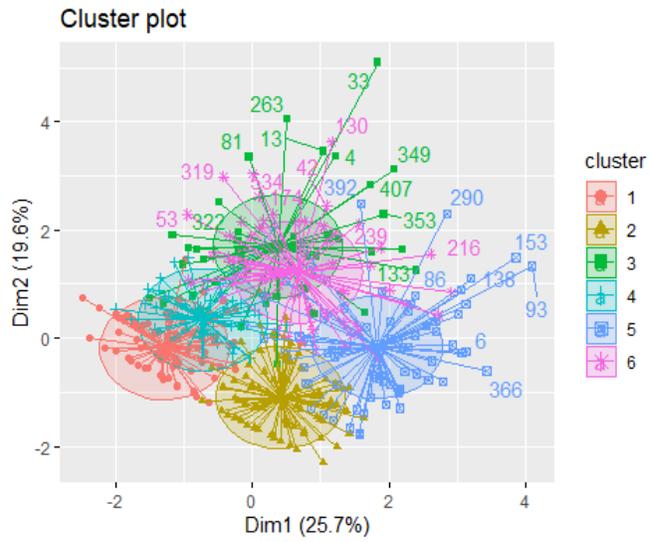
fviz_cluster(k6, data = df, ellipse.type = "euclid", repel = TRUE, star.plot = TRUE) #e
llipse.type= "t", "norm", "euclid"

```

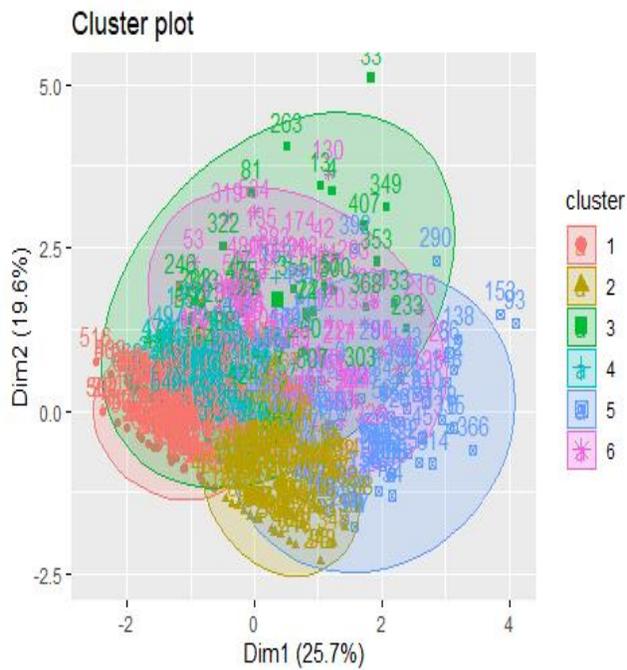
```

## Warning: ggrepel: 525 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

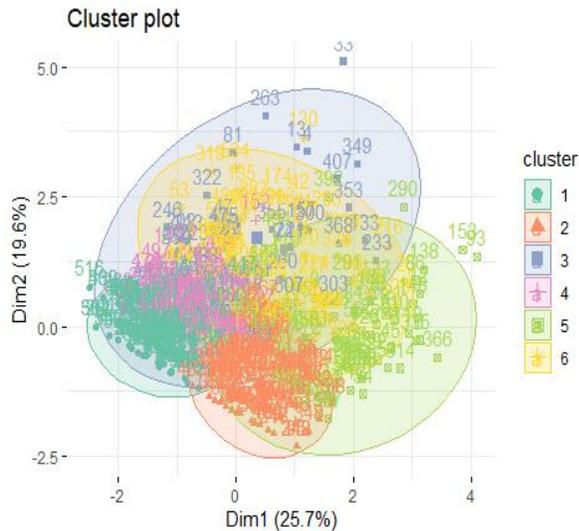
```



```
fviz_cluster(k6, data = df, ellipse.type = "norm")
```



```
fviz_cluster(k6, data = df, ellipse.type = "norm", palette = "Set2", ggtheme = theme_minimal())
```



#para 4 grupos

```
k4 <- kmeans(df, centers = 4, nstart = 25)
k4

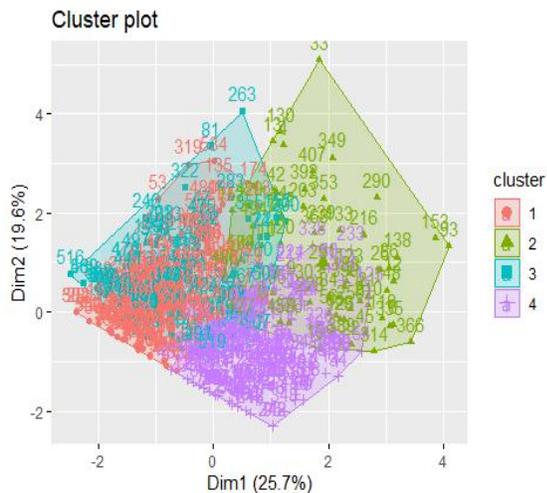
## K-means clustering with 4 clusters of sizes 211, 65, 108, 167
##
## Cluster means:
##   RESPONSABILIDAD_SOCIAL_REESCALADA CUMPLIMIENTO_SERVICIO_REESCALADA
## 1                0.4772203                0.3072977
## 2                -0.1424112               -2.1072483
## 3                -1.4792350                0.3266785
## 4                 0.4091055                0.2206589
##   CONFIANZA_1_REESCALADA NATURALEZA_CONTRATO_REESCALADA
## 1          -0.83775740          -0.5224962
## 2           0.32088984           0.5223974
## 3          -0.04096489          -0.4907570
## 4           0.96007892           0.7742072
##   SEGURIDAD_ECONOMICA_2_REESCALADA REPUTACION_REESCALADA
## 1           0.09992480          -0.1534906
## 2           0.29595620           1.0627312
## 3          -0.34621619          -0.1356899
## 4          -0.01754453          -0.1319551
##
## Clustering vector:
## [1] 3 1 2 2 4 2 1 2 2 4 1 2 2 1 2 3 4 1 1 4 2 2 4 1 2 1 4 1 2 1 1 1 2 4 4 4 1
## [38] 4 4 4 1 2 1 1 2 1 3 4 1 1 4 4 1 4 4 2 3 1 4 4 4 1 1 4 4 1 4 1 4 1 4 1 4 3
## Within cluster sum of squares by cluster:
## [1] 681.3752 445.3545 411.0363 505.9576
## (between_SS / total_SS = 38.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"       "withinss"    "tot.withinss"
## [6] "betweenss"    "size"        "iter"        "ifault"
str(k4)
```

```

## List of 9
## $ cluster      : int [1:551] 3 1 2 2 4 2 1 2 2 4 ...
## $ centers      : num [1:4, 1:6] 0.477 -0.142 -1.479 0.409 0.307 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:4] "1" "2" "3" "4"
## .. ..$ : chr [1:6] "RESPONSIBILIDAD_SOCIAL_REESCALADA" "CUMPLIMIENTO_SERVICIO_REES
CALADA" "CONFIANZA_1_REESCALADA" "NATURALEZA_CONTRATO_REESCALADA" ...
## $ totss       : num 3300
## $ withinss    : num [1:4] 681 445 411 506
## $ tot.withinss: num 2044
## $ betweenss   : num 1256
## $ size        : int [1:4] 211 65 108 167
## $ iter        : int 4
## $ ifault      : int 0
## - attr(*, "class")= chr "kmeans"

```

```
fviz_cluster(k4, data = df)
```

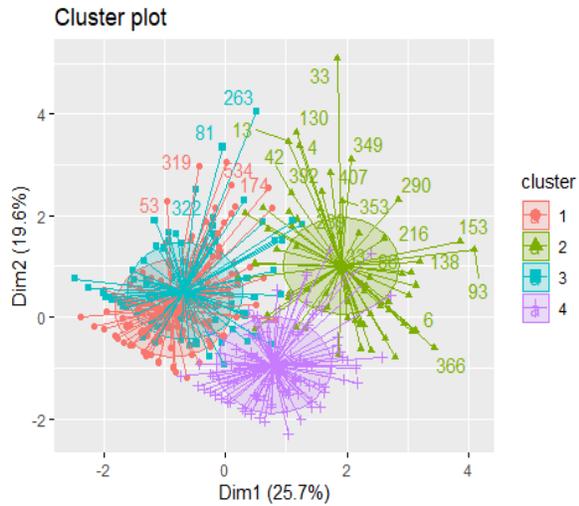


```
fviz_cluster(k4, data = df, ellipse.type = "euclid", repel = TRUE, star.plot = TRUE) #e
llipse.type= "t", "norm", "euclid"
```

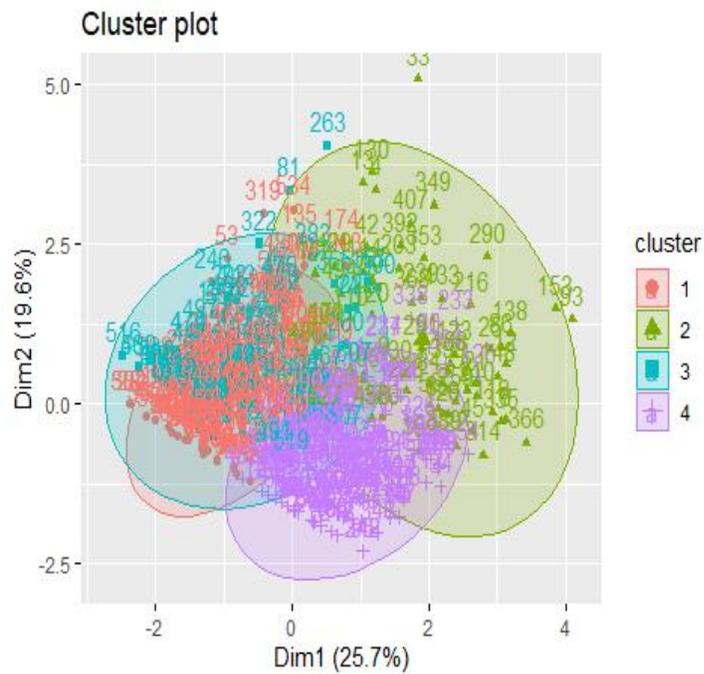
```

## Warning: ggrepel: 525 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

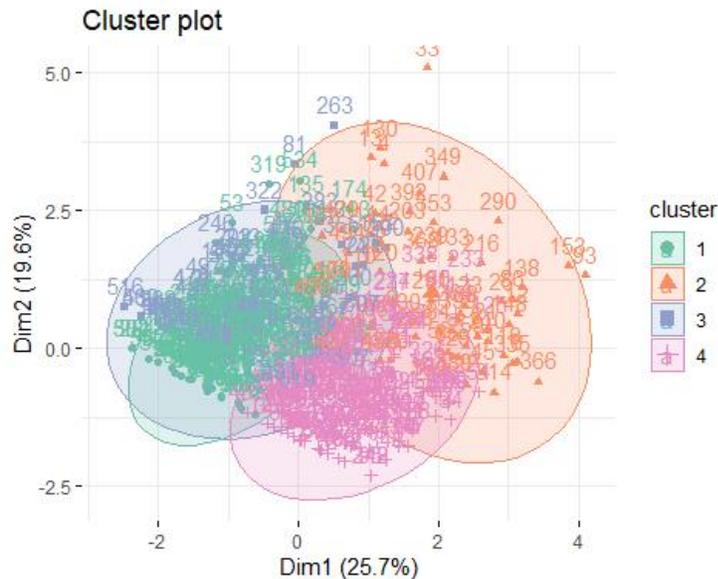
```



```
fviz_cluster(k4, data = df, ellipse.type = "norm")
```



```
fviz_cluster(k4, data = df, ellipse.type = "norm", palette = "Set2", ggtheme = theme_minimal())
```



#ANALISIS PARA 2

```
k2 <- kmeans(df, centers = 2, nstart = 25)
k2

## K-means clustering with 2 clusters of sizes 237, 314
##
## Cluster means:
##   RESPONSABILIDAD_SOCIAL_REESCALADA CUMPLIMIENTO_SERVICIO_REESCALADA
## 1                      0.03728717                  -0.207444
## 2                      -0.02814350                   0.156574
##   CONFIANZA_1_REESCALADA NATURALEZA_CONTRATO_REESCALADA
## 1                0.9365235                   0.7607448
## 2               -0.7068665                  -0.5741927
##   SEGURIDAD_ECONOMICA_2_REESCALADA REPUTACION_REESCALADA
## 1                -0.01495060                   0.09029730
## 2                 0.01128437                  -0.06815433
##
## Clustering vector:
## [1] 2 2 2 2 1 1 2 1 2 1 2 1 2 2 2 2 2 2 1 1 1 1 2 1 2 1 2 2 2 2 1 1 1 2
## [38] 1 1 1 2 2 2 2 1 2 2 1 2 2 1 1 2 2 1 1 1 2 2 1 1 2 2 1 1 2 1 2 1 2 1 1
##
## Within cluster sum of squares by cluster:
## [1] 1203.119 1469.477
## (between_SS / total_SS =  19.0 %)
##
## Available components:
##
## [1] "cluster"       "centers"        "totss"          "withinss"       "tot.withinss"
## [6] "betweenss"     "size"           "iter"           "ifault"
```

```
str(k2)

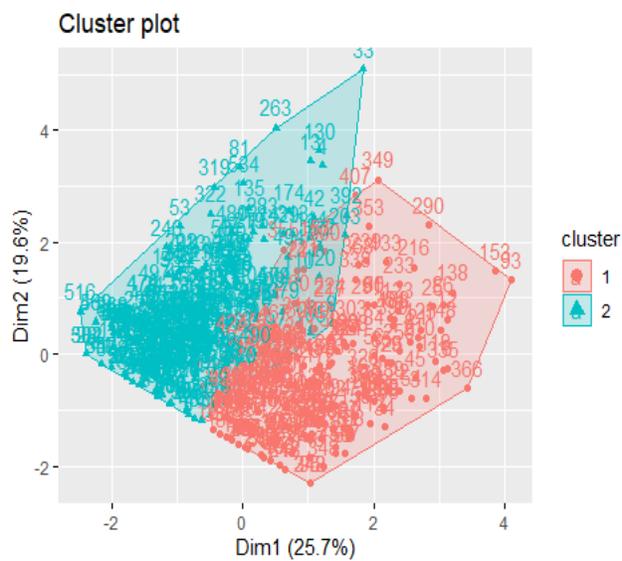
## List of 9
## $ cluster      : int [1:551] 2 2 2 2 1 1 2 1 2 1 2 1 ...
```

```

## $ centers      : num [1:2, 1:6] 0.0373 -0.0281 -0.2074 0.1566 0.9365 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:2] "1" "2"
## .. ..$ : chr [1:6] "RESPONSIBILIDAD_SOCIAL_REESCALADA" "CUMPLIMIENTO_SERVICIO_REES
CALADA" "CONFIANZA_1_REESCALADA" "NATURALEZA_CONTRATO_REESCALADA" ...
## $ totss       : num 3300
## $ withinss    : num [1:2] 1203 1469
## $ tot.withinss : num 2673
## $ betweenss   : num 627
## $ size        : int [1:2] 237 314
## $ iter        : int 1
## $ ifault      : int 0
## - attr(*, "class")= chr "kmeans"

```

```
fviz_cluster(k2, data = df)
```



```

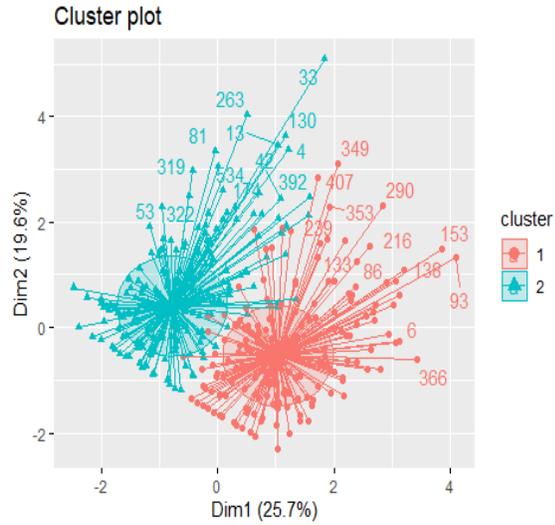
fviz_cluster(k2, data = df, ellipse.type = "euclid", repel = TRUE, star.plot = TRUE) #e
llipse.type= "t", "norm", "euclid"

```

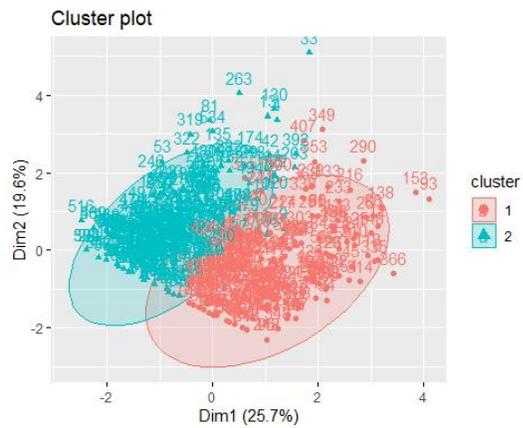
```

## Warning: ggrepel: 525 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

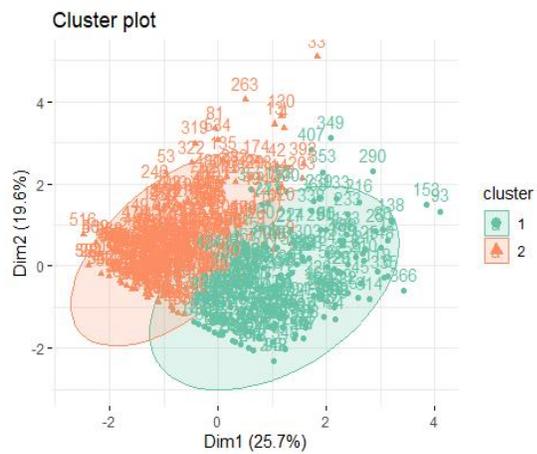
```



```
fviz_cluster(k2, data = df, ellipse.type = "norm")
```



```
fviz_cluster(k2, data = df, ellipse.type = "norm", palette = "Set2", ggtheme = theme_minimal())
```



##quedarme con 3

```

df_T %>%
  mutate(Cluster = k3$cluster) %>%
  group_by(Cluster) %>%
  summarise_all("mean")

## Warning in mean.default(NOMBRE): argument is not numeric or logical: returning
## NA

## Warning in mean.default(NOMBRE): argument is not numeric or logical: returning
## NA

## Warning in mean.default(NOMBRE): argument is not numeric or logical: returning
## NA

## Warning in mean.default(TIPO_PERSONA): argument is not numeric or logical:
## returning NA

## Warning in mean.default(TIPO_PERSONA): argument is not numeric or logical:
## returning NA

## Warning in mean.default(TIPO_PERSONA): argument is not numeric or logical:
## returning NA

## # A tibble: 3 × 11
##   Cluster NUM_P...1 NOMBRE TIPO...2 RESPO...3 CUMPL...4 CONFI...5 NATUR...6 SEGUR...7 REPUT...8
##   <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1  10180.    NA    NA  0.845  0.952  0.393  0.336  0.415  0.0287
## 2     2   8931.    NA    NA  0.886  0.851  0.892  0.648  0.424  0.0486
## 3     3   7782.    NA    NA  0.8    0.780  0.607  0.516  0.640  0.649
## # ... with 1 more variable: PUNTAJE_FINAL <dbl>, and abbreviated variable names
## #   1NUM_PROVEEDOR, 2TIPO_PERSONA, 3RESPONBILIDAD_SOCIAL_REESCALADA,
## #   4CUMPLIMIENTO_SERVICIO_REESCALADA, 5CONFIANZA_1_REESCALADA,
## #   6NATURALEZA_CONTRATO_REESCALADA, 7SEGURIDAD_ECONOMICA_2_REESCALADA,
## #   8REPUTACION_REESCALADA

df <- df_T
df

## # A tibble: 551 × 10
##   NUM_PROVEEDOR NOMBRE TIPO...1 RESPO...2 CUMPL...3 CONFI...4 NATUR...5 SEGUR...6 REPUT...7
##   <dbl> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1 CONSTR... JURIDI...  0.8    1    0.571  0.327  0    0
## 2    19 ACONDI... JURIDI...  1      1    0.429  0.154  1    0
## 3    20 MUEBLE... JURIDI...  1      0.6  0.286  0.731  1    0.2
## 4    53 ASESOR... JURIDI...  0.4    0.4  0.714  0.154  1    0.72
## 5    82 AGENCI... JURIDI...  1      1    0.571  0.712  0.35  0.2
## 6    98 COMPAÑ... JURIDI...  1      0    1      0.731  1    0
## 7   107 C R CO... JURIDI...  1      0.6  0.286  0.462  0.35  0
## 8   110 CENTRO... JURIDI...  1      0.6  0.857  0.846  0.35  0.72
## 9   122 FORMUL... JURIDI...  1      0.6  0      1      1    0.28
## 10  125 R.R. D... JURIDI...  1      0.6  0.857  0.519  0.35  0
## # ... with 541 more rows, 1 more variable: PUNTAJE_FINAL <dbl>, and abbreviated
## #   variable names 1TIPO_PERSONA, 2RESPONBILIDAD_SOCIAL_REESCALADA,
## #   3CUMPLIMIENTO_SERVICIO_REESCALADA, 4CONFIANZA_1_REESCALADA,

```

```

## #      5NATURALEZA_CONTRATO_REESCALADA, 6SEGURIDAD_ECONOMICA_2_REESCALADA,
## #      7REPUTACION_REESCALADA

df$clus<-as.factor(k3$cluster)
df

## # A tibble: 551 × 11
##   NUM_PROVEEDOR NOMBRE TIPO_...1 RESPO...2 CUMPL...3 CONFI...4 NATUR...5 SEGUR...6 REPUT...7
##   <dbl> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1 CONSTR... JURIDI... 0.8 1 0.571 0.327 0 0
## 2 19 ACONDI... JURIDI... 1 1 0.429 0.154 1 0
## 3 20 MUEBLE... JURIDI... 1 0.6 0.286 0.731 1 0.2
## 4 53 ASESOR... JURIDI... 0.4 0.4 0.714 0.154 1 0.72
## 5 82 AGENCI... JURIDI... 1 1 0.571 0.712 0.35 0.2
## 6 98 COMPAÑ... JURIDI... 1 0 1 0.731 1 0
## 7 107 C R CO... JURIDI... 1 0.6 0.286 0.462 0.35 0
## 8 110 CENTRO... JURIDI... 1 0.6 0.857 0.846 0.35 0.72
## 9 122 FORMUL... JURIDI... 1 0.6 0 1 1 0.28
## 10 125 R.R. D... JURIDI... 1 0.6 0.857 0.519 0.35 0
## # ... with 541 more rows, 2 more variables: PUNTAJE_FINAL <dbl>, clus <fct>, and
## # abbreviated variable names 1TIPO_PERSONA, 2RESPONBILIDAD_SOCIAL_REESCALADA,
## # 3CUMPLIMIENTO_SERVICIO_REESCALADA, 4CONFIANZA_1_REESCALADA,
## # 5NATURALEZA_CONTRATO_REESCALADA, 6SEGURIDAD_ECONOMICA_2_REESCALADA,
## # 7REPUTACION_REESCALADA

df$clus<-factor(df$clus)
data_long <- gather(df, caracteristica, valor, RESPONBILIDAD_SOCIAL_REESCALADA:PUN
AJE_FINAL, factor_key=TRUE)
data_long

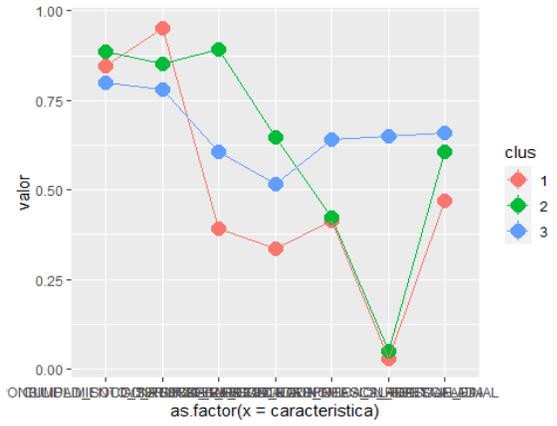
## # A tibble: 3,857 × 6
##   NUM_PROVEEDOR NOMBRE TIPO_...1 clus caract...2 valor
##   <dbl> <chr> <chr> <fct> <fct> <dbl>
## 1 1 CONSTRUCTORA CAMPOS S.A JURIDI... 1 RESPON... 0.8
## 2 19 ACONDICIONAMIENTO DE OFICINAS SOCI... JURIDI... 1 RESPON... 1
## 3 20 MUEBLES METALICOS ALVARADO SOCIEDA... JURIDI... 3 RESPON... 1
## 4 53 ASESORIA EN ELECTRONICA Y COMPUTAC... JURIDI... 3 RESPON... 0.4
## 5 82 AGENCIAS BASICAS MERCANTILES (A B ... JURIDI... 2 RESPON... 1
## 6 98 COMPAÑIA TECNICA Y COMERCIAL SATEC... JURIDI... 2 RESPON... 1
## 7 107 C R CONECTIVIDAD SOCIEDAD ANONIMA JURIDI... 1 RESPON... 1
## 8 110 CENTRO CARS SOCIEDAD ANONIMA JURIDI... 3 RESPON... 1
## 9 122 FORMULARIOS ESTANDARD COSTA RICA SA JURIDI... 3 RESPON... 1
## 10 125 R.R. DONNELLEY DE COSTA RICA SOCIE... JURIDI... 2 RESPON... 1
## # ... with 3,847 more rows, and abbreviated variable names 1TIPO_PERSONA,
## # 2caracteristica

ggplot(data_long, aes(as.factor(x = caracteristica), y = valor,group=clus, colour = c
lus)) +
  stat_summary(fun = mean, geom="pointrange", size = 1)+
  stat_summary(geom="line")

## No summary function supplied, defaulting to `mean_se()`

## Warning: Removed 21 rows containing missing values (`geom_segment()`).

```



```
#geom_point(aes(shape=clus))
```

```
library(openxlsx)  
write.xlsx(df, "DF_cluster.xlsx")
```