HACIENDA
SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO

CNSF
COMISIÓN NACIONAL DE SEGUROS Y FIANZAS

DIRECCIÓN GENERAL DE DESARROLLO E INVESTIGACIÓN

DIRECCIÓN DE ASUNTOS ACTUARIALES

PROCEDIMIENTOS PARA EL SEMÁFORO DE CALIDAD DE LA INFORMACIÓN ESTÁDISTICA

15 de noviembre 2023

Contenido

Introducción 3
Capítulo 1. Generalidades 4
1.1 Objetivo 4
1.2 Antecedentes 4
Capítulo 2. Base de datos 6
2.1 Fuente de información 6
2.2 Base de datos 6
Capítulo 3. Marco teórico 7
3.1. Validaciones 8
3.2 Clustering 12
3.2.1 Algoritmo K-means 12
3.2.2 Método Elbow 13
3.3 Asignación de calificación 13
Capítulo 4. Marco metodológico 15
4.1 Ejemplo práctico para el Semáforo de Calidad de la Información Estadística 15

Introducción

En el marco del Sector de Seguros y Fianzas en nuestro país, la evaluación y mejora continua de la calidad de la información estadística entregada por las compañías se ha convertido en un aspecto crucial. Con el propósito de abordar esta necesidad, se desarrolló un índice que no solo mide la precisión y confiabilidad de los datos proporcionados en el Reporte Regulatorio sobre Información Estadística (RR-8), sino que también asigna una calificación a cada institución.

El objetivo principal de este índice es impactar de manera positiva en el riesgo reputacional de cada entidad, proporcionando una evaluación transparente y objetiva de la calidad de su información. En este contexto, la calificación obtenida por cada institución no solo reflejará su compromiso con la exactitud de los datos, sino que también influirá en cómo son percibidas tanto por el sector como por el público en general.

A través de esta iniciativa, se busca fomentar la entrega de información confiable, verídica y de alta calidad. Este índice no solo beneficiará a las instituciones públicas y al sector asegurador y afianzador, sino que también proporcionará datos fundamentales para la generación de estudios e indicadores sobre el dinamismo mercado de Seguros y Fianzas en nuestro país.

Capítulo 1. Generalidades

1.1 Objetivo

Desarrollar un índice que evalúe la calidad de la información estadística proporcionada por las compañías de seguros y fianzas del país en el Reporte Regulatorio sobre Información Estadística (RR-8). Con el fin de divulgar la calificación de cada institución para impactar en su riesgo reputacional, reflejando la calidad de su información.

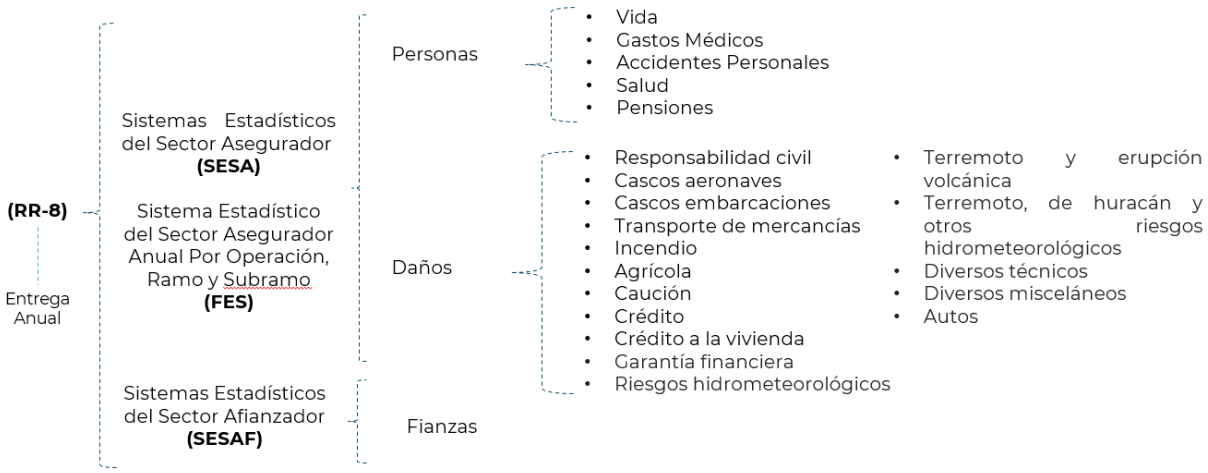
Este índice tiene como finalidad obtener información confiable y verídica, utilizada por instituciones públicas, el sector asegurador, afianzador y el público en general. Esto permitirá generar estudios e indicadores sobre el mercado de seguros y fianzas del país.

1.2 Antecedentes

De acuerdo con la Circular única de Seguros y Fianzas (CUSF) las instituciones aseguradoras y afianzadoras del país deben presentar alguno de los 13 Reportes Regulatorios dependiendo de la actividad de la institución, donde las Instituciones y Sociedades Mutualistas, así como las demás personas y entidades que en los términos de esta Ley están sujetas a inspección y vigilancia por parte de la Comisión Nacional de Seguros y Fianzas.

El área de Asuntos Actuariales se encarga de revisar el Reporte Regulatorio sobre Información Estadística (RR-8) donde se solicita instituciones aseguradoras y afianzadoras del país información relativa a los Sistemas Estadísticos del Sector Asegurador (SESA), los Sistemas Estadísticos del Sector Afianzador (SESAF), Información Estadística por Operación, Ramo o Seguro (ORS), Información Estadística por Ramo y Subramo (RyS) y el Sistema Estadístico del Sector Asegurador Anual Por Operación, Ramo Y Subramo (FES).

Dentro de cada Reporte Regulatorio la entrega de información referente a los ramos que se muestran a continuación se entrega de manera anual a la Comisión Nacional de Seguros y Fianzas.



Mientras que la Información Estadística por Operación, Ramo o Seguro (ORS) e Información Estadística por Ramo y Subramo (RyS) se entregan de manera mensual, es un resumen acumulativo de cómo se encuentra la información estadística de cada institución.

Capítulo 2. Base de datos

2.1 Fuente de información

La información del Reporte Regulatorio sobre Información Estadística (RR-8) se presentará conforme a lo señalado en los Capítulos 39.1, 39.2 y 39.3 de Circular Única de Seguros y Fianzas, deben cumplir con los aspectos generales señalados en el Anexo 38.1.9-a de la Circular Única de Seguros y Fianzas.

La base de datos se compone de información estadística que entregan las compañías en archivo txt dentro de las disposiciones mencionadas en el párrafo anterior y la entregas se realiza de la siguiente manera:

- **Sistemas Estadísticos del Sector Asegurador (SESA)** se entrega de forma anual, la información se entrega conforme a lo establecido en la ley de instituciones de seguros y fianzas en el capítulo 38.1.9.
- **Sistemas Estadísticos del Sector Afianzador (SESAF)** se entrega de forma anual, la información deberá presentarse dentro de los veintinueve días hábiles siguientes al cierre del ejercicio de que se trate.
- **Información Estadística por Operación, Ramo o Seguro (ORS)** se entrega de forma mensual, la información deberá presentarse dentro de los primeros veintidós días hábiles siguientes al cierre del mes de que se trate.
- **Información Estadística por Ramo y Subramo (RyS)** se entrega de forma mensual, la información deberá presentarse dentro de los primeros veintidós días hábiles siguientes al cierre del mes de que se trate.
- **Sistema Estadístico del Sector Asegurador Anual Por Operación, Ramo Y Subramo (FES)** se entrega de forma anual, la información deberá presentarse dentro de los primeros sesenta y cinco días hábiles siguientes al cierre del ejercicio de que se trate.

2.2 Base de datos

La información estadística entregada por las compañías mediante archivos txt se concentran en un servidor de la Comisión Nacional de Seguros y Fianzas.

Capítulo 3. Marco teórico

Para revisar la calidad de la información, se construye un indicador basado en un credit scoring, el cual es definido por Medina y Martí (2012) como un sistema de evaluación crediticia que permite valorar de forma automática el riesgo asociado a cada solicitud de crédito. Se utiliza como referencia el credit scoring debido a que proporciona una calificación a los créditos según el riesgo asociado, y para el caso de medir la calidad de la información entregada, se asignará una calificación de manera similar dependiendo de las inconsistencias de la información estadística entregada.

En un credit scoring, a medida que el puntaje obtenido es más alto, el cliente es considerado mejor para pagar. En el caso del indicador del Semáforo de Calidad de la Información Estadística, un puntaje más elevado indica que la institución aseguradora entregó su información estadística con más inconsistencias; es decir, a mayor score, peor es la calidad de la información estadística entregada.

Para asignar puntaje a la información estadística entregada, se dividen las validaciones en grados con la finalidad de clasificarlas según la importancia de la información proporcionada por las instituciones. Los grados son los siguientes:

Grado 1	Validaciones que sólo involucran variables categóricas[1].
Grado 2	Validaciones que involucran variables categóricas y variables de montos.
Grado 3	Validaciones que únicamente involucran variables de montos.
Factor	Validaciones que revisan que “Otros” no sea utilizado más del 5% de la población total.

Los grados se aplican a las validaciones catalogadas en 4 grupos principales:

1. Validaciones consideradas como tolerantes.
2. Validaciones contables.
3. Validaciones con porcentaje menor al 5% de su población total.
4. Validaciones de consistencia de catálogos entre ejercicios reportados.

[1] Una variable categórica es un tipo de variable estadística que representa diferentes categorías o grupos discretos, generalmente identificados por etiquetas o nombres.

3.1. Validaciones

1. Validaciones consideradas como tolerantes

La forma de evaluar es mediante validaciones de tolerancia cero y validaciones consideradas como tolerantes, las de tolerancia cero son validaciones básicas que se encuentran en el sistema donde entregan las instituciones mediante el cual no se deja entrar la información si no se pasan estas validaciones, para el manual del Semáforo de la Calidad de la Información Estadística el enfoque será únicamente de las validaciones que no son de tolerancia cero.

Las validaciones que no son de tolerancia cero son validaciones creadas para revisar la información que entregan las instituciones aseguradoras y se modifica respecto a experiencia que se va obteniendo año con año y a los casos especiales de las instituciones. A estas validaciones que no son de tolerancia cero se les dará un grado dependiendo de la importancia de la validación.

Las validaciones sin tolerancia cero se suministran en un archivo Excel por el área de Asuntos Actuariales y varían según el ramo que se esté revisando.

La forma de calcular el porcentaje de error de las validaciones consideradas como tolerantes para dar un peso a cada validación es la siguiente:

$$Porcentaje\ de\ error = \frac{Total\ de\ errores\ comentidos\ por\ la\ compañía_i}{Total\ de\ registros\ reportados\ por\ la\ compañía} * 100$$

Donde i es la validación revisada y va de 1, 2, 3..., n dependiendo del ramo.

2. Validaciones contables

Otras validaciones empleadas en el Semáforo de Calidad de la Información Estadística se orientan hacia el Reporte Regulatorio sobre Estados Financieros (RR-7), partiendo del supuesto que ambas variables deben de coincidir debido a que tienen la misma definición y periodicidad. En este contexto, las validaciones consisten en contrastar la información reportada en el RR-7 con el RR-8 para asegurar la congruencia en las siguientes variables:

- Prima emitida
- Prima cedida
- Comisión directa
- Monto del siniestro

- Monto de vencimientos
- Monto de rescate
- Gastos de ajuste
- Monto de dividendos
- Monto de salvamento
- Monto de recuperaciones de terceros
- Monto recuperado de reaseguro

La forma de calcular el porcentaje de error de las validaciones contables es la siguiente:

$$Porcentaje\ de\ error = (1 - \frac{\min(Monto\ contable\ RR8_i, Monto\ contable\ RR7_i)}{\max(Monto\ contable\ RR8_i, Monto\ contable\ RR7_i)}) * 100$$

Donde i es la validación de monto revisada, i va de 1, 2, 3..., n dependiendo del ramo.

3. Validaciones con porcentaje "Otros" mayor al 5% de su población total

Dentro de las variables solicitadas a las instituciones, algunas son variables categóricas y tienen la opción de catalogar la información con la etiqueta de "Otros". Se debe tener cuidado para que las instituciones no cataloguen la mayoría de su información estadística con la etiqueta "Otros". Para evitarlo, se aplica la siguiente regla:

$$Porcentaje\ de\ otros = \frac{Total\ de\ registros\ catalogados\ como\ OTROS_i}{Total\ de\ registros\ reportados\ por\ la\ compañía} * 100$$

Donde i es la validación de catálogo con opción de "OTROS" y va de 1, 2, 3..., n dependiendo del ramo.

Porcentaje de error
- Si x% > 5%, el puntaje es distinto de 0.
- Si x% <= 5%, el puntaje es 0.

Donde x es el porcentaje obtenido en el porcentaje de otros.

4. Validaciones de consistencia de catálogos entre ejercicios reportados.

Las validaciones de consistencia de catálogos entre ejercicios reportados se refieren a revisar que lo reportado en las categorías de validación sea consistente entre los ejercicios reportados. Es decir, se verifica que la información reportada en el ejercicio actual, para cada catálogo y cada registro, sea igual a la reportada en el ejercicio anterior para cada catálogo y cada registro.

$$Catálogo_t = Catálogo_{t-1}$$

Donde t es el periodo del ejercicio reportado.

Se valida la información para cada registro reportado durante el periodo del ejercicio actual.

Nota: Se excluyen variables que, por su naturaleza, pueden sufrir modificaciones en la clave de su catálogo.

El peso de las validaciones de consistencia entre catálogos se asigna mediante el porcentaje de error entre lo reportado en el ejercicio actual y lo reportado en el ejercicio anterior para cada registro en las variables de catálogo.

$$Porcentaje\ de\ error = \frac{(Catálogo_t \neq Catálogo_{t-1})_i}{Total\ de\ registros\ reportados\ por\ la\ compañía} * 100$$

Donde i es la validación de consistencia entre catálogos y va de 1, 2, 3..., n dependiendo del ramo.

Donde t es el ejercicio reportado.

Teniendo en cuenta el porcentaje de error en todas las validaciones, se procede a realizar escenarios para definir los puntajes según el porcentaje de error de la validación. Se lleva a cabo un análisis de frecuencias y se parten los porcentajes de error en cuartiles para observar el comportamiento de los porcentajes de error. A partir de esto, se establecen cuatro cortes, es decir, el porcentaje de error se divide en cuatro, siendo cero un porcentaje excelente, ya que la institución no presentaría errores en su información, y 100% lo peor, ya que indicaría que la institución tendría toda su información estadística errónea.

Considerando los grados y la distribución de los porcentajes de error quedan los siguientes puntos de corte:

Grado 1

Rango de error	**Puntaje**
>= 0%, <= 1%	0
> 1%, <= 5%	1
> 5%, <= 10%	2
> 10%	9

Grado 2

Rango de error	**Puntaje**
>= 0%, <= 1%	0
> 1%, <= 5%	10
> 5%, <= 10%	20
> 10%	90

Para los grados 1 y 2, se considera una tolerancia de hasta el 10% en el porcentaje de error, debido a temas de calidad de la información.

Grado 3

Rango de error	**Puntaje**
>= 0%, <= 0.1%	0
> 0.1%, <= 0.5%	50
> 0.5%, <= 6%	150
> 6%	800

En el caso de las variables de grado 3, es crucial que las instituciones entreguen correctamente la información estadística. Por lo tanto, si el porcentaje de error es mayor al 6%, se le asigna el puntaje más alto.

Para las validaciones con un porcentaje de "Otros" mayor al 5% de su población total, se asigna un puntaje de 90 puntos. Si el porcentaje de "Otros" es menor al 5% de su población total, se otorga un puntaje de 0 puntos.

Después de asignar un puntaje a cada validación, se procede a realizar la suma de los puntajes institución por institución, obteniendo como resultado un puntaje final.

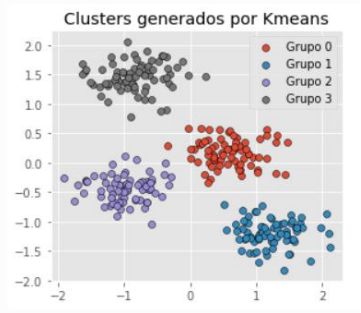
$$\sum_{i=1}^{n} Validación_i$$

Donde i va de 1 hasta n validaciones que existen en cada ramo.

El score de cada compañía es introducido al modelo se segmentación clustering para revisar el número de agrupaciones óptimo.

3.2 Clustering

Una técnica comúnmente empleada para identificar patrones o grupos dentro de un conjunto de observaciones son los clústeres. Se trata de un modelo de segmentación que organiza un conjunto de datos en grupos según características similares, al mismo tiempo que separa aquellos datos que carecen de similitudes en sus características. Este proceso resulta en la asignación de etiquetas a los registros, indicando el grupo al que pertenecen.



3.2.1 Algoritmo K-means

El algoritmo K-means agrupa las observaciones de un número predefinido de K clúster, donde la suma de las varianzas debe ser lo menor posible.

Considérese $C_1, \dots, C_k$ como los conjuntos formados por los índices de las observaciones de cada uno de los clústeres.

Ejemplo:

$$C_1: índices\ de\ las\ obsevaciones\ agrupadas\ pertenencen\ al\ clúster\ 1$$

$$C_2: índices\ de\ las\ obsevaciones\ agrupadas\ pertenencen\ al\ clúster\ 2$$

$$\dots$$

$$C_k: índices\ de\ las\ obsevaciones\ agrupadas\ pertenencen\ al\ clúster\ k$$

Todos los conjuntos deben satisfacer lo siguiente:

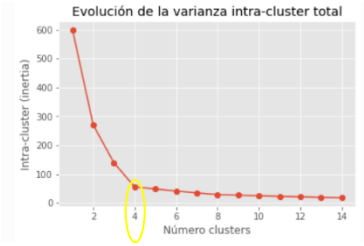
$C_1 \cup C_2 \cup \dots \cup C_k = \{1, \dots, k\}$ Toda observación pertenece a uno de los k clúster.

$C_k \cap C_{k'} \neq \emptyset \; para\; todo\; k \neq k'$ Ninguna observación pertenece a más de un clúster a la vez.

Para cuantificar la varianza interna de cada clúster se emplea la suma de las distancias euclídeas al cuadrado entre cada observación (x_i) y el centroide (μ) de su clúster.

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

El algoritmo de K-means se asegura de reducir la varianza interna del clúster en cada paso hasta alcanzar un óptimo local. Sin embargo, dado que no evalúa todas las posibles distribuciones de las observaciones, sino solo una parte de ellas es importante ejecutar el algoritmo varias veces, cada vez con una asignación aleatoria inicial diferente, y elegir aquella que logre la menor varianza total.



3.2.2 Método Elbow

Un enfoque popularmente conocido como el método del codo se utiliza como medida para determinar el número óptimo de clústeres en un conjunto de datos. Este método implica la representación gráfica de los resultados obtenidos, permitiendo identificar el punto en la curva (conocido como 'codo') donde la mejora en la formación de clústeres deja de ser significativa.

El método del codo calcula la variabilidad total intra-clúster en relación con el número de clústeres, seleccionando como óptimo aquel valor a partir del cual agregar más clústeres apenas resulta en mejoras sustanciales.

Con el método de elbow podemos definir los clústeres óptimos para el conjunto de datos, en la imagen se muestra que el codo es en el número 4, entonces podemos definir que en el conjunto de datos el número óptimo de clúster es 4

3.3 Asignación de calificación

El modelo de segmentación clustering arrojó que lo óptimo es tener 4 clúster, es decir, 4 segmentos. Dado lo anterior se procede a segmentar a las compañías en 4 grupos y se otorga una calificación: Confiable, Aceptable, Poco confiable y Deficiente.

Calificación del Semáforo de Calidad de la Información Estadística	
SCORE	**Calificación**
>= 0, <= 10	Confiable

> 10, <= 171	Aceptable
> 171, <= 890	Poco Confiable
> 890	Deficiente

Aquellas instituciones que no entregan la información estadística serán catalogadas con calificación "Sin entrega", el siguiente cuadro presenta la calificación del Semáforo de Calidad de la Información Estadística con una descripción del significado de cada calificación.

El siguiente cuadro muestra que significa cada calificación del Semáforo de Calidad de la Información Estadística.

Semáforo de Calidad de la Información Estadística	**Descripción**
Confiable	La información entregada por parte de la compañía no presenta errores o éstos no son significativos para el análisis de la información.
Aceptable	La información entregada por parte de la compañía presenta errores poco significativos.
Poco confiable	La información entregada por parte de la compañía presenta errores en variables de importancia en su reporte. Se debe tomar con precaución para realizar análisis estadísticos.
Deficiente	La información entregada por parte de la compañía presenta grandes inconsistencias en los reportes entregados a esta Comisión. Se recomienda no utilizar esta información para fines estadísticos.
Sin entrega	La compañía no cumple con la entrega de información.

Capítulo 4. Marco metodológico

4.1 Ejemplo práctico para el Semáforo de Calidad de la Información Estadística

Cada ramo cuenta con validaciones específicas como ejemplo se tomará la información estadística de una compañía con la finalidad de explicar el proceso del Semáforo de Calidad de la Información Estadística.

1. Validaciones consideradas como tolerantes

Para las validaciones consideradas como tolerantes se revisa un Excel propuesto, y se encuentran validaciones como:

- La fecha de fin de vigencia debe ser mayor a la fecha de emisión
- La diferencia entre la fecha de fin de vigencia e inicio de vigencia debe ser menor o igual a 5 años.

Utilizando una compañía revisamos que la diferencia entre la fecha de fin de vigencia e inicio de vigencia sea menor o igual a 5 años, la base tiene 81 registros de los cuáles 10 tienen este error, siendo la diferencia mayor a 5 años, así que para obtener el porcentaje se utiliza la fórmula propuesta en el apartado 3.1. Validaciones del capítulo anterior.

$$Porcentaje\ de\ error = \frac{Total\ de\ errores\ comentidos\ por\ la\ compañía_i}{Total\ de\ registros\ reportados\ por\ la\ compañía} * 100$$

$$Porcentaje\ de\ error = \frac{10}{81} * 100$$

$$Porcentaje\ de\ error = 12.34\%$$

Para la validación 1 el porcentaje de error es de 12.34%, y de acuerdo con las siguientes definiciones la validación de “diferencia entre la fecha de fin de vigencia e inicio de vigencia sea menor o igual a 5 años” depende únicamente de variables categóricas por lo tanto será de grado 1.

Grado 1	Validaciones que sólo involucran variables categóricas.
Grado 2	Validaciones que involucran variables categóricas y variables de montos.
Grado 3	Validaciones que únicamente involucran variables de montos.

Factor	Validaciones que revisan que la etiqueta “Otros” no sea utilizado más del 5% de la población total.

De acuerdo con la tabla de puntajes de grado 1, la validación tiene un porcentaje de error de 12.34% y el puntaje correspondiente es de 9 puntos.

Grado 1	
Rango del porcentaje de error	**Puntaje**
>= 0%, <= 1%	0
> 1%, <= 5%	1
> 5%, <= 10%	2
> 10%	9

Se realiza el proceso para todas las validaciones consideradas como tolerantes.

2. Validaciones contables

En cuanto a las validaciones contables se obtiene el porcentaje de error con la siguiente formula

$$Porcentaje\ de\ error = (1 - \frac{\min(Monto\ contable\ RR8_i, Monto\ contable\ RR7_i)}{\max(Monto\ contable\ RR8_i, Monto\ contable\ RR7_i)}) * 100$$

Por ejemplo, la compañía reporta una póliza con prima emitida de $23,000 pesos en el reporte regulatorio RR8, mientras que en el reporte regulatorio RR7 se reportó la cantidad de $18,000 pesos. La diferencia entre reportes partiendo supuesto que ambas variables deben de coincidir debido a que tienen la misma definición y periodicidad es de $ 5,000 pesos, se aplica la fórmula para obtener el porcentaje de error:

$$Porcentaje\ de\ error = (1 - \frac{\min(\$23{,}000, \$18{,}000)}{\max(\$23{,}000, \$18{,}000)}) * 100$$

$$Porcentaje\ de\ error = (1 - \frac{\$\ 18{,}000}{\$\ 23{,}000}) * 100$$

$$Porcentaje\ de\ error = (1 - 0.78) * 100$$

$$Porcentaje\ de\ error = (0.22) * 100$$

$$Porcentaje\ de\ error = 22\%$$

De acuerdo con la tabla de grados, la prima emitida es únicamente variable de montos por lo tanto es de grado 3, y la siguiente tabla muestra el puntaje que le corresponde a la validación, debido a que el porcentaje de error es del 22% el puntaje dado es de 800 puntos.

Grado 3	
Rango de error	**Puntaje**
>= 0%, <= 0.1%	0
> 0.1%, <= 0.5%	50
> 0.5%, <= 6%	150
> 6%	800

Se aplica el proceso para todas las validaciones contables:

- Prima emitida
- Prima cedida
- Comisión directa
- Monto del siniestro
- Monto de vencimientos
- Monto de rescate
- Gastos de ajuste
- Monto de dividendos
- Monto de salvamento
- Monto de recuperaciones de terceros
- Monto recuperado de reaseguro

3. Validaciones con porcentaje “Otros” mayor al 5% de su población total

Las validaciones con porcentaje “Otros” mayor al 5% de su población total se obtiene aplicando la regla:

$$Porcentaje\ de\ otros = \frac{Total\ de\ registros\ catalogados\ como\ OTROS_i}{Total\ de\ registros\ reportados\ por\ la\ compañía} * 100$$

$$Porcentaje\ de\ otros = \frac{55}{81} * 100$$

$$Porcentaje\ de\ otros = 0.6790 * 100$$

$$Porcentaje\ de\ otros = 67.90\%$$

Sustituimos x con el 67.90% para obtener el porcentaje de error

Porcentaje de error
- Si x% > 5%, el puntaje es distinto de 0.
- Si x% <= 5%, el puntaje es 0.

Quedando lo siguiente:

Porcentaje de error
- **Si 67.90% > 5%, el puntaje es distinto de 0.**
- Si 67.90% <= 5%, el puntaje es 0.

Se cumple que el 67.90% es mayor al 5% permitido por lo que se aplica un puntaje de 90 puntos.

4. Validaciones de consistencia de catálogos entre ejercicios reportados.

Para las validaciones de consistencia se utiliza la siguiente fórmula:

$$Porcentaje\ de\ error = \frac{(Catálogo_t \neq\ Catálogo_{t-1})_i}{Total\ de\ registros\ reportados\ por\ la\ compañía} * 100$$

En el ejemplo, la institución reportó en su catálogo de moneda que la moneda era nacional en el ejercicio anterior en 81 pólizas, mientras que para el ejercicio actual reportó que la moneda en 10 de sus pólizas fue en dólares. Por tal motivo, no coinciden

10 pólizas con lo que reportó en el ejercicio actual respecto a lo que reportó en el ejercicio anterior.

Aplicando la fórmula, quedaría lo siguiente:

$$Porcentaje\ de\ error = \frac{10}{81} * 100$$

$$Porcentaje\ de\ error = 0.1234 * 100$$

$$Porcentaje\ de\ error = 12.34\%$$

Aplicando importancia grado 1 debido a que la validación depende únicamente de variables categóricas se obtendría un puntaje de 9 puntos.

Grado 1	
Rango del porcentaje de error	**Puntaje**
>= 0%, <= 1%	0
> 1%, <= 5%	1
> 5%, <= 10%	2
> 10%	9

Después de asignar un puntaje a cada validación, se procede a realizar la suma de los puntajes institución por institución, obteniendo como resultado un puntaje final.

$$score = \sum_{i=1}^{n} Validación_i$$

$$score = Validación_1 + Validación_2 + Validación_3 + \cdots + Validación_n$$

La tabla muestra a forma de ejemplo los puntajes obtenidos en las validaciones.

	Validaciones que no son tolerancia cero		Validaciones contables		Validaciones de catálogo con opción de "OTROS"	Validaciones de consistencia entre catálogos	
CLAVE	Validación 1	Validación 2	Validación 3	Validación 4	Validación 5	Validación 6	Validación 7
CIA 1	90	0	800	0	90	0	9

$$score = Validación\ 1 + Validación\ 2 + Validación\ 3 + Validación\ 4 + Validación\ 5 + Validación\ 6 + Validación\ 7$$

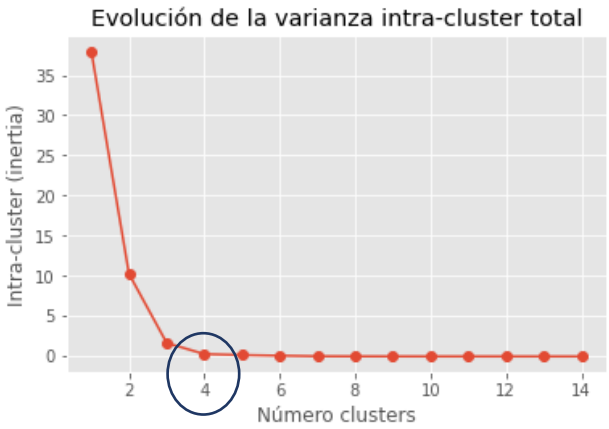
Se sustituye los pesos de cada validación y queda como sigue:

$$score = 90 + 0 + 800 + 0 + 90 + 0 + 9$$

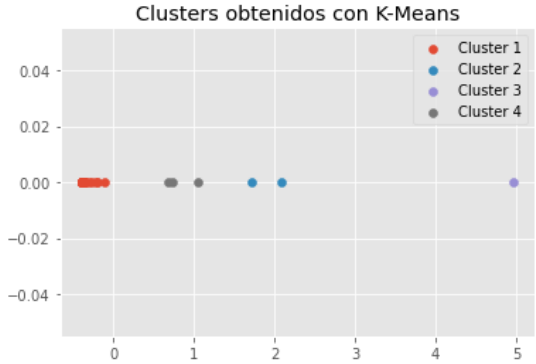
$$score = 989$$

El score se obtiene por institución y se introduce en el modelo de segmentación por clustering para determinar el número óptimo de agrupaciones.

Se aplica el método del codo para determinar el número óptimo de grupos o clústeres para la información estadística. Siguiendo los resultados obtenidos con el método del codo, se determina que debemos tener 4 grupos que segmenten el conjunto de información. Estos 4 grupos separan de manera significativa las diferencias entre ellos y agrupan según sus puntajes.



Además, se considera una mejor opción tener 4 calificaciones para el Semáforo de Calidad de la Información Estadística en lugar de 3, ya que esto proporciona una mayor visibilidad a la mejora de la calificación en el futuro, al ofrecer 4 calificaciones distintas que una institución puede obtener.



Se aplica la segmentación K-means con 4 clúster, con esto se confirma la existencia de 4 calificaciones para las instituciones de acuerdo con su score obtenido.

Calificación del Semáforo de Calidad de la Información Estadística	
SCORE	**Calificación**
>= 0, <= 10	Confiable
> 10, <= 171	Aceptable
> 171, <= 890	Poco Confiable
> 890	Deficiente

De acuerdo con el apartado de asignación de calificación del capítulo 3, tabla “Calificación del Semáforo de Calidad de la Información Estadística” y a que la institución obtuvo un score de:

$$score = 989$$

El score es mayor a:

$$score > 890$$

Por lo tanto,

$$\boldsymbol{score = Deficiente}$$

La institución obtuvo una calificación “Deficiente”, es decir, la información entregada por parte de la compañía presenta grandes inconsistencias en los reportes entregados a esta Comisión. Se recomienda no utilizar esta información para fines estadísticos.

Antes de enviar una calificación a la institución aseguradora o afianzadora, se revisan las cartas que envían a la compañía con la idea de entender el motivo del envío de su información estadística o si tienen alguna aclaración respecto a lo enviado. Muchas veces, presentan argumentos válidos o casos especiales atípicos en comparación con las validaciones generales que se aplican para todo el sector.

Después, con las calificaciones del Semáforo de Calidad de la Información Estadística, se emiten oficios de requerimiento y emplazamiento. Para posteriormente, el Semáforo de Calidad de la Información Estadística se publica en diversos canales de difusión. Es importante destacar que solo se proporciona el nombre de cada compañía junto con su calificación: Confiable, Aceptable, Poco confiable, Deficiente o Sin entrega.