UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO

FACULTA DE CIENCIAS

Análisis de Componentes Principales: un modelo actuarial aplicado al impacto de la caducidad en la Reserva de Seguros de Vida Individual

Tesis presentada por:

Rafael Pérez Aguilera

ACTUARÍA

Tutor: M. En F. Alfonso Parrao Guzmán

CIUDAD UNIVERSITARIA, CD. MX., 2022

INDICE

JUSTIFICACIÓN DE LA TESIS

Los seguros de Vida son un método de protección (económica para los beneficiarios) ante un hecho natural como la muerte provocada por factores aleatorios con el proceso de la misma muerte, es decir, es una protección contra la incertidumbre del momento de ocurrencia de la muerte por las pérdidas económicas derivadas de ésta, que puedan causar para los familiares o a quienes haya designado como beneficiarios, así como los planes a futuro de sus dependientes mermados o detenidos por este hecho. Dicho lo anterior y dada la incertidumbre de los posibles hechos aleatorios que pueden dar como resultado la muerte, es sumamente importante que las empresas aseguradoras tengan una metodología del cálculo robusta de la reserva (pasivos de la empresa) así como indicadores de dicha reserva para saber qué variables impactan en mayor medida a ésta.

PLANTEAMIENTO DEL PROBLEMA

Actualmente en México los seguros en sus diferentes ramos, están tomando mayor auge por diferentes cuestiones como sigue: Los seguros de gastos médicos (mayores) debido a las condiciones actuales de diferentes Instituciones de Salud, los seguros de automóviles debido a las leyes mexicanas así como una mejor protección ante los riesgos de daños tanto al auto como a las personas y sus bienes, así como la inseguridad de nuestro país, los seguros de Daños como consecuencia de sismos, terremotos, incendios y diferentes hechos que puedan mermar el patrimonio de cada persona y por último los seguros de vida como una protección contra las pérdidas económicas que pueda causar la muerte.

Para una empresa aseguradora es de vital importancia el cálculo de la reserva de riesgos en curso, la cual tiene la finalidad de cubrir la esperanza de pérdida de los riesgos en vigor y que dependen del tipo de ramo o seguro de que se trate, en este caso la muerte. Dicho lo anterior uno de los problemas dentro de las empresas de seguros de Vida es determinar qué variables impactan directamente y en mayor medida en las reservas y cómo se puede explicar la variación de la reserva en diferentes momentos de la *valuación*[1] así como cuántas pólizas puede tener la empresa, que impacten directamente en las reservas.

Una solución técnica a estos dos problemas es un modelo de análisis de componentes principales, el cual permite analizar y determinar las variables que más afecten o tengan una relación alta en la variación de la reserva de un momento a otro y un modelo de caducidad que permita predecir las pólizas que una empresa tendrá en vigor para años posteriores.

OBJETIVO

El objetivo de este trabajo es dar a conocer un indicador actualizable año con año, para llevar a cabo controles de la variabilidad de la reserva que la explique, así como un modelo de caducidad de pólizas y que genere las proyecciones necesarias para la reserva.

[1] Las carteras de Vida se valúan de acuerdo con la normativa mexicana, al menos mensualmente. Valuación es la forma de cuantificar el monto que la Aseguradora deberá reservar para hacer frente a la media de las reclamaciones, de aquellas pólizas que en ese momento se encuentren en vigor.

HIPÓTESIS

Como hipótesis se considera que las principales variables que afectan los cambios de la reserva son: la prima, edad del asegurado, tipo de plan y plazo del seguro. Probablemente existen otras no menos importantes; como: el salario promedio, cultura financiera, etc., sin embargo, estas variables no se encuentran disponibles en la información en general de las pólizas.

CAPÍTULO I. Solvencia II y Reservas en México para los seguros de Vida Individual

Conceptos Básicos

Los seguros de vida individual son una forma de protección para los beneficiarios del asegurado, ante un hecho natural como la muerte provocada por hechos aleatorios y fortuitos que tengan o no que ver con la naturalidad de la misma muerte, es decir, es una protección ante los posibles riesgos, amenazas o peligros. Derivado de la introducción de éstos, en donde la transferencia del riesgo pasa a las compañías aseguradoras, es necesario que éstas sean **solventes** para con el pago de sus **obligaciones** hacia los beneficiarios del asegurado.

Solvencia

Financieramente, cuando se habla de solvencia se refiere a la capacidad de una organización o entidad para cumplir con todas sus obligaciones de pago sin importar de cuando tenga que afrontarlas. Ahora bien, se habla de una solvencia técnica, cuando la empresa por sí misma es capaz de generar autofinanciación suficiente para hacer frente a sus deudas, ya que, de lo contrario, tendrá que recurrir a otras opciones como financiación adicional, es decir, créditos o venta de activos. Por último, se habla de solvencia efectiva cuando se cumplen con las obligaciones en tiempo pactado. De este modo se dice que una empresa es solvente, cuando lleve a cabo los factores que fundamentan la solvencia técnica y efectiva.

Para el sector asegurador se debe tener en mente que se habla de una entidad que recibe ingresos en forma de prima para después realizar el pago de sus obligaciones, es decir, los siniestros por muerte en el caso de los seguros de vida, siniestros de atención médica en el caso de los seguros de gastos médicos mayores y el pago de la suma asegurada en pérdidas de bienes para los seguros de daños. Con lo anterior se puede decir que el recibir los ingresos antes de realizar los pagos, los cuales son inciertos, pone en primer lugar el objetivo de ser una entidad solvente.

Por consiguiente, para asegurar una estabilidad de todos los asegurados, es necesario que al precio del servicio o transferencia del riesgo se le agregue un componente, a este componente se le conoce como recargo técnico.

En términos más simples, la solvencia en la compañía de seguros puede ser catalogado como otro pasivo que esté compuesto por recursos propios y reservas, que han sido constituidos mediante recursos ajenos y por esta precisa razón es de gran importancia para la entidad contar con un régimen adecuado, mismo que le permita asegurar su permanencia en el mercado en el largo plazo ~~como~~ para la protección de los asegurados. Es así como, para poder determinar y evaluar su posición financiera, se deben tomar en cuenta de manera coherente los riesgos a los que como aseguradores están expuestos, así como las consecuencias que traerían la manifestación de éstos. Una situación de insolvencia para este tipo de entidad ocasionaría elevados costos económicos y sociales, ya que son éstas, las que, mediante las funciones de protección y compensación, administran los recursos de terceros. Dadas las condiciones antes mencionadas, a continuación, se presentan algunos aspectos que las aseguradoras deberán considerar para contar con un nivel de solvencia adecuado:

- La aseguradora debe hacer frente a las obligaciones adquiridas bajo las circunstancias previstas, tanto en el corto como en el largo plazo.
- Se deben evaluar los factores de riesgo a los que están expuestas y su posible impacto.
- Se necesita prudencia explícita en requerimientos regulatorios.
- Los valores entre activos y pasivos que se obtienen al llevar a cabo la evaluación de solvencia deben ser consistentes.
- Se debe considerar un régimen de solvencia específico en la determinación de reservas.
- Debe existir claridad en el costo esperado de las responsabilidades asumidas.

Bajo el contexto anterior, se puede observar que es de suma importancia que la entidad aseguradora mantenga un nivel de control en sus operaciones, de tal forma que le permita continuar con sus funciones de manera confiable y segura. Es por ello, que se debe tomar en cuenta el hecho de que las actividades del seguro son de naturaleza aleatoria, lo que ocasiona desviaciones del mismo carácter, trayendo como resultado problemas en la estabilidad del ente asegurador. En consecuencia, las compañías comenzaron a elaborar esquemas que les permitieran minimizar este riesgo, tomando como referencia las propuestas de Basilea para contar con una mejor solidez financiera.

Solvencia I y II

Solvencia I tuvo sus comienzos en los años 70 y su aplicación se prolongó hasta que se introdujo el modelo de Solvencia II, en sus inicios supuso un gran avance en el sector asegurador. Solvencia I trata un modelo sólido, sencillo de aplicar y de fácil seguimiento y control, de ahí que los principios sobre los que está basado este marco regulatorio son:

- Se definen reglas de cálculo para provisiones técnicas mediante la aplicación de fórmulas en un ambiente de prudencia.
- El cálculo del margen de solvencia se realiza a través de porcentajes de siniestros, de provisiones y primas.

Este modelo establece por un lado una relación directa entre el capital de las instituciones de seguros y su volumen de negocios, lo cual quiere decir que a mayor dimensión tenga la empresa, ésta debía contar con mayores recursos de capital respecto a aquellas pequeñas empresas.

A diferencia de Solvencia I, el esquema de Solvencia II plantea la necesidad de que las entidades cuenten con recursos suficientes para hacer frente a sus obligaciones, para lo cual contempla un modelo basado en el perfil de riesgos de cada institución, de manera que una entidad con una mayor exposición al riesgo debería contar con más recursos financieros que una entidad con un perfil de riesgo menor. Dicho lo anterior y con el objetivo de lograr su meta Solvencia II se basa en tres procesos de disciplina:

- Regulatoria: basada en el establecimiento de conjunto de normas que reduzcan la probabilidad de insolvencia.
- La auto-disciplina: consiste en las normas de auto impuestas por las entidades de seguros a través de un gobierno corporativo las cuales proporcionan a las compañías a mantener la solvencia.
- Disciplina del mercado.

Estos procesos se encuentran agrupados en los tres pilares del marco de solvencia II, mismos que se presentan en la siguiente figura:

Imagen I. Modelo Mexicano de Supervisión Basado en Riesgos tipo Solvencia II, Herrera Conteras Fernando, CNSF.

Dado que este trabajo está basado en las Reservas Técnicas, se enfocará en la explicación del Pilar I. Dentro del marco regulatorio mexicano, en este pilar se establecen los requerimientos para garantizar la solvencia de las instituciones en 5 conceptos:

- Reservas Técnicas
- Requerimiento de Capital de Solvencia
- Fondos Propios admisibles
- Inversiones
- Reaseguro

Aunado a lo anterior, contempla la generación de un “Balance Económico” en donde se busca que los activos y pasivos deben ser valuados considerando su valor de mercado.

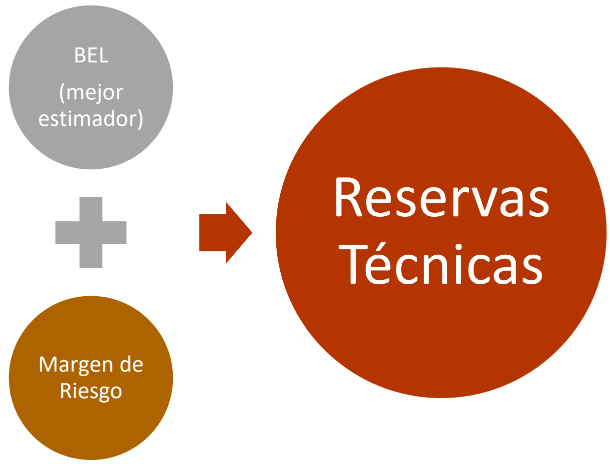
Reservas Técnicas

Las reservas técnicas son las provisiones necesarias que las entidades de seguros y fianzas deben de constituir para que tengan recursos de tal forma que puedan hacer frente a sus obligaciones. Por esta razón se deben constituir las reservas técnicas obligatorias establecidas en la Ley de Instituciones de Seguros y Fianzas (LISF) en sus artículos 216 y 220 dependiendo del ramo que operan, mismas que se mencionan a continuación:

- Reserva de riesgos en curso
- Reserva de obligaciones pendientes por cumplir
- Reserva matemática especial
- Reserva para fluctuación de inversiones
- Reserva de contingencia
- Reserva de riesgos catastróficos
- Reserva de fianzas en vigor
- Reserva de contingencia de fianzas

Aunado a lo anterior las instituciones deben considerar lo siguiente:

- Constituir y valuar las reservas técnicas de forma prudente, confiable y objetiva, utilizando métodos actuariales.
- El monto de las reservas técnicas deberá constituirse como la suma de **la mejor estimación** (BEL por sus siglas y traducción en inglés) y de un **margen de riesgo**.
- La mejor estimación debe realizarse, de acuerdo con la legislación, con información oportuna, confiable, homogénea y suficiente, utilizando hipótesis apegadas a la realidad, así como métodos actuariales y técnicas estadísticas en apego a los estándares actuariales. Así mismo la mejor estimación está compuesta por el valor esperado de los flujos futuros entendiéndose como el valor esperado de la totalidad de los ingresos y egresos en términos brutos, necesarios para hacer frente a las obligaciones de los contratos de seguro y reaseguro durante su periodo de vigencia.
- El margen de riesgo es el monto que al sumarse con el mejor estimador garantice que el monto de las reservas técnicas sea equivalente al que las instituciones de seguros requerirán para asumir y hacer frente a sus obligaciones. El margen de riesgo también es visto como el costo de capital que sobre una tasa libre de riesgo esperaría ganar la empresa a la que se le transfiere el riesgo.



Reserva de Riesgos en Curso y BEL

Como ya se explicó con anterioridad la Reserva de Riesgos en Curso es una provisión realizada por las compañías de seguros para hacer frente a los siniestros ocurridos y reclamados durante la vigencia de la póliza. En México y de acuerdo con la legislación se valúa a partir de un modelo interno que realiza cada compañía de seguros en donde este modelo es desarrollado, explicado y autorizado mediante una Nota Técnica por la Comisión Nacional de Seguros y Fianzas, todo esto significa que la compañía puede aplicar sus métodos propios, sin embargo, existen algunas condiciones que deben cumplirse de manera estricta:

- Se debe contar con información estadística suficiente, confiable y oportuna.
- Todos los métodos y resultados se deben validar mediante pruebas de backtesting.

Lo anterior genera una sinergia con el Pilar II debido a que, como son procesos y metodologías internas en cada entidad, éstas deben de ser revisadas constantemente para su correcto funcionamiento y es aquí donde entra la **función actuarial** la cual se encarga de valorar y certificar la correcta utilización de las hipótesis así como del funcionamiento del cálculo de la reserva de riesgos en curso, esta función actuarial debe ser capaz de medir riesgos potenciales con conocimientos sólidos y juicio independiente en dichas reservas.

Ahora bien, dentro de la normativa se menciona que la Reserva de Riesgos en Curso debe corresponder a la "mejor estimación" o, en palabras más actuariales, al valor esperado de las obligaciones derivadas de los contratos de seguros en vigor, más un "margen de riesgo". En este sentido y utilizando el esquema de Solvencia II, se establece que la mejor estimación debe corresponder al valor esperado de los flujos de las obligaciones futuras sin considerar un margen de riesgo dentro de la misma estimación.

Para el negocio de vida individual, que es donde está enfocado este trabajo es de suma importancia tener los movimientos que puede tener una póliza en su operación y de forma general es posible expresar la reserva de riesgos en curso denotado por RRC de la siguiente manera:

$$RRC = \sum_j \frac{(Sin_j + Resc_j + Div_j + Gadmin_j + CAdq_j - Primas_j)}{\prod_{i=1}^{j}(1 + i_j)} + MR$$

En donde:

Sin_j: Siniestralidad esperada al tiempo j

$Resc_j$: Rescates esperados al tiempo j

Div_j: Dividendos esperados al tiempo j

$Gadmin_j$: Gastos de administración esperados al tiempo j

$Cadq_j$: Costos de adquisición esperados al tiempo j

$Primas_j$: Primas de tarifa netas esperados al tiempo j

i_j: tasa de interés al tiempo j

MR: Margen de Riesgo

Al observar la ecuación anterior resulta lógico pensar que, al ser estimaciones en cada uno de los conceptos, estos deben de tener una metodología que no necesariamente debe de ser la misma para cada concepto. Por lo anterior debe de existir una base matemática y estadística sólida dentro de las compañías de seguros y como se ha mencionado anteriormente esta debe tener información oportuna, confiable, homogénea y suficiente en donde las hipótesis para cada concepto deben de ser suficientemente realistas. De esta forma y a manera de ejemplo si se trata de un seguro de vida, donde los rescates es la probabilidad de que se cancele la póliza por el comportamiento del asegurado, el monto estimado de las obligaciones de la entidad aseguradora por el número de cancelaciones es:

$$Resc_i = SA_i * cad_i$$

$cad_i =$ es el valor estimado de salidas por cancelación (caducidad).

Es de suma importancia mencionar en este punto que la caducidad es únicamente una salida por cancelación por lo que al plantear un modelo de esta caducidad se debe excluir la mortalidad y que si bien resulta simple leer la ecuación para la estimación la aplicación es bastante complicada, derivado de que existen pólizas con diferentes sumas aseguradas y claramente con diferentes años póliza en donde la caducidad debería ser decreciente.

Es aquí donde es posible observar la complejidad de Rescates y en general de la RRC en donde se le exige al actuario conocimientos sólidos en todo ámbito. En este trabajo se abundará en el siguiente capítulo una forma de realizar el cálculo de las tasas estimadas de caducidad para un correcto funcionamiento dentro de la RRC.

CAPÍTULO II. Caducidad

¿Qué es la caducidad?

La caducidad es una métrica para el caso de seguros, que indica la probabilidad de que una póliza sea cancelada por cualquier causa, a excepción del término de la póliza y el siniestro. La caducidad es una medida de suma importancia para las empresas aseguradoras, ya que las cancelaciones de las pólizas afectan directamente a las reservas matemáticas de la institución las cuales corresponden al pasivo de las mismas, es decir, afectan de manera directa a las obligaciones de la empresa, así como al ingreso de primas que tiene la compañía a lo largo del plazo del pago de primas de las pólizas vigentes que permiten a la aseguradora una eficiente recuperación de sus gastos de administración y adquisición, mantener condiciones favorables de solvencia y liquidez, así como generar utilidades.

La caducidad, dentro de una compañía de seguros, también sirve como un indicador que puede utilizar el área de *pricing* y *suscripción* ya que, a partir de éste, el área de *pricing* puede identificar las áreas de oportunidad de ciertos negocios ya que una cancelación mayor a otros negocios puede ser un indicativo de que un negocio específico no está captando venta requerida o existe algo en el producto que no es atractivo para con el asegurado lo cual genera cancelaciones y por consiguiente rendimientos negativos para la aseguradora.

Otra función de la caducidad es que puede ser un indicador de malas prácticas y "focos rojos". Para hacer más claro este punto, suponga que se tiene un producto nuevo en una compañía, en el cual, al agente de seguros se le da una comisión del 50% sobre la prima que va a pagar el asegurado al comprar la póliza, además suponga que los agentes encargados de vender estas pólizas venden mil pólizas pero de estas mil pólizas pasado un año se cancelan 900 de ellas, al segundo año de las 100 restantes se cancelan 20 de ellas y al tercer año de las 80 se cancelan 20.

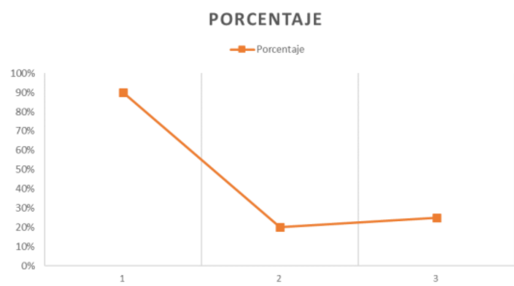
Utilizando la definición de caducidad se calcula para este caso, como:

$$Caducidad = \frac{número\ de\ pólizas\ canceladas\ al\ final\ del\ año\ de\ estudio}{número\ de\ pólizas\ totales\ al\ inicio\ del\ año\ de\ estudio}$$

Siguiendo la ecuación anterior, se obtiene la siguiente información en forma de tabla.

Años Póliza	Pólizas Vigentes	Caducidad	Porcentaje
1	100	900/1000	90%
2	80	20/100	20%
3	60	20/80	25%

Gráficamente se observa de la siguiente manera



Se podría preguntar ¿Por qué se tiene una caducidad tan grande para el primer año? Contestar esta pregunta es difícil sin realizar un análisis profundo y sin contar con la información necesaria, ya que las causas pueden ser variadas, sin embargo, dada la información con la que se cuenta (resultado de la caducidad), se tiene algo de intuición sobre el problema de ofrecer al agente de seguros una comisión tan grande sobre la prima pagada por el asegurado. Lo único que está sucediendo, es que el agente está enfocado en venta nueva más no en la conservación de la cartera (renovación de pólizas) contemplando el porcentaje de comisión que tiene en el esquema de venta nueva y de renovación.

Dado que este trabajo está enfocado a los seguros de Vida Individual, es necesario mencionar algunos de los aspectos más relevantes, principalmente en cuanto a la empresa y las modalidades de aseguramiento que se presenta por tipo de riesgo y plazo, de igual manera es necesario hacer la clasificación de los seguros en el mercado a los que pertenecen:

- Seguros Tradicionales: son aquellos seguros que se caracterizan por la constitución de reservas matemáticas con base en primas constantes y riesgos variables en el tiempo, dentro de estos se encuentran los seguros ordinarios de vida, seguros dotales, seguros dotales mixtos, seguros temporales e incluso los diferidos.
- Seguros Universales: Este seguro consiste en la protección contra la mortalidad, más un componente financiero al ofrecer fondos de inversión sobre primas excedentes.
- Seguro Flexible: Es aquel que combina una cobertura de seguro por fallecimiento que tiene ligado un fondo, el cual es administrado por la propia compañía de seguros y únicamente le ofrece una tasa garantizada. La vigencia de este tipo de seguros es variable y se encuentra altamente ligado a la suficiencia de la reserva matemática y su fondo.

Además de la clasificación anterior también existe la clasificación en general de Cartera Básica y Cartera de Beneficios adicionales.

La primera de ellas se compone de las coberturas que se asocian a los riesgos de fallecimiento o supervivencia de las pólizas. La segunda comprende aquellas coberturas contra los riesgos de accidente, invalidez, morbilidad de alto riesgo del asegurado e incluso de pérdidas orgánicas del mismo asegurado, entre otros riesgos adicionales.

Método de Cálculo

En el apartado anterior dentro del ejemplo se tocaron dos puntos fundamentales para el método de cálculo de la caducidad, uno de ellos tiene que ver con el hecho de que se ofreció a la venta un nuevo producto, esto brinda una pauta para el cálculo de caducidades, es decir, se debe tener información de productos en condiciones/planes homogéneos (condiciones que ya se han mencionado en el apartado anterior dentro de las modalidades), ya que por sus características existen productos que son más susceptibles de cancelación que otros, por ejemplo, se tienen diferentes caducidades de un seguro vitalicio con plazo de pago 20 a un seguro vitalicio con pagos vitalicios, el por qué puede no ser tan claro pero esto se debe a que en el vitalicio con pago de 20 años mientras más se acerque al último año menor es la probabilidad de que se cancele la póliza ya que el año 20 sería el último que se tendría que pagar y para el asegurado la cobertura es para toda la vida, por otro lado para el vitalicio con pagos durante toda la vida, es probable que en algún momento del vigor de la póliza, el asegurado ya no pueda (o no desee) pagar la prima a los 65 años o a los 70 dependiendo de las condiciones en que se encuentre dicho asegurado.

La segunda característica tiene que ver con el año de vigencia de la póliza, éste es un factor fundamental que influye en el cálculo. Por ejemplo, regresando al ejemplo del punto anterior supóngase que se tiene un seguro vitalicio con plazo de pago 20 en donde el pago 20 fue el día de hoy, en este caso no tendría sentido para el asegurado cancelar la póliza porque como tal, ya está pagado. De acuerdo con diversos estudios actuariales se conoce que existe una relación importante entre la antigüedad de las pólizas y la probabilidad de cancelación. Típicamente, las probabilidades de cancelación de los planes se pueden construir haciendo un análisis en un mismo año, el nivel de cancelaciones observadas respecto de las pólizas expuestas con diferentes años de antigüedad[2].

Aunado a lo anterior dentro del cálculo de caducidad se encuentran involucrados movimientos dentro de la cartera que pueden afectar al conjunto de pólizas dentro del estudio. A continuación, se describirán aquellos movimientos que pueden afectar el alza o baja en una cartera:

- Aumentos de pólizas:
    - Emisión: La emisión sucede una vez que el área de suscripción y reaseguro de una compañía con base en la solicitud del seguro, criterios de asegurabilidad y sus debidas pruebas, aceptan el riesgo mediante un contrato de seguro, en donde se estable el momento exacto en donde la compañía aseguradora adquiere el riesgo y por cuanto tiempo, la emisión puede ser anticipada, es decir, antes de asumir el riesgo.
    - Rehabilitaciones: Se dan cuando un contrato ha sido dado de baja de la cartera de pólizas en vigor, debido a la falta de pago de primas por parte del asegurado y después de un periodo de tiempo el asegurado decide reactivar sus derechos

[2] Tasas de Caducidad: Guía de apoyo para la Construcción y Aplicación, Act. Pedro Aguilar Beltrán. Asociación Mexicana de Instituciones de Seguros

y obligaciones, mediante el pago de primas que hubieran vencido desde la fecha que operó la cancelación hasta la fecha de rehabilitación del seguro.

- Disminución de pólizas
    - Vencimientos: Esto se presenta cuando el asegurado llega con vida al final del plazo contratado (seguro temporal y dotal). En este caso sólo es aplicable para planes dotales y en línea con el plan la aseguradora procede a pagar la suma asegurada establecida y al mismo tiempo a dar de baja la póliza con estatus de terminada o expirada.
    - Rescates: Es el valor garantizado que tiene el asegurado por Ley (LSCS[3]) cuando la temporalidad del seguro es mayor o igual a 10 años (en el caso de rescates para planes de menor temporalidad es facultativo para la reaseguradora) donde la compañía debe otorgarlo a petición del asegurado después de un cierto tiempo en que el seguro se ha mantenido vigente (de acuerdo con las condiciones generales y/o particulares de la póliza), este rescate puede ser parcial o total. Cuando es total se traduce en una cancelación del contrato de seguro previa solicitud del asegurado, lo que generaría un pago en efectivo (o conversión a un seguro saldado o prorrogado[4] con ese valor de rescate como prima única) derivado de un porcentaje de la reserva matemática constituida.
    - Expiración: Este concepto se aplica a los planes temporales que cubren el riesgo de fallecimiento, cuando el asegurado llegue con vida al final del plazo pactado en el contrato con la diferencia de que en esta ocasión no se le paga la suma asegurada, pero se cancela la póliza y se libera la reserva.
    - Siniestro: Este concepto se aplica cuando el asegurado ha fallecido por alguna causa procedente del contrato por lo que la aseguradora tiene que pagar la suma asegurada a los beneficiarios y marca la póliza con estatus siniestrada generando una disminución en las pólizas vigentes.
    - Cancelación: este concepto como su nombre lo indica, ocurre cuando una póliza se cancela por alguna razón diferente a las anteriores. Cuando la póliza se cancela antes de que termine el plazo el asegurado tiene derecho al valor de rescate[5].

En este trabajo se dará una generalización del procedimiento para hacer el cálculo de caducidad utilizando las consideraciones anteriormente puntualizadas:

- Primero se procederá a dividir las pólizas por producto y planes homogéneos (con las características antes mencionadas) además de tomar en consideración la moneda

[3] Ley Sobre el Contrato de Seguro
[4] Saldado es un seguro en donde se mantiene el plazo original de la póliza, pero con la suma asegurada disminuida y prorrogado mantiene la suma asegurada original, pero disminuye el plazo de seguro.
[5] El valor de rescate es un concepto que se refiere al monto que le otorgará la aseguradora al asegurado o beneficiario, en caso de que la póliza sea cancelada. Comúnmente es un concepto aplicable en seguros de vida de largo plazo, sin embargo, el concepto puede ser aplicado a otros tipos de seguros de largo plazo, especialmente a los seguros de accidentes y enfermedades. Valores de Rescate, Act. Pedro Aguilar Beltrán

en que fueron emitidas, es decir, se procederá a clasificar la cartera en planes tradicionales, flexibles y universales, así como la moneda pactada.

- Hecho lo anterior, se calculará el porcentaje de conservación por producto-plan-moneda y de esta manera por complemento, se obtendrá la caducidad
    - El porcentaje de conservación se obtiene como:

$$\%Conservación = \frac{Base\ Conservada}{Base\ a\ Conservar}$$

Y por complemento la caducidad será:

$$\%Caducidad = 1 - \%Conservación$$

Un cuestionamiento de las fórmulas anteriores puede ser ¿Cómo obtener la Base a Conservar y la Base Conservada? Para obtener la base a conservar se debe tomar la información al inicio de cada periodo de medición sin hacer distinción por año observación y tomando en cuenta sus respectivos movimientos en el periodo de gracia correspondiente.

Es decir, se establece una única base de experiencia estadística como si cada año formara parte de una misma cartera expuesta a las mismas condiciones. Por ejemplo, la base a conservar de pólizas de antigüedad uno está formada por las pólizas de antigüedad uno al inicio del periodo de medición, suponga Diciembre de 2007 (base a conservar en el periodo 2008) más las pólizas con antigüedad uno al inicio del periodo de medición de Diciembre de 2008 (base a conservar en el periodo 2009) y así sucesivamente hasta el año actual, sin embargo el actuario debe de saber que en algunas ocasiones no se tiene la información suficiente para un estudio de 10 años, sin embargo, esta misma información es necesaria para que el estudio sea suficientemente robusto y preciso.

De la manera expuesta anteriormente para la base a conservar, se obtiene la base conservada con la única variante de que la base conservada será al final del periodo de medición. Por ejemplo, tomando las mismas pólizas con antigüedad uno, se observaron aquellas que llegaron vigentes al final de un año de medición, para lo cual se utilizó la información de Diciembre 2007 (base conservada del periodo 2007) a Diciembre de la fecha del presente trabajo.

Cabe mencionar que la metodología descrita es para frecuencia, es decir, número de pólizas, sin embargo, la misma metodología se puede realizar utilizando la prima e incluso la suma asegurada cuidando aquellas veces que se incremente la prima o suma asegurada.

Una vez hecho lo anterior, se tendrán las tasas de conservación por año inicial hasta el año final de estudio, el siguiente paso es obtener la media de estas tasas de conservación por año póliza de la forma usual para así obtener la conservación media de la cartera por año póliza:

$$\overline{x_{ij}} = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=l}^{l-n} x_{ij}$$

Donde:

i: años póliza, con *i=1,2, 3, …, m*

j: años observados, con *j= l-n*

l: año observado final

n: año observado inicial

Utilizando la formula $\%Caducidad = 1 - \%\ de\ Conservación$ se obtendrá la caducidad real de la cartera por año póliza.

Hasta este punto se ha calculado la caducidad real promedio, sin embargo, como en todo estudio de tasas, se busca tener una curva que se ajuste lo mejor posible a estas tasas para hacer predicciones sobre las mismas. Para hacer esto, se debe tener en cuenta que las tasas de caducidad decrecen conforme pasa el tiempo, es decir, los años pólizas siempre deben de ser decrecientes y en el mejor de los casos (sobre todo en la práctica) convexa por lo que al ajustar una curva, estas deben de tener el mismo comportamiento, es por esta razón que una forma sencilla para modelar la caducidad a partir de la conservación es por medio de una función de supervivencia ya que esta es decreciente (mayoritariamente convexa) y para ajustar la función se puede utilizar la regresión lineal y así predecir años póliza futuros. Sin embargo, al momento de elegir un modelo como la función de supervivencia es necesario demostrar que justamente es de supervivencia, lo cual se demuestra de la siguiente manera:

1. Una función de supervivencia es una función S(x) definida en el intervalo [0,w] no creciente tal que S(0) = 1 y S(x) = 0 cuando x → ∞

De esta forma, ya que se tenga la función de Supervivencia, a partir de ella se harán las transformaciones necesarias para que se aplique la regresión lineal cuidando las propiedades de la regresión y sobre todo el coeficiente de determinación (R) ajustado que sea estadísticamente cercano al valor 1.

Ejemplo:

Suponga que se cuenta únicamente con pólizas temporales con plazos de seguro menor a 10 años, además se cuenta con una historia de estas pólizas de 2012 a 2016, como se sugiere en el siguiente cuadro.

	2012	2013	2013	2014	2014	2015	2015	2016
Año Póliza								
1	1023	576	1033	552	1272	642	1021	590
2	467	289	572	313	552	306	642	377
3	231	185	285	213	313	249	306	211
4	175	112	184	135	212	164	248	200
5	57	48	84	74	72	56	65	49
6	42	33	48	40	74	67	56	46
7	34	31	33	29	40	32	67	59
8	5	4	31	26	29	21	32	27
9	0	0	4	3	26	21	21	18
10	0	0	0	0	0	0	0	0

Siguiendo los pasos anteriores se puede decir que la segunda columna denominada “2012” es **la base a conservar** por año póliza y la tercera columna denominada “2013” es **la base conservada** por año póliza, es decir, de todas las pólizas de 2012, cuántas de ellas quedaron a finales de 2013. La cuarta columna es la nueva base **a conservar en 2013** la cual contiene las todas las pólizas conservadas de 2012 más la nueva emisión de pólizas del 2013 por lo tanto la quinta columna denominada “2014” es la **base conservada** del año 2013. De esta forma se definen las demás columnas, aunque cabe señalar que, al pasar de una base conservada a una nueva base a conservar del año siguiente, debe existir una lógica en cuanto a los años póliza se refiere. Véase el siguiente cuadro:

	2012	2013	2013	2014	2014	2015	2015	2016
Año Póliza								
1	1023	576	1033	552	1272	642	1021	590
2	467	289	572	313	552	306	642	377
3	231	185	285	213	313	249	306	211
4	175	112	184	135	212	164	248	200
5	57	48	84	74	72	56	65	49
6	42	33	48	40	74	67	56	46
7	34	31	33	29	40	32	67	59
8	5	4	31	26	29	21	32	27
9	0	0	4	3	26	21	21	18
10	0	0	0	0	0	0	0	0

En el cuadro anterior se puede ver que en 2012 el número de pólizas con plazo de seguro diez y que está corriendo su primer año póliza son 1023 de las cuales al final de 2013 únicamente se conservan 576. Dicho lo anterior, debe considerar que esas 576 pólizas ya no corresponden al primer año póliza porque teóricamente al final de 2013 ya cruzaron vigencia, es por esta razón que en la cuarta columna se encuentran 572. El motivo de que sean menos puede variar dependiendo de los supuestos que toma cualquier compañía para la elaboración de una nueva base a conservar, sin embargo, lo que nunca puede suceder es que sean más pólizas que las que se conservaron en la base conservada, a lo más pueden ser iguales. Es por este motivo que, al observar cada base conservada con una nueva base a conservar, se tiene que son menores o iguales. Véase el siguiente cuadro el cual contiene sombreadas en verde y amarillo la base conservada y la nueva base a conservar.

	2012	2013	2013	2014	2014	2015	2015	2016
Año Póliza								
1	1023	576	1033	552	1272	642	1021	590
2	467	289	572	313	552	306	642	377
3	231	185	285	213	313	249	306	211
4	175	112	184	135	212	164	248	200
5	57	48	84	74	72	56	65	49
6	42	33	48	40	74	67	56	46
7	34	31	33	29	40	32	67	59
8	5	4	31	26	29	21	32	27
9	0	0	4	3	26	21	21	18
10	0	0	0	0	0	0	0	0

En verde se encuentran las bases conservadas y las amarillas son las nuevas bases a conservar en donde las amarillas son menores o iguales a las verdes del año n-1.

Una vez validada esta consistencia se procederá a calcular el porcentaje de Conservación por año de historia y año póliza como se mencionó anteriormente (base conservada entre la base a conservar) y se promediará para tener una sola curva de tasas reales de conservación y de caducidad utilizando el complemento. De esta forma quedan los siguientes resultados.

Año Poliza	2012	2013	2014	2015	Conservacion Promedio	Caducidad
1	56%	53%	50%	58%	54%	46%
2	62%	55%	55%	59%	58%	42%
3	80%	75%	80%	69%	76%	24%
4	64%	73%	77%	81%	74%	26%
5	84%	88%	78%	75%	81%	19%
6	79%	83%	91%	82%	84%	16%
7	91%	88%	80%	88%	87%	13%
8	80%	84%	72%	84%	80%	20%
9	NA	75%	81%	86%	80%	20%
10	NA	NA	NA	NA	0%	0%

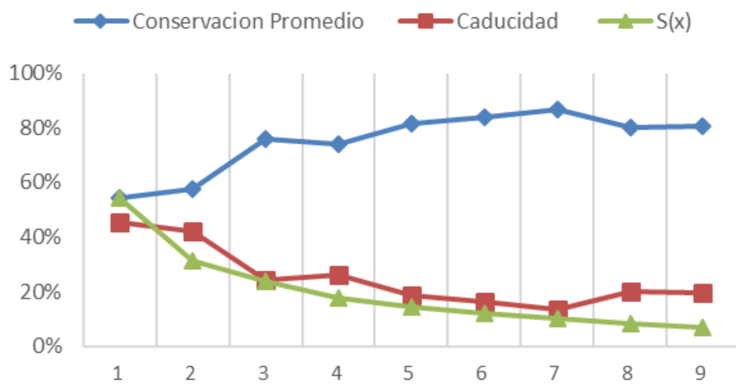
Siguiendo la definición de supervivencia y utilizando la esperanza o el promedio de la conservación, el cual es a partir de la información real, se construirá la función de supervivencia de la siguiente manera:

2. Sea x el año póliza con x =1, 2, ...,10 (esto es porque son pólizas temporales con plazo máximo de seguro de 10 años)

$$S(x) = \begin{cases} Conservación_x & cuando\ x = 1 \\ S(x-1) * Conservación_x & cuando\ x > 1\ y\ x \leq 10 \\ 0 & cuando\ x \geq 10 \end{cases}$$

Año Poliza	2012	2013	2014	2015	Conservacion Promedio	Caducidad	S(x)
1	56%	53%	50%	58%	54%	46%	54%
2	62%	55%	55%	59%	58%	42%	31%
3	80%	75%	80%	69%	76%	24%	24%
4	64%	73%	77%	81%	74%	26%	18%
5	84%	88%	78%	75%	81%	19%	14%
6	79%	83%	91%	82%	84%	16%	12%
7	91%	88%	80%	88%	87%	13%	10%
8	80%	84%	72%	84%	80%	20%	8%
9	NA	75%	81%	86%	80%	20%	7%
10	NA	NA	NA	NA	0%	0%	0%

Graficando los resultados:



Observando la gráfica se puede notar que la función de supervivencia a partir de la conservación es convexa decreciente. Centrándose en esta función ahora se requiere encontrar una función que se ajuste a estas tasas, la que se comporta como tal se observa en la gráfica que es la función exponencial, sin embargo, también se debe tomar en cuenta que se ajustará mediante una regresión lineal, por lo que una función que puede servir para ajuste es la siguiente:

$$S(x) = \exp\left(-\alpha x^{\beta}\right) \, con \; \alpha \neq 0 \; y \; \beta \neq 0$$

Para comprobar que es función de supervivencia se demuestra lo siguiente:

1. S(0) = 1
2. S(x) = 0 cuanto x → ∞

Demostración

P.D. 1)

$$Sea \; x = 0 \; \rightarrow S(0) = \exp\left(-\alpha 0^{\beta}\right) \rightarrow como \; \alpha \neq 0 \; y \; \beta \neq 0 \, , 0^{\beta} = 0 \; y - \alpha * 0 = 0$$

$$\rightarrow S(0) = \exp(0) = 1$$

$\square$

P.D. 2)

Cuando $x \geq 10 \; S(x) = 0$

$\square$

Utilizando esta curva de ajuste, se realizarán las transformaciones necesarias para que se pueda aplicar una regresión lineal. Las transformaciones son las siguientes:

- $x' = \ln{(x)}$
- $y = \ln{(-\ln{(S(x))})}$

El por qué se debe a la siguiente demostración:

$$S(x) = \exp\left(-\alpha x^{\beta}\right) \leftrightarrow \ln\left(S(x)\right) = -\alpha x^{\beta} \leftrightarrow -\ln\left(S(x)\right) = \alpha x^{\beta}$$

Aplicando logaritmos nuevamente, así como utilizando sus propiedades:

$$\ln\big(-\ln\big(S(x)\big)\big) = \ln(\alpha) + \beta\ln\,(x)$$

Renombrando la función se tiene:

$y = \alpha' + \ \beta x'$

Lo cual representa un modelo de regresión lineal simple en donde $\alpha' y\ \beta$ representan los parámetros a estimar de la siguiente manera:

$$\beta = \frac{\sum x' \sum y - n \sum x'y}{(\sum x')^2 - n \sum x'^2} = \frac{\sum(x' - \bar{x}')(y - \bar{y})}{\sum(x' - \overline{x'})^2}$$

$$\alpha' = \frac{\sum y - \beta \sum x'}{n} = \bar{y} - \beta\overline{x'}$$

Dicho lo anterior se tienen los siguientes resultados:

Año Poliza	S(x)	x=ln(x)	y=ln(-ln(S(x)))	(x-E(x))	(y-E(y))	(x-E(x))^2
1	54%	0.000	-0.4993	-1.51	-0.97	2.28
2	31%	0.693	0.1459	-0.82	-0.32	0.67
3	24%	1.099	0.3603	-0.41	-0.11	0.17
4	18%	1.386	0.5521	-0.12	0.08	0.02
5	14%	1.609	0.6643	0.10	0.19	0.01
6	12%	1.792	0.7522	0.28	0.28	0.08
7	10%	1.946	0.8169	0.44	0.35	0.19
8	8%	2.079	0.9101	0.57	0.44	0.32
9	7%	2.197	0.9938	0.69	0.52	0.47
10	0%	2.303	0.0000	0.79	-0.47	0.63

Por lo que los parámetros estimados resultan ser los siguientes:

$$\beta = 0.465579$$

$$\alpha' = \ -.02336$$

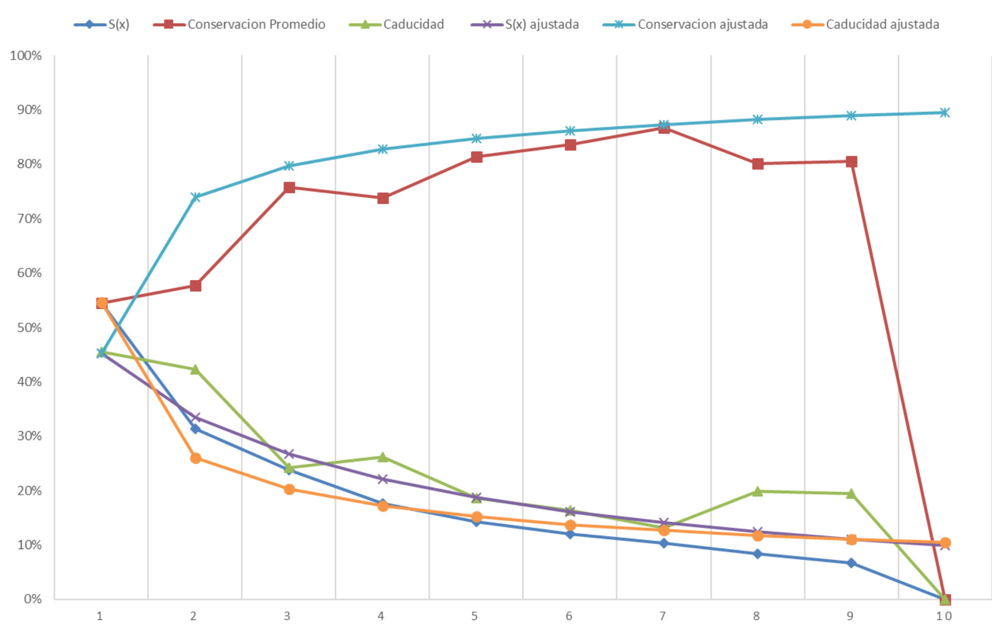
$$\therefore y = -.02336 + \ 0.465579x$$

R ajustada = .5267

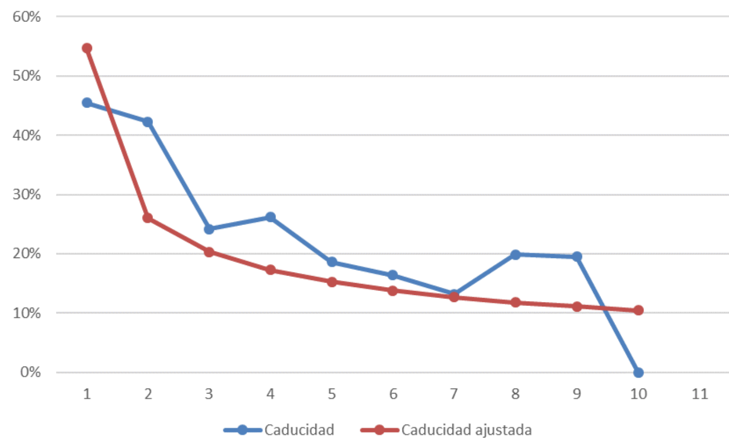
Ahora simplemente se regresa la transformación utilizando las funciones a la inversa a partir de las estimaciones, obteniendo así la función de Supervivencia ajustada, posteriormente se obtiene la conservación ajustada y haciendo 1- % de conservación se obtiene la caducidad ajustada. Lo anterior arroja los siguientes resultados:

Año Poliza	S(x)	Conservacion Promedio	Caducidad	S(x) ajustada	Conservacion ajustada	Caducidad ajustada
1	54%	54%	46%	45%	45%	55%
2	31%	58%	42%	34%	74%	26%
3	24%	76%	24%	27%	80%	20%
4	18%	74%	26%	22%	83%	17%
5	14%	81%	19%	19%	85%	15%
6	12%	84%	16%	16%	86%	14%
7	10%	87%	13%	14%	87%	13%
8	8%	80%	20%	12%	88%	12%
9	7%	80%	20%	11%	89%	11%
10	0%	0%	0%	10%	90%	10%

Al observar y analizar gráficamente todas estas tasas la S(x) ajustada y la caducidad ajustada es mayor que la real, esto se debe a que al realizar la regresión lineal toma un ajuste punto por punto tomando en cuenta las variaciones más grandes dentro de la función de supervivencia. Otro de los puntos a analizar muy claramente es que, si bien la información puntualiza que en realidad no existe caducidad en el punto 10, sin embargo, existe una tasa de caducidad al estimar en ese año y esto es claramente debido a la función exponencial con los parámetros generados a partir de la información. También es claro que la conservación y caducidad tanto ajustada como real son complemento una de otra y gráficamente se puede observar eso.



Graficando la caducidad real versus la caducidad ajustada por el modelo a partir de una regresión lineal utilizando como curva de ajuste una función exponencial se tiene lo siguiente.



En este ejercicio si se analiza la gráfica y la *R ajustada*, se observa que la curva no se ajusta de manera adecuada a la caducidad, sin embargo, es aquí cuando se tiene que analizar el porqué de ello. Analizando el ejercicio se pueden puntualizar diferentes respuestas:

- La Homogeneización de las temporalidades puede ser que no sea correcto.
- Se tiene poca información, es decir, la cartera es pequeña.
- La función de ajuste no es la adecuada

Si el primer punto fuera el caso, se requiere un análisis profundo de la cartera para hacer una homogeneización correcta acerca de plan, temporalidad e incluso la moneda con que se emitió sea la más adecuada para segmentar la cartera.

Si el segundo punto fuera el caso, puede ser porque se tienen pocas pólizas con la segmentación elegida y es necesario una reestructuración en la homogeneización respecto a la temporalidad. También puede ser que no se tengan suficientes pólizas con una moneda específica por lo que dependiendo del comportamiento de la cartera se buscará una moneda que se comporte de igual manera suavizando así la curva para aplicar un modelo como el anterior.

Para estar posicionado en el tercer punto, se deben arreglar los dos puntos anteriores para que, de esta forma, se tenga la certeza de que ya no es problema de homogeneización y falta de información si no de la función de ajuste, la cual no tiene el comportamiento adecuado por lo que el único camino es encontrar otra función de supervivencia.

CAPÍTULO III. Análisis de Componentes Principales

Uno de los problemas de la caducidad es que se pueden encontrar muchas variables que logran influir en ésta, por lo que la mayoría de las compañías utilizan el juicio experto para decidir qué variables utilizar en la homogeneización y segmentación de la cartera al proceder con la metodología de la caducidad. En este sentido y teniendo en cuenta que la problemática es conocer ¿cuál de las n-variables influyentes se deben utilizar sin recurrir o adolecer de un juicio experto? El análisis de componentes principales es la respuesta a esta problemática ya que es una técnica del análisis multivariado que ayuda a identificar las variables que más influyen dentro de la caducidad. La finalidad de utilizar esta metodología es identificar las componentes que expliquen la mayor parte de la información de los datos iniciales con objeto de hacer una homogeneización y segmentación de la cartera adecuada.

Modelo de Análisis de Componentes Principales

Al inicio de este capítulo se explicó para qué es necesario el análisis de componentes principales en este trabajo, utilizando las definiciones y propiedades de matrices es posible definir esta metodología como sigue. El análisis de componentes principales transforma el conjunto original de variables correlacionadas en otro conjunto de variables no correlacionadas, que son combinaciones lineales de las variables originales que explican la mayor parte de la varianza del conjunto original, en palabras más informales: de todas las variables que afectan a la caducidad se encontrarán aquellas que explican la mayor parte de la variabilidad en la caducidad. De esta forma, se enfocará en dichas variables para hacer una correcta homogeneización y segmentación, para posteriormente proceder con la aplicación a la metodología de la caducidad.

Las componentes principales son obtenidas de tal manera que la primera componente principal, denotada por Y_1 es la combinación lineal de las variables X_j, con $j = 1, 2, \ldots, p$

$$Y_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p$$

Donde los valores $a_{11}, a_{21}, \ldots, a_{p1}$ son elegidos de tal forma que maximicen la varianza de Y_1. La segunda componente principal Y_2 es aquella combinación lineal de las variables que no están correlacionadas con Y_1 misma que acumula la máxima cantidad de la variación total restante que no ha sido considerada por Y_1. Por lo que, en general, la m-ésima componente principal es aquella combinación lineal de $X's$ tal que:

$$Y_m = a_{1m}X_1 + a_{2m}X_2 + \cdots + a_{pm}X_p$$

La cual contiene la varianza más grande de todas las combinaciones lineales restantes que no estén correlacionadas con todas las componentes principales previamente obtenidas.

Para obtener las componentes principales primero se supondrá que se tiene un vector de p variables correlacionadas X con las características de sus distribuciones desconocidas, pero considerando que:

$$E[X] = \mu = (\mu_i)_{px1} \quad y \quad Var[X] = \Sigma = (\sigma_{ij})_{p\,x\,p}$$

Donde Σ es la matriz de varianzas y covarianzas. El desarrollo se realizará bajo estas hipótesis, de tal forma que el análisis se procederá a realizar con los estimadores a partir de la matriz de datos.

La primera componente Y_1 debe ser una combinación lineal de las variables originales de tal forma que esta variable contenga la máxima varianza posible, por lo que:

$$Y_1 = a'_1 X$$

Donde: $a'_1 = \left(a_{11}, a_{12}, \dots, a_{p1}\right)$, y $X' = (X_1, X_2, \dots, X_p)$, de esta forma:

$$Var[Y_1] = Var[\acute{a}_1 X] = a'_1 \Sigma a_1$$

Por lo que se plantea el problema:

$$Max\ a'_1 \Sigma a_1 \quad \text{sujeto a} \quad a'_1 a_1 = 1$$

La restricción se plantea para tener unicidad en la solución además de limitar la varianza de Y_1, pues cuanto más grandes sean los elementos de a_1, la varianza de Y_1 crecerá arbitrariamente.

Para resolver el problema anterior, se utilizará el método de los multiplicadores de Lagrange. Este método consiste en plantear la siguiente función:

$$\varphi = a'_1 \Sigma a_1 - \lambda(a'_1 a_1 - 1)$$

El cual lleva al problema de encontrar el valor máximo de la función anterior, por lo que derivando con respecto a a_1, se tiene:

$$a'_1 \Sigma a_1 = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} \begin{bmatrix} \sigma_{11} & \dots & \sigma_{n1} \\ \vdots & \ddots & \vdots \\ \sigma_{1n} & \dots & \sigma_{nn} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

$$= \sigma_{11} a_1^2 + \sigma_{21} a_1 a_2 + \cdots + \sigma_{n1} a_1 a_n + \sigma_{12} a_1 a_2 + \sigma_{22} a_2^2 + \cdots + \sigma_{n2} a_2 a_n + \cdots + \sigma_{1n} a_1 a_n + \sigma_{1n} a_2 a_n + \cdots + a_{nn} x_n^2$$

Derivando parcialmente la expresión anterior respecto a cada uno de los elementos de *a* se tiene:

$$\frac{da'_1 \Sigma a_1}{da_1} = 2\sigma_{11} a_1 + (\sigma_{12} + \sigma_{21}) a_2 + \cdots + (\sigma_{1n} + \sigma_{n1}) a_n$$

$$\frac{da'_1 \Sigma a_1}{da_2} = (\sigma_{21} + \sigma_{12}) a_1 + 2\sigma_{22} a_2 + \cdots + (\sigma_{2n} + \sigma_{n2}) a_n$$

Siguiendo este procedimiento hasta el *n*-ésimo término de a:

$$\frac{da'_1 \Sigma a_1}{da_n} = (\sigma_{n1} + \sigma_{1n}) a_1 + (\sigma_{n2} + \sigma_{2n}) a_2 + \cdots + 2\sigma_{nn} a_n$$

Reuniendo en un vector las derivadas respecto del escalar $a'_1 \Sigma a_1$ para cada elemento de a_1, se obtiene la derivada de dicho escalar con respecto al vector a_1:

$$\frac{da'_1\Sigma a_1}{da_2}=\left\{\begin{bmatrix}\sigma_{11} & \cdots & \sigma_{n1}\\ \vdots & \ddots & \vdots\\ \sigma_{1n} & \cdots & \sigma_{nn}\end{bmatrix}+\begin{bmatrix}\sigma_{11} & \cdots & \sigma_{n1}\\ \vdots & \ddots & \vdots\\ \sigma_{1n} & \cdots & \sigma_{nn}\end{bmatrix}\right\}\begin{bmatrix}a_{11}\\ \vdots\\ a_{nn}\end{bmatrix}$$

$$=(\Sigma+\Sigma')a_1$$

Pero como Σ es una matriz simétrica $\Sigma=\Sigma'$ por lo que el resultado anterior es:

$$\frac{da'_1\Sigma a_1}{da_n}=2\Sigma a_1 \qquad (1)$$

Por otro lado:

$$\lambda\left(a'_1 a_1-1\right)$$

$$=\lambda\left(\sum_{i=1}^{n}a_i^2-1\right)=\lambda((a_1^2+a_2^2+a_3^2+\cdots+a_n^2)-1)$$

Derivando el resultado anterior en cada entrada:

$$\frac{d\lambda(a'_1a_1-1)}{a_1}=\lambda(2a_1)$$

$$\frac{d\lambda(a'_1a_1-1)}{a_2}=\lambda(2a_2)$$

$$\vdots$$

$$\frac{d\lambda(a'_1a_1-1)}{a_n}=\lambda(2a_n)$$

Reuniendo en un vector las derivadas respecto del escalar se tiene:

$$\frac{d\lambda(a'_1a_1-1)}{a_1}=2\lambda\begin{bmatrix}a_1\\ \vdots\\ a_2\end{bmatrix} \qquad (2)$$

Tomando los resultados de (1) y (2), se tiene que la derivada es:

$$\Rightarrow\frac{d\varphi}{da_1}=2\Sigma a_1-2\lambda a_1$$

Igualando el vector restante al valor *0* se tiene lo siguiente:

$$2\Sigma a_1-2\lambda a_1=0\Leftrightarrow$$

$$2(\Sigma a_1-\lambda a_1)=0\Leftrightarrow$$

$$(\Sigma a_1-\lambda a_1)=0\Leftrightarrow$$

$$(\Sigma-\lambda I)a_1=0$$

Esta última ecuación tiene solución no trivial sí $(\Sigma - \lambda I)$ es una matriz singular, por lo que su determinante debe de ser igual al valor 0, es decir $|\Sigma - \lambda I| = 0$. Esto último ya se explicó anteriormente y es conocido como el polinomio característico de Σ y su grado es p, siendo éste el orden de Σ.

Como ya se ha visto:

$$(\Sigma a_1 - \lambda a_1) = 0 \Leftrightarrow$$

$$\Sigma a_1 = \lambda a_1$$

Por lo que para resolver este problema sólo basta con encontrar los valores propios y vectores propios de Σ.

Ahora, la varianza de Y_1 va a estar dada por:

$$Var[Y_1] = Var[\acute{a}_1 X] = a'_1 \Sigma a_1$$

Por otro lado:

$$(\Sigma a_1 - \lambda a_1) = 0 \Leftrightarrow$$

$$\Sigma a_1 = \lambda a_1 \Leftrightarrow$$

Multiplicando la igualdad anterior por a'_1 por la izquierda:

$$a'_1 \Sigma a_1 = \lambda a'_1 I a_1$$

Y $\lambda a'_1 I a_1 = \lambda$, ya que $a'_1 a_1 = 1$, lo que implica que:

$$Var[Y_1] = a'_1 \Sigma a_1 = \lambda a'_1 I a_1 = \lambda$$

$$\therefore Var[Y_1] = \lambda$$

Con el resultado anterior, la varianza de Y_1 es una raíz del polinomio $|\Sigma - \lambda I| = 0$, por lo que para formar el vector a Y_1 se tomará el vector propio asociado al valor propio más grande. De esta forma se tiene la primera componente principal:

$$Y_1 = a'_1 X$$

Ahora bien, dado que las componentes principales son combinaciones lineales no correlacionadas, ordenadas de acuerdo con su varianza decreciente, la segunda componente Y_2 tendrá la segunda varianza más grande y no estará correlacionada con Y_1:

$$Y_2 = a'_2 X$$

Donde: $a'_2 = \left(a_{12}, a_{22}, \dots, a_{p2}\right)$, y $X' = (X_1, X_2, \dots, X_p)$, de esta forma, para obtener a_2, se necesita calcular la varianza de Y_2 así como la covarianza de Y_1 y Y_2:

$$Var[Y_2] = Var[\acute{a}_2 X] = a'_2 \Sigma a_2,$$

Por definición la covarianza de dos variables aleatorias es:

$$Cov[Y_1, Y_2] = E[(Y_2 - E[Y_2])(Y_1 - E[Y_1])']$$

$$Cov[Y_1, Y_2] = E[(a'_2X - E[a'_2X])(a'_1X - E[a'_1X])']$$

$$Cov[Y_1, Y_2] = E[(a'_2X - a'_2E[X])(a'_1X - a'_1E[X])']$$

$$Cov[Y_1, Y_2] = E[(a'_2X - a'_2\mu)(a'_1X - a'_1\mu)']$$

$$Cov[Y_1, Y_2] = E[a'_2(X - \mu)(X - \mu)'a_1]$$

$$Cov[Y_1, Y_2] = a'_2Var[X]a_1$$

$$Cov[Y_1, Y_2] = a'_2\Sigma a_1$$

Como se mencionó anteriormente, la covarianza de estas dos variables debe de tener el valor *0* por lo que:

$$Cov[Y_1, Y_2] = a'_2\Sigma a_1 = 0$$

Por otra parte, del desarrollo de la primera componente principal se tiene que:

$$(\Sigma a_1 - \lambda a_1) = 0$$

Multiplicando por la izquierda a'_2 se tiene que:

$$a'_2\Sigma a_1 - \lambda a'_2a_1 = 0 \Leftrightarrow$$

$$a'_2\Sigma a_1 = \lambda a'_2a_1 \quad \text{pero} \quad a'_2\Sigma a_1 = 0$$

$$\therefore a'_2a_1 = 0$$

De esta forma se observa una restricción que será planteada al problema, esta restricción es que a_2 debe ser ortogonal con respecto a_1 de lo contrario, se tendría una correlación, por lo que el problema quedará planteado de la siguiente manera:

$$Max\ a'_2\Sigma a_2 \quad \text{sujeto a} \quad a'_2a_2 = 1 \quad \text{y} \quad a'_2a_1 = 0$$

Para resolver este problema, nuevamente se utilizarán los multiplicadores de Lagrange, definiendo la siguiente función:

$$\Psi = a'_2\Sigma a_2 - \lambda(a'_2a_2 - 1) - \gamma a'_2a_1$$

Al igual que en la primera componente principal basta con encontrar el máximo de Ψ y esto se logra derivando Ψ siguiendo el mismo procedimiento con respecto a_2 e igualando al vector a cero de la siguiente forma:

$$\Psi = a'_2\Sigma a_2 - \lambda(a'_2a_2 - 1) - \gamma a'_2a_1$$

$$\frac{d\Psi}{da_2} = 2\Sigma a_2 - 2\lambda a_2 - \gamma a_1$$

$$2\Sigma a_2 - 2\lambda a_2 - \gamma a_1 = 0$$

Multiplicando por a'_1 por la izquierda se tiene:

$$2a'_1\Sigma a_2 - 2\lambda a'_1 a_2 - \gamma a'_1 a_1 = 0$$

Donde se observa que $\gamma = 0$ debido a que la covarianza de Y_1 y Y_2 es cero lo que implica que $a'_1\Sigma a_2 = 0$ lo que a su vez implica que $a'_1 a_2 = 0$ y $a'_1 a_1 = 1$, por lo que la ecuación anterior queda reducida a:

$$\frac{d\Psi}{da_2} = 2\Sigma a_2 - 2\lambda a_2$$

$$2\Sigma a_2 - 2\lambda a_2 = 0$$

$$2(\Sigma a_2 - \lambda a_{2)} = 0$$

$$(\Sigma - \lambda I)a_2 = 0$$

Igualmente, para obtener Y_1 basta con tomar el segundo valor propio más grande de λ_2 y a_2 su vector propio correspondiente. De esta forma, la segunda componente principal será:

$$Y_2 = a'_2 X$$

Si se continúa con este procedimiento hasta la *m*-ésima componente principal estará dada por:

$$Y_m = a'_m X$$

Donde a'_m es el vector propio asociado al *m*-ésimo valor propio de Σ:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq \cdots \geq \lambda_p$$

Y su varianza será:

$$Var[Y_m] = Var[a'_m X] = a'_m Var[X] a_m = a'_m \Sigma a_m$$

Pero como $a'_m \Sigma a_m = \lambda_m a'_m I a_m = \lambda_m$. Entonces todos los componentes Y se pueden expresar como el producto de una matriz formada por los eigenvalores, multiplicada por el vector X que contiene las variables originales $x_1, \dots, x_p$

$$Y = Ax$$

Donde:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}, \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$$

Como:

$$Var(Y_1) = \lambda_1$$

$$Var(Y_2) = \lambda_2$$

$$\vdots$$

$$Var(Y_p) = \lambda_p$$

La matriz de varianzas y covarianzas será:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_p \end{bmatrix}$$

Esto debido a que cada componente principal no está correlacionada, por lo que su covarianza es igual a cero y los elementos sobre la diagonal son los eigenvalores que a su vez son la varianza de cada Y_j.

Por otra parte:

$$\Lambda = \mathrm{Var[Y]} = \mathrm{Var[A'X]} = \mathrm{A'Var[X]A} = \mathrm{A'\Sigma A}$$

$$\therefore \Lambda = \mathrm{A'\Sigma A}$$

Al despejar Σ se tendrá:

$$\Lambda = \mathrm{A'\Sigma A} \Leftrightarrow$$

$$\mathrm{A\Lambda} = \mathrm{AA'\Sigma A} \Leftrightarrow$$

Pero $\mathrm{AA'} = I$, por lo que $\mathrm{A\Lambda} = \mathrm{\Sigma A}$, y multiplicando por la derecha por $\mathrm{A'}$:

$$\mathrm{A\Lambda A'} = \mathrm{\Sigma AA'} \Leftrightarrow$$

$$\mathrm{A\Lambda A'} = \mathrm{\Sigma I}$$

$$\therefore \mathrm{A\Lambda A'} = \Sigma$$

Calculando la traza[6] de la matriz Σ, se puede comprobar que:

$$tr(\Sigma) = tr(\mathrm{A\Lambda A'})$$

$$tr(\Sigma) = tr(\mathrm{A'A\Lambda})$$

$$tr(\Sigma) = tr(\Lambda)$$

$$= \sum_{i=1}^{p} \lambda_i = \sum_{i}^{p} Var[Y_i]$$

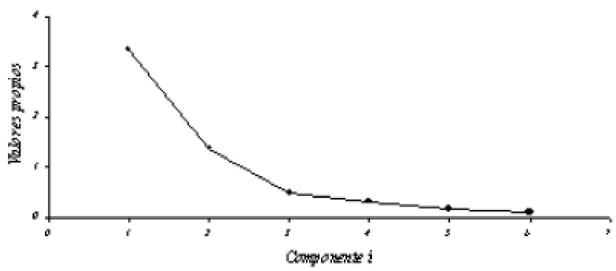
Lo que indica que la varianza total de las variables originales, expresada como la suma de varianzas, es igual a la suma de varianzas de todas las p componentes principales. La igualdad anterior arroja un resultado interesante: "*la i-ésima componente principal aporta una proporción $\lambda_i / \sum_{i=1}^{p} \lambda_i$ de la variación total de las variables originales*", por lo que las primeras m componentes principales aportan la proporción de variación total:

[6] Traza de una matriz $tr(A) = \sum_{i=1}^{n} a_{ij}$, es decir, es la suma de los valores de la diagonal.

$$\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{p} \lambda_i}$$

Llegado a este punto, es necesario saber: ¿Cuántas componentes retener? Lamentablemente la respuesta a esta pregunta es que no existe un método formal para decidir cuántas componentes retener, todo depende del juicio del que esté realizando la investigación, sin embargo es posible utilizar la lógica y decidir con respecto a la variación total, es decir, incluir solo las componentes suficientes para explicar el porcentaje significativo de variabilidad, por ejemplo de un 90% de variabilidad sería aceptable, sin embargo, también depende de la información y la búsqueda que realice el investigador teniendo en cuenta que entre más componentes mayor porcentaje de variabilidad se tiene sobre el total.

Otro método consiste en realizar una gráfica de codo, la cual se construye graficando λ_i versus i y seleccionar como el número de componente el valor de *i* correspondiente a un "codo" en la curva, esto es, el punto donde los valores propios "grandes" terminan y comienzan los valores propios "pequeños".



Gráfica 1. Componentes principales

Enfocándose en la gráfica anterior lo más adecuado es retener dos componentes principales, pues a partir de la tercera componente comienza una disminución notable en los valores propios, lo cual indica que las primeras dos componentes explican la mayor parte de la variación total.

Una problemática que existe al hacer el análisis de componentes principales a partir de la matriz de varianzas y covarianzas es que la varianza de una variable no es independiente de la escala. Supóngase que se tiene un conjunto de *p* variables las cuales se encuentran en diferente escala, en este caso, la estructura de las componentes principales derivadas de la matriz de varianzas y covarianzas dependerá de la selección arbitraria de las unidades de medición; adicionalmente, si hay grandes diferencias entre las varianzas de $X_1, X_2, \ldots, X_p$, las variables que tengan las varianzas más grandes tendrán que dominar las primeras componentes principales. Esta dificultad se puede salvar realizando una modificación de tal manera que las componentes no se vean afectadas. Esto se puede conseguir estandarizando las variables para que cada una tenga varianza igual a uno, esto significa que se utilizara la matriz de correlaciones **P** en lugar de la matriz de varianzas y covarianzas.

Para encontrar las componentes principales primero se obtendrá la matriz P a partir de la matriz de varianzas y covarianzas. Primero se define **D** a la matriz diagonal cuyos elementos sobre la diagonal son el recíproco de las desviaciones estándar de las variables X_i:

$$D = \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}$$

De esta forma la matriz **P** será de la forma:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$$

Cuyos elementos serán los coeficientes de correlación entre las variables X_i y X_j, por lo tanto, la matriz **P** será:

$$\mathbf{P} = \mathbf{D\Sigma D} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{pmatrix}$$

La derivación matemática para encontrar las componentes principales es la misma que la metodología con la matriz de varianzas y covarianzas. Sin embargo, las componentes principales obtenidas a partir de la matriz **P** no son las mismas que las obtenidas por la matriz de varianzas y covarianzas.

En la matriz de correlación, los términos sobre la diagonal son todos iguales a uno. De esta forma la traza de **P** se calcula:

$$tr(P) = p = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} Var[Y_i]$$

Entonces la proporción que aporta a la variación total la *j*-ésima componente es $\frac{\lambda_j}{p}$.

El decidir cuál de las dos matrices utilizar depende prácticamente de las diferencias importantes entre las unidades de medición de cada variable o entre sus varianzas.

EJEMPLO:

Sea X una matriz de tres variables con 10 observaciones en cada variable. Cada observación consiste en 3 medidas de agua: espesor, desplazamiento horizontal y desplazamiento vertical, misma que se represente gráficamente como se sigue:

$$X = \begin{bmatrix} 7 & 4 & 3 \\ 4 & 1 & 8 \\ 6 & 3 & 5 \\ 8 & 6 & 1 \\ 8 & 5 & 7 \\ 7 & 2 & 9 \\ 5 & 3 & 3 \\ 9 & 5 & 8 \\ 7 & 4 & 5 \\ 8 & 2 & 2 \end{bmatrix}$$

El objetivo: realizar un análisis de componentes principales.

Utilizando la herramienta de Excel, se obtienen las medias de cada variable x_1, x_2, x_3:

Media	**6.9**	**3.5**	**5.1**

Posteriormente se construye la siguiente tabla para disminuir la dificultad de los cálculos en Excel:

comp1	comp2	comp3	$(x_1 - \overline{x_1})$	$(x_2 - \overline{x_2})$	$(x_3 - \overline{x_3})$
7	4	3	0.1	0.5	-2.1
4	1	8	-2.9	-2.5	2.9
6	3	5	-0.9	-0.5	-0.1
8	6	1	1.1	2.5	-4.1
8	5	7	1.1	1.5	1.9
7	2	9	0.1	-1.5	3.9
5	3	3	-1.9	-0.5	-2.1
9	5	8	2.1	1.5	2.9
7	4	5	0.1	0.5	-0.1
8	2	2	1.1	-1.5	-3.1

Recordando el cálculo de la matriz de varianzas y covarianzas se tiene lo siguiente:

$$S = \begin{bmatrix} 2.322 & 1.6111 & -.4333 \\ 1.6111 & 2.5 & -1.2778 \\ -.4333 & -1.2778 & 7.8778 \end{bmatrix}$$

Se procede a encontrar los valores propios y valores propios con el programa “R”:

Eigenvalores	
λ_1	8.2739
λ_2	3.6761
λ_3	0.7499

	Eigenvectores		
comp1	-0.1375708	0.6990	-0.70172743
comp2	-0.2504597	0.6609	0.70745703
comp3	0.9583028	0.2731	0.08416157

Por lo tanto, las componentes principales serán:

$$Y_1 = -.1375708X_1 - .2504597X_2 + .9583028X_3$$

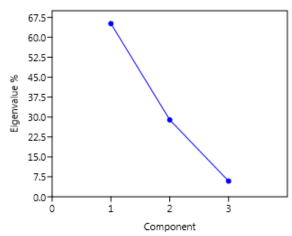
$$Y_2 = .6990X_1 + .6609X_2 + .2731X_3$$

$$Y_3 = -.70172743X_1 + .70745703X_2 + .08416157X_3$$

Y los porcentajes de variabilidad explicados por cada componente son:

Variable	Eigenvalor	% Variación
λ_1	8.2739	65.15
λ_2	3.6761	94.09
λ_3	0.7499	100.00

Por último, se realiza una gráfica de codo:



Para concluir este ejercicio, se observa de la gráfica anterior que el investigador se puede quedar con las dos primeras componentes principales debido a que son las que acumulan la mayor variabilidad de las tres componentes.

CAPÍTULO IV. Modelo de Análisis de Componentes Principales Aplicado a Seguros de Vida Individual y su Impacto en la Caducidad

En este capítulo, se aplicará un modelo de análisis de componentes principales a una base de datos reales de la compañía de seguros "A" con el fin de identificar qué variables afectan más a la reserva y así realizar con estas variables el cálculo de las tasas de caducidad. Posteriormente, una vez que se identifiquen las principales variables que afectan a la reserva se utilizará la correcta homogeneización con las variables obtenidas para realizar un estudio de caducidad adecuado para predecir la probabilidad de cancelación de las pólizas.

Datos

La información recopilada de la compañía de seguros "A" se divide en tres grandes carteras que conforman la reserva de Vida Individual: Seguros Tradicionales (seguros temporales, dotales y ordinarios de vida), Seguros Flexibles (seguro con fondo a partir de una tasa garantizada) y Seguros Universales (seguro con fondo a partir de un portafolio de inversión); dentro de ellos existen diferentes temporalidades de plazo de seguro, moneda y diferentes productos para cada cartera con su propia nota técnica, siendo la cartera de Tradicionales la más grande en reserva con cerca de 14 mil millones de pesos, seguros Universales con aproximadamente 1,200 millones de pesos y por ultimo flexibles con 1,100 millones.

Las bases que se utilizarán para cada cartera y que se trabajarán mediante un manejador de bases de datos llamado Visual Fox Pro son las siguientes:

- Trad_mmmaaaa.dbf
- Univ_mmmaaaa.dbf
- Flex_mmmaaaa.dbf

Estas bases serán anuales por cada diciembre, sin embargo, para el análisis de componentes principales se utilizará la base de cada cartera de diciembre de 2018 y para el estudio de caducidad se realizará un compilado de bases a conservar y base conservada con un histórico de 10 años.

Identificación de Variables

Para realizar el procedimiento de análisis de componentes principales, como primer paso se debe identificar las variables que afectan al cálculo de la reserva utilizando la hipótesis planteada al principio de este trabajo, por lo que las variables identificadas que afectan a la reserva son:

- Cartera: no es un componente influyente en la reserva, pero es una segmentación necesaria para la identificación de la reserva en la compañía y en donde se puede observar en qué cartera existe mayor movimiento.
- Temporalidad: es un componente importante dentro de la reserva ya que éste rige el plazo que tendrá la póliza, sin embargo, no es posible dar una temporalidad segmentada, es decir aquellos que tienen plazo *1,2, 3, …,99*, por lo que se segmentará en dos observando cuántas

pólizas existen por plazo obteniendo el porcentaje acumulado por plazo y lograr la segmentación más adecuada.

- Tipo Plan: el tipo de plan es un componente de segmentación de cada cartera ya que indica si existen pólizas únicamente temporales o dotales u ordinarios de vida. Por lo que esta segmentación será la anterior mencionada
- Prima: es un componente principal de la reserva ya que sobre la prima se calcula la reserva como se observó en el primer capítulo. La prima se segmentará en dos de acuerdo con la concentración de prima por todo el universo de carteras utilizando la media.
- Edad del asegurado: es un componente necesario para el cálculo de reserva, ya que este ayudará a identificar la probabilidad de muerte del asegurado.
- Moneda: es otro de los componentes importante dentro de la reserva ya que en la práctica a la reserva se le aplica una tasa de interés técnica por moneda, así como las variaciones por tipo de cambio que la reserva pudiera tener. En este caso se tendrá la segmentación de pesos, dólares y UDI[7] que son las monedas que utiliza la compañía "A".
- Género: un componente que no suele ser tomado en cuenta para el análisis de carteras en cuanto a reservas se refiere, sin embargo, es importante debido a que en la suscripción se suele dar un descuento al ser mujer, el cual afecta directamente a la reserva por la probabilidad de muerte de la persona. Generalmente la probabilidad de muerte de los hombres es más grande a la probabilidad de muerte de las mujeres dependiendo del estudio para la tabla de mortalidad en donde se aplique.
- Forma de pago: aunque no es un componente que afecte directamente a la reserva, es una variable que va ligada con la prima ya que la prima se puede pagar mensual, trimestral, semestral o anual. Y en muchas ocasiones, la forma de pago y el tamaño de la prima, es lo que hace que un asegurado deje de pagar y por consiguiente se cancele su póliza lo cual impacte la reserva con un flujo de salida.

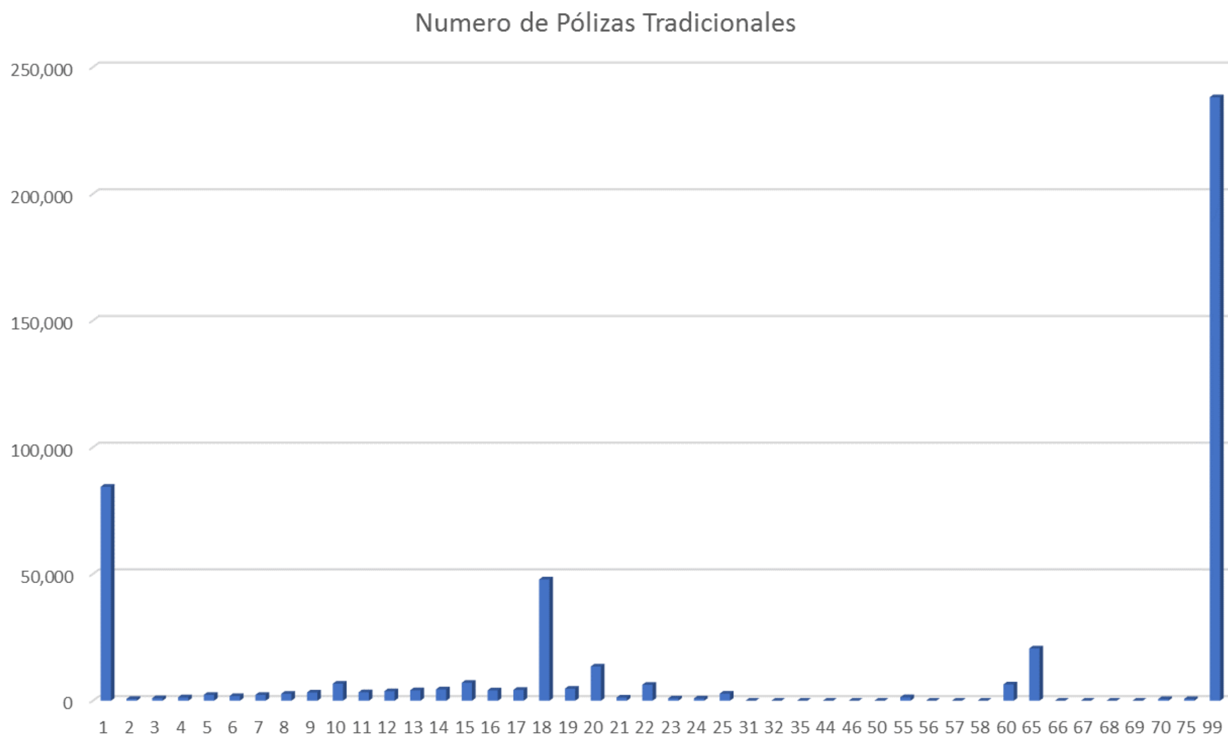
Si bien es cierto que existen otras variables que afectan directamente a la reserva como la probabilidad de muerte, la extra prima que se liga a un padecimiento médico o al giro laboral que tenga el asegurado, salario promedio, cultura financiera como ahorro entre otras; desafortunadamente no se cuenta con esa información y/o no existen dentro de las bases.

Análisis de las bases de datos

Segmentación del Plazo del Seguro

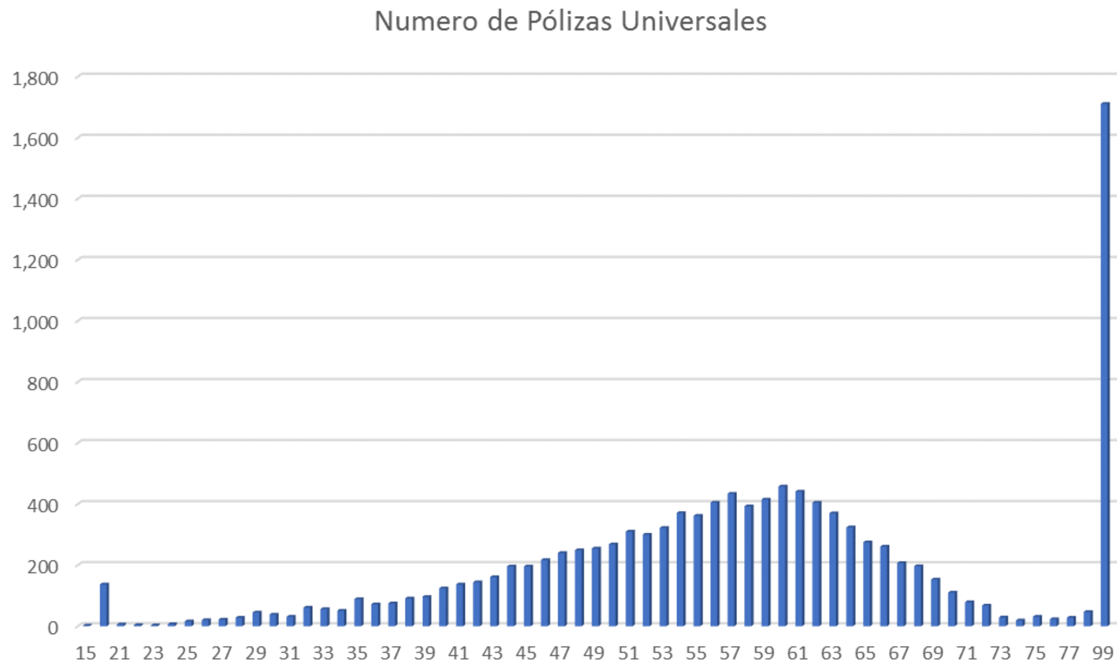
Dicho lo anterior se comenzará con el análisis para obtener la segmentación del plazo del seguro, empezando por la cartera más grande: Tradicionales. La cartera de Tradicionales tiene una combinación de plazos y temporalidades. Existen seguros temporales con diferentes plazos desde temporal 1 renovable automáticamente hasta temporal 30 y 40. Dotales corto plazo (temporalidad uno), dotal 20 en planes educacionales, dotal 30, dotal 40 y así sucesivamente hasta el dotal 60 y ordinarios de vida. Lo anterior se puede observar en la siguiente gráfica.

[7] Las Unidades de Inversión (UDI) son unidades de valor que establece el Banco de México para solventar las obligaciones de los créditos o de cualquier acto mercantil o financiero. Su valor incrementa diariamente para mantener el poder adquisitivo del dinero utilizando las variaciones del Índice Nacional de Precios al Consumidor, en otras palabras, la inflación



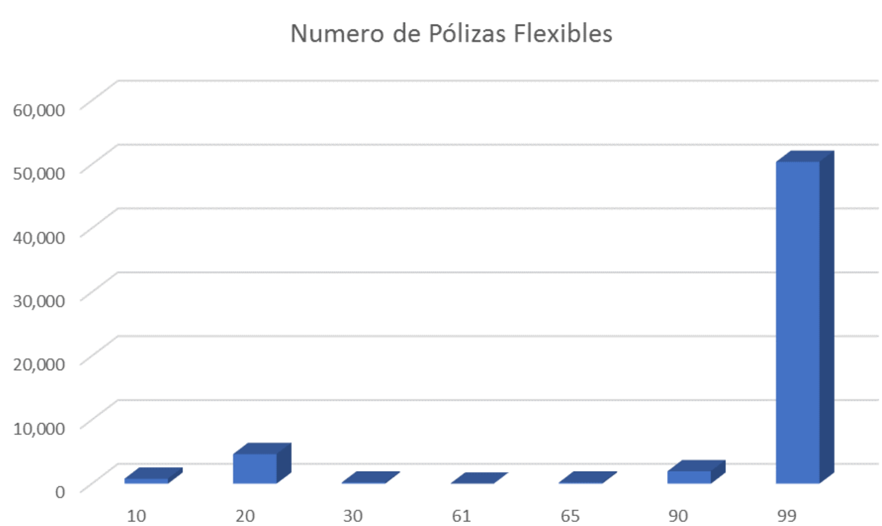
Al observar la gráfica anterior, se puede notar que la mayor cantidad de pólizas-certificados se encuentran aquellos con plazo 99, seguido por el plazo 1, seguido por plazo 18. También al observar esta misma grafica se puede notar que la mayor parte de la cartera la controlan los seguros ordinarios de vida, seguido de seguros dotales y por último los seguros temporales.

Siguiendo con la segunda cartera más grande: Universales; esta cartera al igual que la cartera de Tradicionales tiene diferentes temporalidades, sin embargo, cabe resaltar que a diferencia con Tradicionales la cartera de Universales no cuenta con planes dotales, sólo cuenta con temporales y ordinarios de vida ya que a diferencia de los dotales que cubre la supervivencia si es dotal puro, este tipo de seguros funcionan como una cuenta de ahorro en donde en cualquier momento se puede hacer uso del componente del fondo de ahorro al que están asociados, el cual se alimenta de aportaciones por parte del asegurado por lo que un plan vitalicio puede funcionar como fondo de ahorro que se puede retirar a una cierta edad y aun seguir asegurado si cuentan con el plazo de seguro 99. Lo anterior se puede observar en la siguiente gráfica:



Al observar la gráfica anterior es posible observar todas las temporalidades que existen. Lo primero que notamos es que no existen plazos menores a 15 años, y que tiene un comportamiento de numero de pólizas semejante a una campana salvo por el plazo 99 el cual tiene la mayor concentración de pólizas.

Por último, se sigue con la cartera más antigua de la compañía y también la más pequeña en cuanto a reservas se refiere: Flexibles. Esta cartera al igual que Universales no tienen planes dotales, únicamente tienen planes temporales y ordinarios de vida con sus diferentes plazos de seguros. Lo anterior se puede observar en la siguiente gráfica:

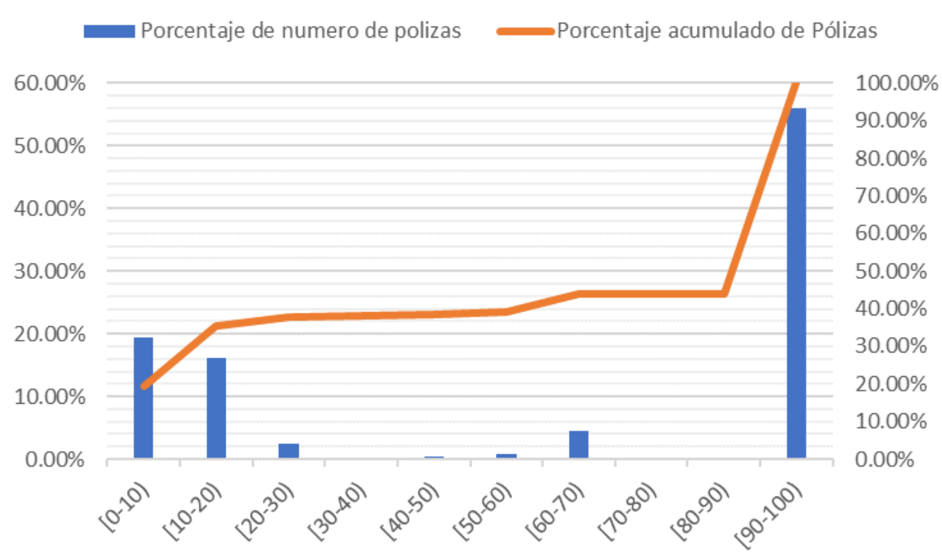


En la gráfica anterior se puede observar que la mayor concentración de pólizas-certificados se encuentra en el plazo 99 que nos lleva a los planes ordinarios de vida, seguido por el plazo 20 el cual es un seguro temporal y por último seguro vitalicio clasificado operativamente como un temporal 90.

Con lo anterior se realiza una generalización, uniendo todas las carteras y por consiguiente uniendo todos los plazos de seguros, haciendo intervalos para acumular pólizas-certificados en plazos de 10 años desde plazo 0 el cual se sabe que no existe hasta el plazo 100 que de igual manera no existe. Una vez acumuladas la cuenta de pólizas se calcula un porcentaje de pólizas por intervalo y la acumulación de los mismos porcentajes por intervalo, se puede observar esto mismo en la siguiente tabla.

Plazo de Seguro	Numero de Pólizas	Porcentaje de numero de polizas	Porcentaje acumulado de Pólizas
[0-10)	100,059	19.31%	19.31%
[10-20)	83,879	16.19%	35.50%
[20-30)	12,303	2.37%	37.88%
[30-40)	625	0.12%	38.00%
[40-50)	1,797	0.35%	38.35%
[50-60)	4,714	0.91%	39.26%
[60-70)	23,521	4.54%	43.80%
[70-80)	939	0.18%	43.98%
[80-90)	0	0.00%	43.98%
[90-100)	290,244	56.02%	100.00%

Por último, se hace un histograma de frecuencias utilizando el porcentaje de pólizas por intervalo, incluyendo una línea de tendencia asociada al porcentaje acumulado de Pólizas.



Al observar la gráfica anterior la mayor cantidad de pólizas se encuentran en el intervalo 90 a 100 y esto se debe porque en cada cartera la mayor cantidad de pólizas se localizan en el plazo 99 es decir la mayor cantidad de pólizas son aquellas de plan Ordinario de Vida. Si se focaliza en la línea de tendencia se puede notar que la primera gran acumulación de pólizas es hasta el plazo de seguro 30, sin embargo, la última parte contiene más del 50% de la acumulación se ve reflejada en el plazo 90 a 100 por lo que debido esto, se tomó la decisión de partir el plazo en este trabajo en aquellas menores o iguales a 70 años y aquellos que son mayores a 70 años.

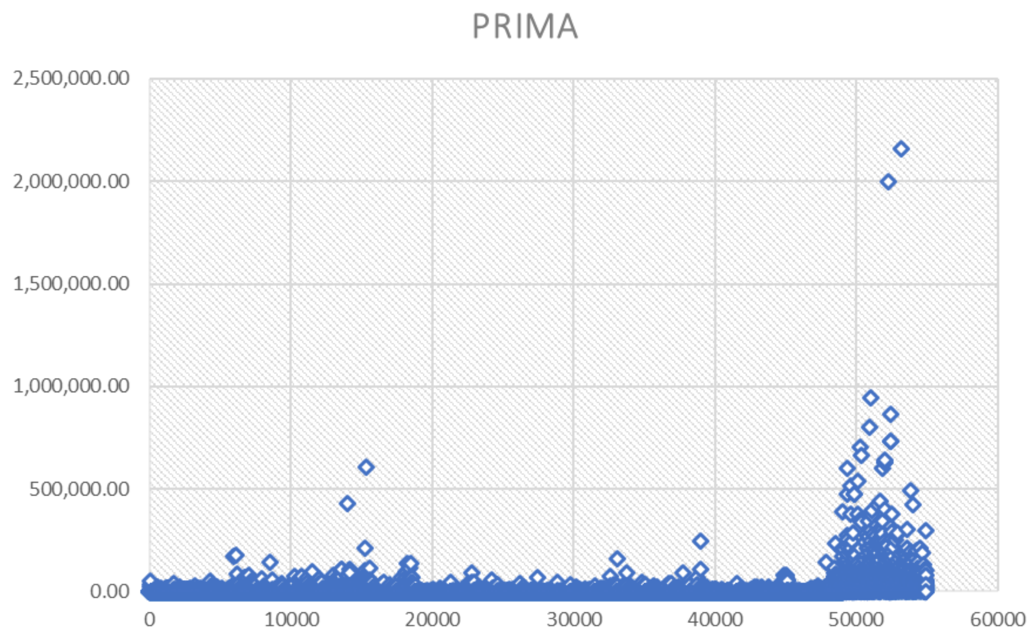
Segmentación de la Prima

Iniciando con la cartera de Flexibles, usando la base “Flex_dic2018.dbf” se identifica que el monto más grande de prima consolidada a pesos es de 2,158,615.54 millones de pesos y el monto más chico es de 20.57 pesos, por lo que se decide hacer una segmentación en primera instancia particionando el monto más grande en 10 para que se tengan 10 intervalos y de esta forma se tiene lo siguiente:

Intervalos	Número de Pólizas	Porcentaje de Número de Polizas
[0-215862)	54,883	99.86%
[215862-431724)	56	0.10%
[431724-647586)	13	0.02%
[647586-863448)	5	0.01%
[863448-1079310)	2	0.00%
[1079310-1295172)	0	0.00%
[1295172-1511034)	0	0.00%
[1511034-1726896)	0	0.00%
[1726896-1942758)	0	0.00%
[1942758-2158620)	2	0.00%
Total	54,961	100%

Analizando detenidamente la tabla anterior se tiene que en total existen 78 “outliers[8]” en la información. Si bien son relevantes en que representan aproximadamente el 15% de la prima total para esta cartera, no así en la agrupación de pólizas por primas ya que estas pólizas no representan ni el 1% de la cartera completa, por lo que es posible eliminarlas para decidir a partir de qué intervalo hacer la partición para la extracción de la información. Dicho lo anterior se tiene la siguiente gráfica y resultados.

[8] Un outlier es una observación anormal en una muestra estadística o serie de tiempo de datos que puede afectar potencialmente a la estimación de parámetros del mismo. En palabras informales un outlier es una observación dentro de una muestra o datos que no es consistente con el resto.

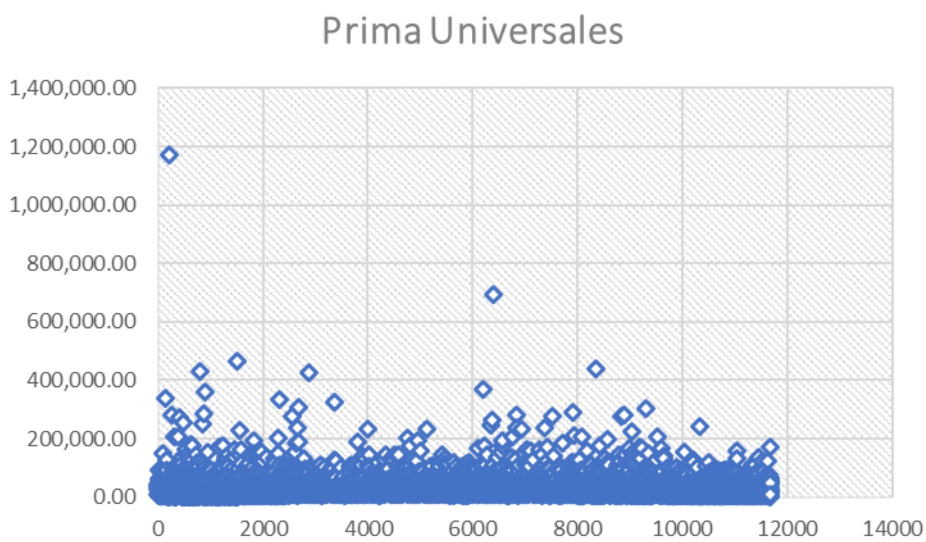


Media	3,870.95
Mediana	58.25
Desviación Estandar	23,447.59
Varianza	549,779,698.36

Observando la gráfica anterior, así como los resultados descriptivos, se puede notar que la mayoría de la prima se encuentra alrededor de la media o esperanza, sin embargo, si también se toma en cuenta la mediana, se puede notar que en la mitad de los datos ordenados de manera ascendente se encuentra en 58.25. Con lo anterior puede ser raro que la mayoría de las primas sean menores a 100 pesos mexicanos, sin embargo, esto sucede por distintas situaciones:

- Son planes que datan del año 1985 por lo que la conversión de la moneda influye dentro de la misma
- El seguro es de ahorro por lo que el componente del seguro es demasiado chico ya que está pensado en primas excedentes que genere un mayor fondo y por consiguiente mayores dividendos.

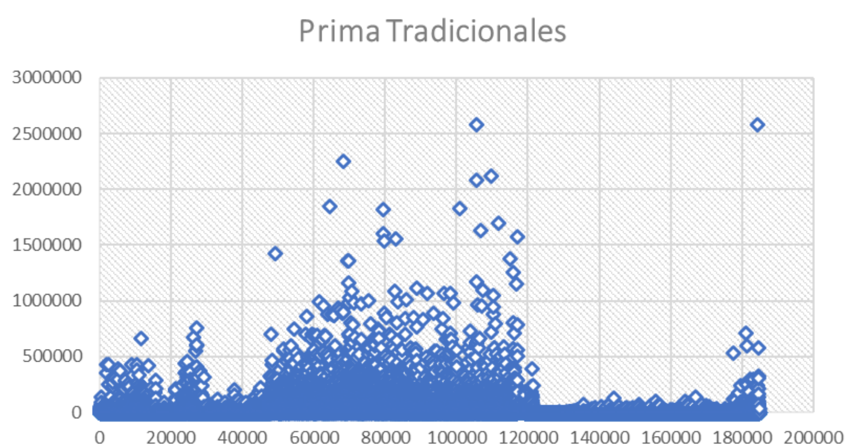
Continuando con el análisis, se prosigue con otra cartera de ahorro, pero ligado a un fondo de inversión: Seguros Universales. Los seguros universales, así como los Flexibles, tienen una prima de seguro pequeña como se muestra en el siguiente diagrama de dispersión.



Media	29,940.56
Mediana	22,402.17
Desviación Estandar	29,663.98
Varianza	879,951,606.19

Teniendo en cuenta el diagrama de dispersión, así como la mediana la mayoría de la prima se encuentra entre 20 mil y 30 mil pesos.

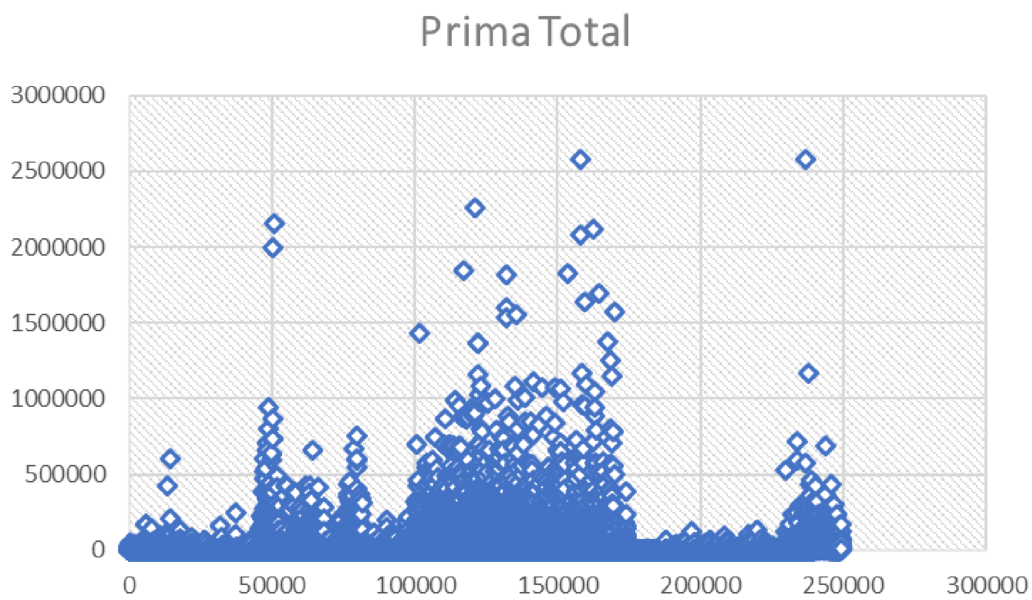
Por último, se tiene la cartera más grande la cual corresponde Tradicionales, el cual tiene los siguientes resultados:



Media	14,406.99
Mediana	452.00
Desviación Estandar	42,255.18
Varianza	1,785,500,069.87

Lo principal al ver los resultados, es que los dotales a corto plazo afectan mucho a la prima y a la dispersión de los datos, por lo cual, al ver la esperanza, la mediana y la desviación estándar, se puede deducir que los dotales a corto plazo estar captando mucha parte de la información, lo cual se ve reflejado en desviación y por consiguiente tanto la desviación como la varianza es mayor que en las carteras anteriores.

Por último, si se observa el universo de pólizas-certificados de forma general obtenemos los resultados siguientes:

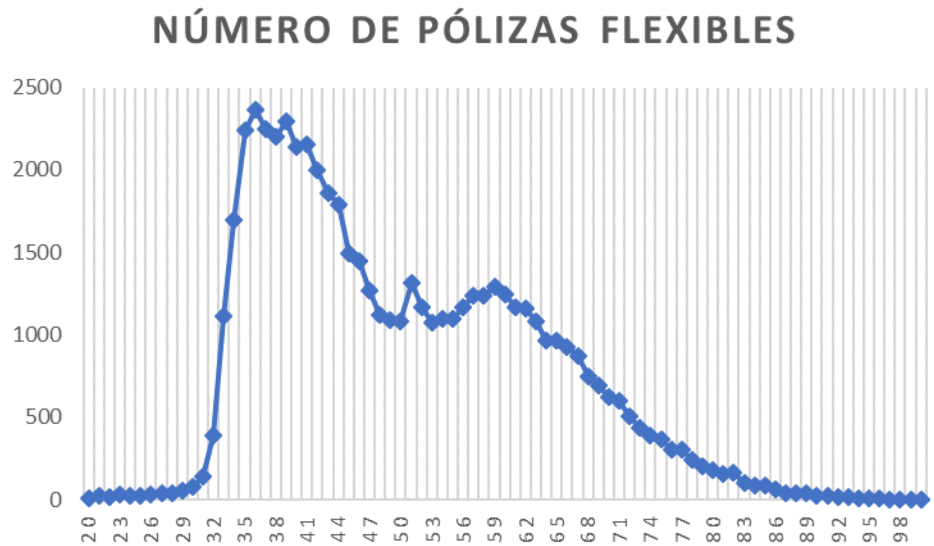


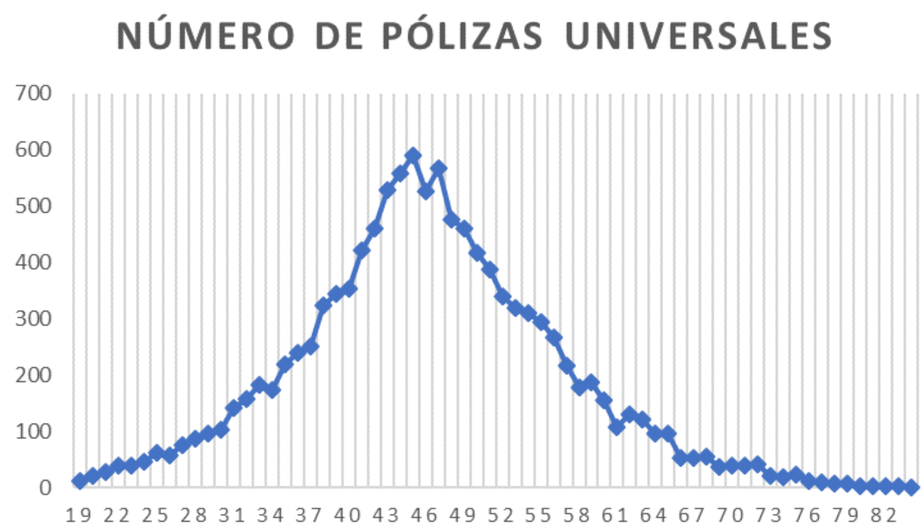
Media	12,947.63
Mediana	194.18
Desviación Estandar	39,299.94
Varianza	1,544,485,373.49

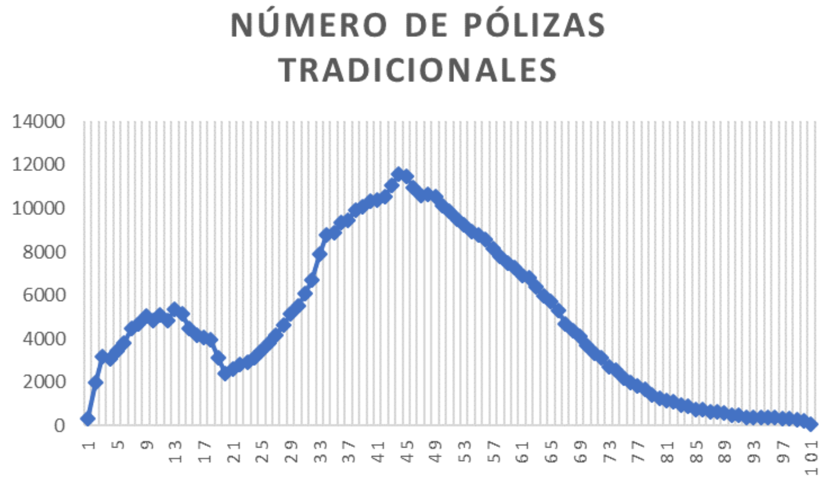
En donde se puede observar por medio de la esperanza y la varianza que los datos están agrupados alrededor de los 12,947.63 pesos, sin embargo, también es de notar que la dispersión de los datos es bastante amplia debido a las sumas aseguradas grandes de todas las carteras.

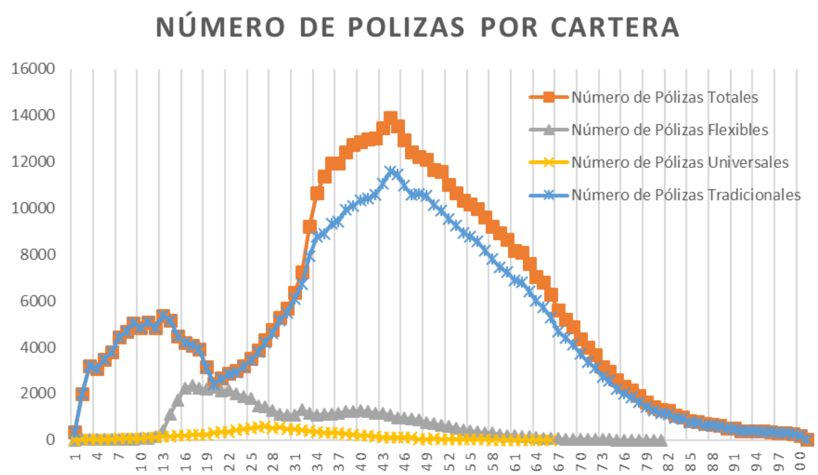
Edad del Asegurado

Otro componente bastante interesante es la edad del asegurado, esta edad indica cual será la probabilidad de muerte del asegurado por año. Iniciando con la cartera de Flexibles, la cual como se puntualizó anteriormente, esta cartera es muy longeva por lo que la mayoría de las pólizas vigentes dentro de esta cartera son de personas de edad avanzada, excluyendo claramente aquellas pólizas familiares en donde el titular de la póliza aseguraba de igual forma a menores de edad. Continuando con la cartera de Universales también es una cartera con personas de edad avanzada, sin embargo, se tienen demasiadas pólizas en edades de 19 a 40 años debido a que este tipo de producto está dedicado sobre todo a planes de ahorro para la vejez. Por último, se encuentra la cartera de Tradicionales, la cual contiene planes educacionales y planes dotales, así como ordinarios de vida los cuales hacen que tenga un comportamiento particular en cuanto edades se refiere, por un lado, muchas pólizas aunque no significativas dentro de edades tempranas y por otro lado planes dotales y ordinarios de vida, lo cuales tienen la mayor captación de la cartera y edades longevas; principalmente de edades 30 a 70 años. Lo anterior se puede observar en los siguientes gráficos.









Debido a que Tradicionales es la cartera más grande en cuanto a número de pólizas se refiere, esto se ve reflejado en número de pólizas por edad de todas las carteras, es por esta razón, que al observar la última grafica de todas la anteriores se nota que la línea marcada con naranja (que representa todas las pólizas de las tres carteras) es similar a la línea azul (cartera Tradicionales), aunque en ciertos segmentos de edad aumenta debido a las otras dos carteras.

Ahora bien, al hacer una agrupación por intervalos de edad, es posible observar en qué grupos de edad se concentran todas las pólizas logrando identificar cuál es el grupo de edad en donde se concentran más pólizas y utilizando la media aritmética dentro de los datos agrupados es posible decidir la segmentación para llevar a cabo la extracción de la información. Los resultados de lo ya mencionado son los siguientes:

Media	44.27
Desviación	19.116
Varianza	38.23

Intervalo de Edad	Número de Pólizas	Porcentaje de Polizas por Intervalo	Porcentaje Acumulado de Pólizas por Intervalo	Marca de Clase	Marca de Clase * Frecuencia
[0-10)	30,004	5.41%	5.41%	5	150,020
[10-20)	45,085	8.14%	13.55%	15	676,275
[20-30)	35,951	6.49%	20.04%	25	898,775
[30-40)	99,553	17.96%	38.00%	35	3,484,355
[40-50)	129,395	23.35%	61.35%	45	5,822,775
[50-60)	103,142	18.61%	79.97%	55	5,672,810
[60-70)	68,328	12.33%	92.30%	65	4,441,320
[70-80)	28,753	5.19%	97.48%	75	2,156,475
[80-90)	9,767	1.76%	99.25%	85	830,195
[90-100)	3,914	0.71%	99.95%	95	371,830
[100-110)	259	0.05%	100.00%	105	27,195
Total	**554,151**	**100%**	**707.30%**	**605**	**24,532,025**

Si se observa únicamente la media es claro que la segmentación debe de ser 44 años, sin embargo, hay que reafirmar este resultado y esto es posible gracias a los resultados del porcentaje acumulado. Por lo que en el intervalo de edad 40 a 50 se nota que el 61.35% de la información total está hasta este intervalo, por lo que se afirma que el promedio aritmético se debe a buena segmentación para la información, es decir, la segmentación es de **44 años**.

Hecho el análisis anterior, tomando cada uno de los análisis como agrupación queda el siguiente cuadro teniendo como base la reserva:

Cartera	TEM	OV	DOTAL	<= 70 años plazo	>70 años plazo	Pesos	Dolares	Udis	Femenino	Masculino	<= 12,947 pesos	> 12,947 pesos	<= 44 años de edad	> 44 años de edad
Tradicionales	47.89	2,154.63	14,473.97	14,525.15	2,151.34	2,491.08	13,777.95	407.46	5,444.10	11,232.39	5,181.20	11,495.29	7,887.08	8,789.41
Flexibles	1,115.51	60.29	0.00	785.06	390.73	246.57	928.99	0.23	220.29	955.50	447.91	727.88	117.23	1,058.56
Universales	1,309.87	0.00	0.00	5.69	1,304.18	311.83	988.55	9.48	387.91	921.96	523.20	786.67	439.09	870.78

Tabla 6. Agregado de cifras con concepto a evaluar en el Análisis de Componentes Principales. Cifras expresadas en millones de pesos

Análisis de componentes principales para identificar variables más influyentes en R

Como se explicó en la parte teórica la metodología para el Análisis de Componentes Principales lo primero es tener la matriz de datos, la cual ya está dada en la Tabla 6. Posteriormente de acuerdo con la metodología es obtener una matriz de varianzas y covarianzas, sin embargo, para no tener variaciones demasiado grandes dentro del análisis de componentes, se procederá a obtener la matriz de correlaciones la cual quitará las variaciones más grandes entre datos normalizando la información.

La matriz de correlaciones es posible obtener la en R mediante la siguiente función:

>Cor(matriz)

La función anterior contiene como "default" la matriz de correlación de coeficientes de Pearson.

Para guardar esta matriz en una variable se utiliza la siguiente función:

> correlacion <-cor(matriz)

Y por último para copiar la matriz de correlaciones se utiliza la siguiente función:

> write.csv(correlacion, file=”nombre_del_archivo”)

Realizando las funciones anteriores con la matriz de datos en la Tabla 6 se tiene la siguiente matriz de correlaciones:

	TEM	OV	DOTAL	men.70.años.plazo	may70.años.plazo	Pesos	Dolares	Udis	Femenino	Masculino	men.12947.pesos	may.12947.pesos	men.44.años.de.edad	may.44.años.de.edad
TEM	1	-0.99293437	-0.989719432	-0.9954164	-0.7719771	-0.98574343	-0.98913591	-0.98667593	-0.98528745	-0.99011905	-0.9876383	-0.989030178	-0.9838202	-0.99247756
OV	-0.99293437	1	0.999698182	0.99973175	0.84195191	0.99874453	0.99959123	0.99901096	0.99860622	0.99976352	0.99926073	0.999570462	0.998128713	0.99999282
DOTAL	-0.98971943	0.99969818	1	0.99886102	0.85495306	0.99967375	0.9999919	0.99980181	0.99960145	0.99999602	0.99990361	0.999988759	0.999329691	0.9997841
men.70.años	-0.9954164	0.99973175	0.998861019	1	0.82922957	0.99731641	0.99866094	0.99771313	0.99711595	0.99899168	0.99810227	0.998623558	0.99644473	0.99963681
may70.años.	-0.7719771	0.84195191	0.854953062	0.82922957	1	0.86792298	0.85703335	0.86511022	0.86925539	0.85348604	0.86207239	0.857402849	0.873368909	0.84399051
Pesos	-0.98574343	0.99874453	0.999673746	0.99731641	0.86792298	1	0.99976843	0.99998412	0.99999639	0.99959769	0.99993202	0.999783615	0.999938711	0.99892719
Dolares	-0.98913591	0.99959123	0.999991904	0.99866094	0.85703335	0.99976843	1	0.99987382	0.99970696	0.99997657	0.99995138	0.999999743	0.999468908	0.9996924
Udis	-0.98667593	0.99901096	0.999801808	0.99771313	0.86511022	0.99998412	0.99987382	1	0.99996536	0.99974165	0.99998185	0.999884964	0.999860445	0.99917228
Femenino	-0.98528745	0.99860622	0.999601453	0.99711595	0.86925539	0.99999639	0.99970696	0.99996536	1	0.99951782	0.99989705	0.999724067	0.999964865	0.99879906
Masculino	-0.99011905	0.99976352	0.999996019	0.99899168	0.85348604	0.99959769	0.99997657	0.99974165	0.99951782	1	0.99986046	0.9999714	0.999222416	0.99983875
men.12947.p	-0.9876383	0.99926073	0.999903612	0.99810227	0.86207239	0.99993202	0.99995138	0.99998185	0.99989705	0.99986046	1	0.999958202	0.999741639	0.99939924
may.12947.p	-0.98903018	0.99957046	0.999988759	0.99862356	0.85740285	0.99978361	0.99999974	0.99988496	0.99972407	0.9999714	0.9999582	1	0.999492033	0.99967434
men.44.años	-0.9838202	0.99812871	0.999329691	0.99644473	0.87336891	0.99993871	0.99946891	0.99986045	0.99996487	0.99922242	0.99974164	0.999492033	1	0.99835327
may.44.años	-0.99247756	0.99999282	0.999784102	0.99963681	0.84399051	0.99892719	0.9996924	0.99917228	0.99879906	0.99983875	0.99939924	0.999674344	0.998353268	1

Tabla 7. Matriz de Correlaciones a partir de la tabla de datos en R.

Analizando la matriz anterior es posible observar que en la diagonal toma el valor de uno en todas sus entradas lo cual es correcto. En segunda instancia es posible observar que todas las variables están correlacionadas, lo cual tiene sentido debido a la construcción y a que cada uno forma parte de la metodología de la reserva o es parte de la afectación a la reserva. Por otro lado, es posible notar que existe correlación negativa de la etiqueta TEM (la cual es temporal) hacia todas las demás variables. Por último, existen algunas variables que están correlacionadas pero su correlación no es tan fuerte como las demás las cuales se encuentran la mayoría en .99. Al observar esto es posible seguir con el análisis posterior.

Como siguiente paso se procederá a realizar el Análisis de Componentes Principales en “R” por medio de la siguiente función:

>Prcomp(matriz_de_datos, scale=TRUE)

Dentro de “scale” es posible escribir “TRUE” o “FALSE”, con la primera realiza en análisis por medio de una matriz de correlaciones, la segunda utiliza una matriz de varianzas y covarianzas de los datos.

Es posible guardar la función anterior, en una variable de la siguiente forma:

>Nombre_variable<-prcomp(matriz_datos, scale=TRUE)

Por último, para obtener un resumen del análisis de componentes principales, así como los vectores propios de cada componente, se utilizan las siguientes funciones:

>Summary(nombre_variable)

>nombre_variable

Utilizando los datos, se tiene las siguientes salidas de R:

```
Importance of components:
                          PC1     PC2       PC3
Standard deviation     3.7030 0.53636 4.621e-16
Proportion of Variance 0.9795 0.02055 0.000e+00
Cumulative Proportion  0.9795 1.00000 1.000e+00
```

```
 Variances:


 Comp. 1    Comp.2    Comp. 3
   13.712209  0.28768205  2.13536E-31
```

Lo anterior puntualiza que solo existen tres componentes principales, en donde el primer componente contiene la mayoría de la variación acumulada con 97.95%. Lo cual se puede corroborar con las varianzas o eigenvalores y con la gráfica de "scree".

De esta forma es posible corroborar que la primera componente es la que tiene el mayor porcentaje de variabilidad ya que contiene la mayor varianza de las tres componentes

Las cargas o vectores propios son que corresponden a las correlaciones de las variables originales con las componentes y pueden observarse en la siguiente tabla:

Loadings:

	PC1	PC2	PC3
TEM	0.2662131	0.31317859	-0.68834224
OV	-0.269715	-0.09286865	-0.36539942
DOTAL	-0.2699641	-0.04709404	0.24734953
men.70.años.plazo	-0.2693311	-0.1359716	0.13685572
may70.años.plazo	-0.2343449	0.92651182	0.2246418
Pesos	-0.2700502	0.00052738	-0.04004321
Dolares	-0.2699893	-0.03959384	-0.04469696
Udis	-0.2700464	-0.00997895	0.05894932
Femenino	-0.270049	0.00554052	-0.23985509
Masculino	-0.2699438	-0.05235295	-0.14608916
men.12947.pesos	-0.2700328	-0.02121209	-0.3389865
may.12947.pesos	-0.2699934	-0.03825636	0.05099129
men.44.años.de.edad	-0.2700328	0.02116895	-0.23822945
may.44.años.de.edad	-0.2697641	-0.0858115	0.04559577

Si bien en cierto que la variable que más tiene correlación con la segunda componente principal es may70.años.plazo tomar esta componente no nos da mucha información en cuanto que variables son las que más afectan ya que los vectores propio son muy bajos en toda la componente, también es cierto que se ha disminuido todas las variables a una sola dimensión, sin embargo, el objetivo aún sigue siendo el encontrar ¿Qué variables contienen la mayor variabilidad? Esto es posible observando los pesos de cada variable obtenidos mediante los vectores propios. Con el análisis anterior al disminuir a una sola dimensión todas las variables ahora solo basta con obtener el valor absoluto de los pesos de esta componente y ordenar de mayor a menor para así obtener las variables con mayor peso que influyen en la variabilidad del primer componente.

Loadings:

Pesos	0.2700502
Femenino	0.2700490
Udis	0.2700464
men.12947.pesos	0.2700328
men.44.años.de.edad	0.2700328
may.12947.pesos	0.2699934
Dolares	0.2699893
DOTAL	0.2699641
Masculino	0.2699438
may.44.años.de.edad	0.2697641
OV	0.2697150
men.70.años.plazo	0.2693311
TEM	0.2662131
may70.años.plazo	0.2343449

Al observar los datos anteriores se tienen resultados interesantes, el primero de ellos son las variables de las tres monedas (Pesos, Udis, Dólares), lo cual indica que para el estudio de caducidad se deben de tomar la segmentación de la moneda. Aunado a lo anterior aparece la variable “Femenino” la cual indica que nuestro estudio solamente se tomara en cuenta aquellas pólizas con sexo femenino, posteriormente se tiene la variable “men.12947.pesos” y “may.12947.pesos” por lo que el estudio de caducidad necesita contemplar toda la prima y por último la variable “men.44.años.de.edad” la cual indica que el estudio de caducidad solo debe incluir a las personas menores a 44 años.

En resumen, se tiene lo siguiente para el Estudio de Caducidad:

- Segmentación por moneda: Pesos, Dólares y Udis.
- Solo se tomarán en cuenta las pólizas del sexo Femenino
- El estudio se realizará por prima ya que es una variable con mayor variabilidad
- Solo se tomarán en cuenta aquellas pólizas cuya edad sea menor a 44 años.

Elaboración de bases a conservar y conservadas para estudio de caducidad

En el apartado anterior, se encontraron las variables que segmentarán y/o utilizarán en el estudio de caducidad. Estas variables tienen el objetivo de que sea la mejor segmentación del universo de pólizas para calcular las tasas de caducidad las cuales se aplicarán a toda el vigor para poder estimar cómo posiblemente se observara dicha reserva.

Sin embargo, antes de realizar el estudio de lleno la segmentación de las variables, se debe tener una lógica. Por ejemplo, tiene sentido realizar el estudio de caducidad por prima ya que capta la variabilidad entre primas muy chicas y grandes. Siguiendo con las pólizas únicamente de sexo femenino, en esta cartera son las que tiene mayor variabilidad según el Análisis de Componentes Principales, sin embargo, el investigador tiene que tomar en cuenta si tiene mucha o poca información de pólizas de sexo femenino debido a que esto es un factor importante en el estudio de caducidad por lo que puede dar resultados inesperados o con demasiada variabilidad de un año a otro o en general en el estudio. Por otro lado, se encuentra la moneda, la cual tiene mucho sentido, ya que normalmente en la práctica se tiene una tasa de interés técnica aplicable a las reservas diferentes para cada moneda, y por último al igual que en el sexo el ACP[9] arroja que sólo se tienen que tomar aquellas pólizas de edad menor a 44 por lo que el investigador tiene que verificar que cuenta con suficiente información para pólizas menores a 44 años, en caso contrario se deberían discriminar estas variables en el ACP por el simple hecho de que el ACP arroja demasiada variabilidad por tener poca información y así volver a realizar el estudio.

Todo esto se realizará por medio de Fox Pro (manejador de datos o con cualquier manejador dado que se trabaja con más información de la que Excel puede manejar). Como primer punto se procederá a realizar una consulta para todas las bases de datos de las tres carteras, posterior a lo anterior, se hará una consulta para un correcto manejo de datos con la segmentación de plan, prima, moneda y plazo de seguro con una historia de 10 años[10], por último, se obtendrá la salida por cartera y se compilaran todos estos outputs en una hoja Excel.

Al igual que en el capítulo II se elaborará una base a conservar y una base conservada por año en donde la base a conservar contendrá todas las primas de las pólizas en vigor al año t y la base conservada contendrá todas las primas de las pólizas que se mantuvieron vigentes al final de $t+1$ sin incluir pólizas que se emitieron de t a $t+1$, esto se realizará para cada año. Posteriormente se calculará un cociente por año para generar una media de todos los años por año-póliza para que de esta forma

[9] ACP es la abreviatura del “Análisis de Componentes Principales”

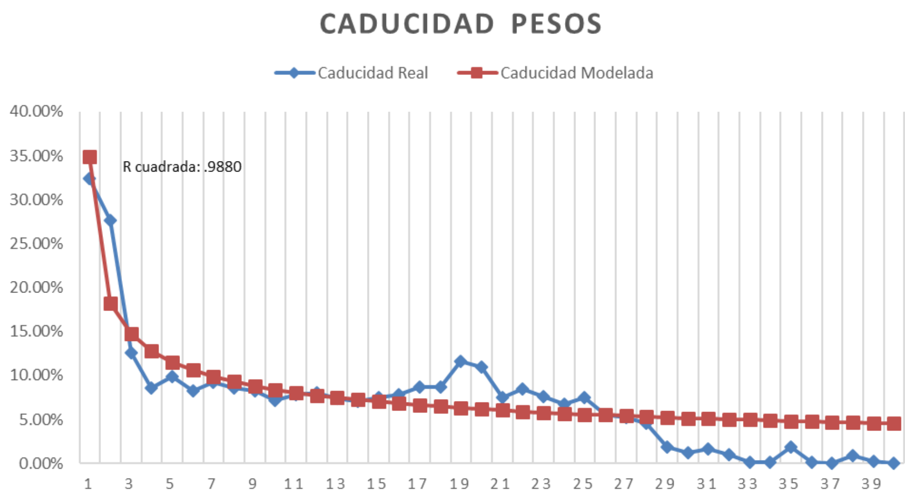
[10] De acuerdo con la “Tasas de Caducidad: Guía de Apoyo para la Construcción y Aplicación” menciona que el periodo de desarrollo de tasas de caducidad es de 10 a 15 años. Debido a que solo se tienen 10 años de historia disponible, este será la historia que se utilizará

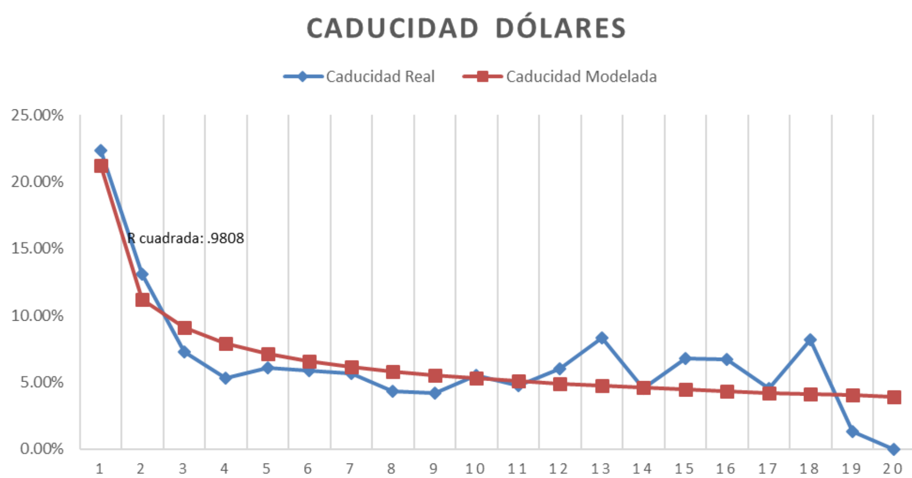
se realice el estudio por medio de supervivencia con un modelo asociado a una variable aleatoria exponencial de la siguiente manera $S(x) = \exp\left(-\alpha x^{\beta}\right)$ y una regresión lineal. Los resultados son los siguientes:

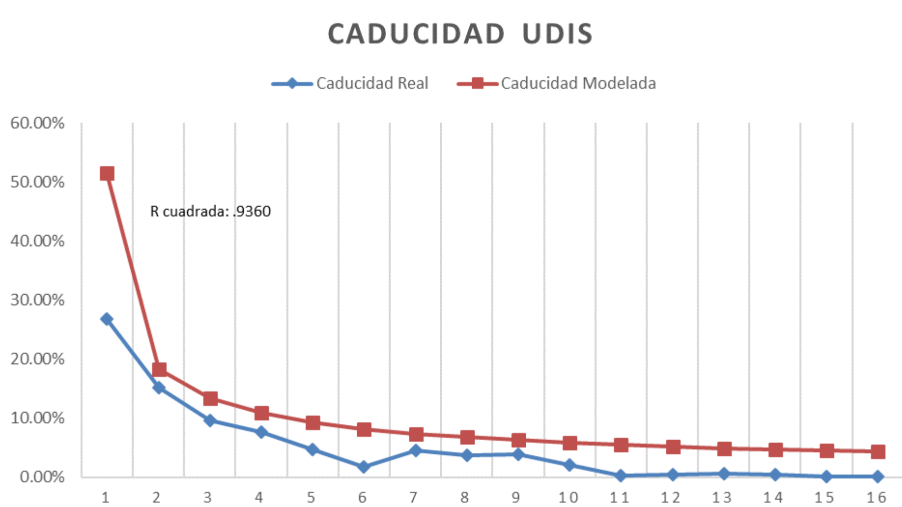
Moneda	Año poliza	Caducidad Real	Caducidad Modelada
Pesos	1	32.38%	34.93%
	2	27.65%	18.27%
	3	12.57%	14.74%
	4	8.58%	12.80%
	5	9.86%	11.52%
	6	8.28%	10.59%
	7	9.21%	9.87%
	8	8.59%	9.28%
	9	8.26%	8.80%
	10	7.20%	8.40%
	11	7.82%	8.05%
	12	8.02%	7.74%
	13	7.44%	7.47%
	14	7.05%	7.23%
	15	7.45%	7.01%
	16	7.80%	6.81%
	17	8.66%	6.63%
	18	8.69%	6.46%
	19	11.66%	6.31%
	20	10.93%	6.17%
	21	7.53%	6.04%
	22	8.45%	5.92%
	23	7.62%	5.80%
	24	6.75%	5.69%
	25	7.52%	5.59%
	26	5.54%	5.50%
	27	5.18%	5.40%
	28	4.53%	5.32%
	29	1.88%	5.24%
	30	1.19%	5.16%
	31	1.66%	5.09%
	32	0.99%	5.01%
	33	0.18%	4.95%
	34	0.15%	4.88%
	35	1.87%	4.82%
	36	0.12%	4.76%
	37	0.06%	4.70%
	38	0.87%	4.65%
	39	0.27%	4.60%
	40	0.00%	4.55%

Moneda	Año poliza	Caducidad Real	Caducidad Modelada
Dólares	1	22.37%	21.21%
	2	13.10%	11.18%
	3	7.28%	9.07%
	4	5.28%	7.91%
	5	6.07%	7.15%
	6	5.86%	6.59%
	7	5.63%	6.16%
	8	4.34%	5.81%
	9	4.19%	5.52%
	10	5.54%	5.28%
	11	4.73%	5.07%
	12	6.02%	4.88%
	13	8.32%	4.72%
	14	4.56%	4.58%
	15	6.79%	4.44%
	16	6.69%	4.32%
	17	4.55%	4.22%
	18	8.18%	4.11%
	19	1.30%	4.02%
	20	0.00%	3.94%

Moneda	Año poliza	Caducidad Real	Caducidad Modelada
Udis	1	26.75%	51.58%
	2	15.12%	18.25%
	3	9.56%	13.32%
	4	7.62%	10.83%
	5	4.70%	9.27%
	6	1.81%	8.18%
	7	4.58%	7.37%
	8	3.63%	6.74%
	9	3.88%	6.23%
	10	2.00%	5.81%
	11	0.27%	5.46%
	12	0.49%	5.16%
	13	0.59%	4.89%
	14	0.49%	4.66%
	15	0.02%	4.45%
	16	0.01%	4.27%


CADUCIDAD PESOS
Caducidad Real
Caducidad Modelada
R cuadrada: .9880
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39

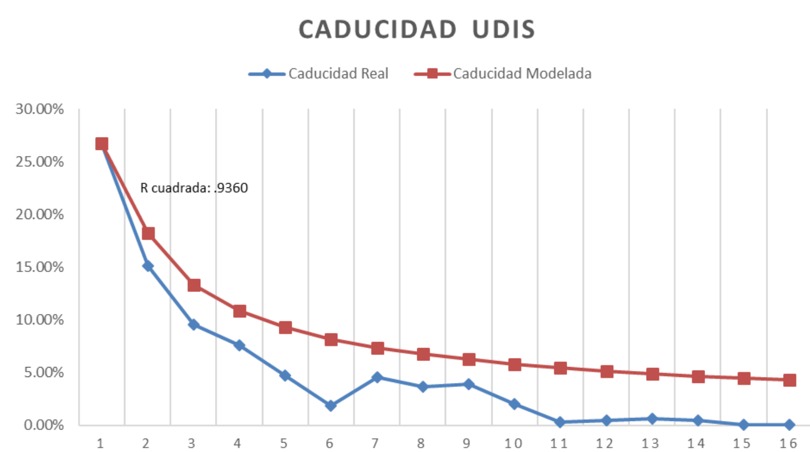

CADUCIDAD DÓLARES
Caducidad Real
Caducidad Modelada
R cuadrada: .9808
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20


CADUCIDAD UDIS
Caducidad Real
Caducidad Modelada
R cuadrada: .9360
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Observando las tablas anteriores, así como las gráficas de la caducidad de cada moneda es posible puntualizar lo siguiente:

- La moneda Pesos y Dólares son los que tienen más captación en cuanto a monto de prima se refiere y esto se ve reflejado en el número de años póliza que tiene cada moneda junto con su caducidad, sin embargo, la moneda pesos es la que tiene más información de las tres monedas.
- La moneda Udi es la que menor información contiene de las tres.
- En todas las monedas el coeficiente de determinación (R cuadrada) es mayor a 0.9 lo cual significa que el ajuste es muy confiable, es más, es posible decir que el modelo explica al menos el 90% de la variabilidad real.
- En algunos casos es posible observar que la caducidad real es cero y en la caducidad modelada es distinta de cero, esto se debe a que el modelo recoge la variabilidad entre puntos ajustando una curva por lo que dentro del modelo al seguir una curva exponencial decreciente no arroja el valor igual a cero.
- Otro punto importante es que la caducidad indicará que la única distinción al aplicarse a todo el universo de pólizas (sea la cartera que sea), sólo tendrá la distinción de la moneda y no se va a aplicar únicamente a la segmentación obtenida por el ACP.
- La caducidad puede beneficiar a algunos productos y puede afectar a algunos otros, por lo que esta caducidad dependiendo de las necesidades del negocio puede cambiar tanto en el ACP.

Por último, es relevante observar la caducidad de la moneda Udis, esta caducidad y en especial la modelada tiene la peculiaridad de que en el primer año-póliza es muchísimo mayor a la real, por lo que es posible hacer que en el primer año al menos tenga la caducidad real. Lo anterior debido a que afectaría mucho a la reserva y en segundo lugar porque no es algo que esté reflejando la realidad al menos en el primer año póliza. Por lo que al tener el primer año-póliza igual tanto real como modelada se tendría lo siguiente:



Es de suma importancia hacer notar que la caducidad puede llegar a variar en resultados cuando se utiliza el número de pólizas en vez de la prima, o la suma asegurada en vez de la prima. El cuándo utilizar una de la otra es complicado debido a que el investigador puede desear captar variación en prima emitida que va ligada a la utilidad de la empresa, o en suma asegurada que va ligada a la variación de la reserva o algo más conservador en donde solamente tenga variación en la cartera en cuanto a pólizas se refiere.

Valuación de Reservas considerando el índice y tasas de caducidad calculadas

Ya teniendo las tasas de caducidad calculadas con el modelo ACP y el estudio de caducidad por medio de una regresión lineal ajustada a un modelo de supervivencia exponencial, es necesario valuar las mismas para observar el impacto por caducidad dentro de las mismas reservas.

Hasta antes de realizar la valuación de la reserva con las nuevas tasas de caducidad se tiene los siguientes saldos de reservas en la compañía "A":

Tipo de Cambio Cierre	
Dólares	19.6512
Udis	6.226631

Flexibles	
Moneda	Reserva
1	246,574,734.18
2	47,273,922.92
3	36,452.28
Total	1,175,791,023.16

Universales	
Moneda	Reserva
1	311,832,124.03
2	50,304,950.90
3	1,523,139.65
Total	1,309,868,803.72

Tradicionales	
Moneda	Reserva
1	2,491,081,012.59
2	701,124,968.34
3	65,438,583.87
Total	16,676,489,905.42

Utilizando la nueva caducidad en el modelo de reservas locales se tiene lo siguiente:

Tipo de Cambio Cierre	
Dólares	19.6512
Udis	6.226631

Flexibles	
Moneda	Reserva
1	246,021,086.00
2	47,212,414.61
3	36,452.28
Total	1,174,028,662.88

Universales	
Moneda	Reserva
1	311,815,547.87
2	50,304,437.81
3	1,523,139.65
Total	1,309,842,144.72

Tradicionales	
Moneda	Reserva
1	2,329,392,876.62
2	682,630,415.67
3	53,941,502.03
Total	16,079,773,529.75

De acuerdo con lo anterior se puede observar que en todas las carteras se libera reserva: Flexibles y Universales lo hacen en menor medida debido a que están ligadas al fondo y se tiene un máximo entre el fondo y la reserva. En la cartera de Tradicionales es donde se puede observar un mayor impacto debido a que es la cartera que concentra la mayor parte de la reserva, lo anterior sucede debido a que las tasas de caducidad modeladas en este trabajo son más grandes a las originales, sin embargo, no por ser mayores las tasas de caducidad significa que estas son incorrectas debido a que al observar los coeficientes de correlación estos se ajustan con una gran confiabilidad, contrario a lo que había por segmento de las tasas de caducidad originales.

Técnicamente hablando un aumento en la caducidad únicamente dentro de los flujos que sirven para calcular el BEL o reserva de riesgos en curso significa que los rescates de pólizas aumentan, es decir, la probabilidad de que las pólizas sean rescatadas es más grande por lo que las obligaciones de la empresa hacia el asegurado disminuyen en cuanto a estimación se refiere, entendiendo que el método BEL implica una estimación de los ingresos y egresos futuros traídos a valor presente.

Dicho lo anterior si observamos un P&L con esta variación encontramos que se tendría una liberación en el rubro de los pasivos de reservas lo cual implica que al ver el resultado técnico se obtendría una mejora debido a que la obligación de la empresa disminuye, por lo que el resultado antes de impuestos también tendría una mejora traduciéndose en una utilidad mayor antes de impuestos, sin embargo, al solo pensar en una reducción en las obligaciones de la compañía es posible realizar una disminución en la suma asegurada, no obstante, es necesario analizar disminución suma asegurada contra aumento de caducidad: disminuir suma asegurada implica también una reducción del BEL, que tan grande es esta disminución depende de que tan grande es la reducción de suma asegurada en la cartera en general, por otro lado, aumento en caducidad afecta directamente a que tan rentable es un producto no así la suma asegurada, por lo que es necesario analizar que no es lo mismo bajar un 30% la suma asegurada contra un aumento de 3% en la caducidad ya sea en cada nodo o tan solo en el primer año póliza. Por lo que ahora es necesario entender que la caducidad afecta directamente a la rentabilidad en un producto y a la utilidad de una empresa.

Conclusiones

El desarrollo de este trabajo se centró en explicar un método estadístico para encontrar las tasas de caducidad adecuadas para un portafolio de carteras en una compañía de seguros y del tratamiento que se le puede dar a las mismas, siguiendo la legislación mexicana en cuanto a la información para el cálculo de las reservas de riesgos en curso. Para ello, fue necesario utilizar dos métodos de cálculo, uno enfocado a encontrar aquellas variables que afectan más a la reserva por medio de la correlación entre estas variables y otro para calcular las tasas de caducidad a partir de las variables que afectan más a las reservas, las cuales fueron previamente encontradas. Lo anterior resulta una herramienta muy útil ya que se quita de por medio el juicio experto, dejando únicamente fundamentos matemáticos y estadísticos, así como una metodología sumamente confiable para el cálculo de las tasas de caducidad. Dicho lo anterior y analizando únicamente la caducidad estudiada se tiene que tomar en cuenta que dentro de la reserva de riesgos en curso o comúnmente conocida como BEL afecta a los flujos de rescates que afecta directamente a las primas de la pólizas: al subir las tasas de caducidad aumentan los rescates hacia la prima lo cual genera menores obligaciones para la compañía, en caso contrario al bajar las tasas de caducidad aumentan las obligaciones de la compañía aumentando el BEL y por consiguiente una disminución en las utilidades de la empresa. Todo lo anterior da como resultado que este trabajo siga pasos estadisticos que pueden indicarnos que la reserva o "*Best estimate*" no sea el que por normas y método de calculo de la caducidad dada por el regulador mexicano CNSF sea el adecuado, este método se adecua estadísticamente a las carteras de cada compañía con las estructuras que las compañías tengan para su negocio.

En cada capítulo se encuentra una interpretación y justificación de los elementos utilizados, con la finalidad de proporcionar las herramientas necesarias para la utilización de este trabajo, esperando que sea de motivación para profundizar este tema tan importante dentro de las compañías de seguros, una nueva aplicación en el desarrollo del cálculo de las tasas de caducidad puede aplicarse con modelos estocásticos para crear vigores dinámicos que permitan hacer una mejor estimación de las posibles obligaciones de las compañías de seguros, esto solo por mencionar algunos tópicos que no se tocan en este trabajo por pertenecer aún trabajo más extenso en el estudio de las probabilidades de cancelación de pólizas.

Como comentario final toda esta metodología puede utilizarse en diferentes tópicos y pueden ser de mucha utilidad, por ejemplo: en la disciplina de la medicina se puede observar que de una enfermedad a la que puede aplicarse diferentes medicinas cuál de ellas son las más efectivas y a partir de esas variables calcular una probabilidad de falla de estas a partir de la metodología de la caducidad. Aunado a lo anterior la caducidad puede utilizarse como indicador de ventas y como indicador de riesgos por malas prácticas. La intención de este comentario es concientizar de lo que se puede realizar con un tema en gran medida sencillo de entender y de enfocar, la cual tiene un amplio campo de estudio.

ANEXOS

Matrices

Sea K un campo[11] y sean n, m dos enteros ≥ 1. Un arreglo de números en K:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}$$

Se denomina matriz en K. Se dice que es una matriz de orden m por n, o matriz de m x n en donde la matriz tiene m renglones y n columnas. Se puede abreviar la forma gráfica de la matriz escribiendo la siguiente notación (a_{ij}), $i = 1, 2, ..., m$ y $j = 1, 2, \ldots, n$.

La primera columna es:

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}$$

El segundo renglón es:

$$(a_{21}, a_{22}, \ldots, a_{2n}).$$

Cuando se utiliza la notación a_{ij}, se hace referencia a la *ij*-ésima entrada o *ij-ésimo* componente de la matriz. Aunado a lo anterior se puede denotar por A a la matriz anterior por lo que el *i*-ésimo renglón se denota como A_i, y es definido como:

$$A_i = (a_{i1}, a_{i2}, \ldots, a_{in}).$$

La *j*-ésima columna se denota como A^j y es definida como:

$$A^j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Por otra parte si (a_{ij}), $i = 1, 2, ..., m$ y $j = 1, 2, \ldots, n$ una matriz y n = m se dice que es una **matriz cuadrada**. Un ejemplo de esto se muestra a continuación.

[11] Se dice que K es un campo si satisface las siguientes condiciones:

- Si x, y son elemento de K, entonces $x+y$ y $x*y$ también son elementos de K
- Si x∈K, entonces $-x$ también es un elemento de K. Además, si $x \neq 0$ entonces x^{-1} es un elemento de K
- Los elementos 0 y 1 también son elementos de K

$$\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & -1 & 5 \\ 2 & 1 & -1 \\ 3 & 1 & -1 \end{pmatrix}$$

Es de suma importancia no confundir los valores dentro de la matriz con los valores de i y j. En estos dos ejemplos la primera matriz tiene dimensiones $i = 2 = j$ mientras que en la segunda matriz tiene las dimensiones $i = 3 = j$.

Otra matriz sumamente importante y que se usa mucho en la práctica, es la denominada **matriz transpuesta**. Sea $A = (a_{ij})$ una matriz de m x n. Se define la matriz transpuesta como la de n x m $B = (b_{ji})$ tal que $b_{ji} = a_{ij}$. A la matriz B se le conoce como transpuesta de A.

$$\begin{pmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{m1} \\ a_{12} & a_{22} & a_{32} & \cdots & a_{m2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{mn} \end{pmatrix}.$$

Para tener un poco más de claridad con este tipo de matriz véase el siguiente ejemplo:

Sea $A = \begin{pmatrix} 2 & 7 & 1 \\ 3 & 4 & 9 \end{pmatrix}$ entonces su matriz transpuesta es ${}^{t}A = \begin{pmatrix} 2 & 3 \\ 7 & 4 \\ 1 & 9 \end{pmatrix}$

Una matriz que se deriva directamente de la matriz transpuesta es la denominada **matriz simétrica**. Una matriz simétrica es una matriz que es igual a su transpuesta y esta matriz es necesariamente cuadrada. Véase el siguiente ejemplo:

$$\begin{pmatrix} 1 & -1 & 2 \\ -1 & 0 & 3 \\ 2 & 3 & 7 \end{pmatrix}$$

La transpuesta de la matriz anterior es exactamente la misma.

La **matriz nula** es una matriz en donde $a_{ij} = 0$ para toda i, j. Su forma es la siguiente:

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Sea $A = (a_{ij})$ una matriz cuadrada, se denomina al vector $a_{11}, \ldots, a_{nn}$ como componentes de la diagonal. Se dice que una matriz es diagonal cuando todos sus componentes son cero excepto aquellos que pertenecen a la diagonal, es decir $a_{ij} = 0$ si $i \neq j$. Toda matriz diagonal es simétrica, su forma algebraica es la siguiente:

$$\begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_n \end{pmatrix}.$$

Por último, se define la **matriz identidad** de n x n como una matriz que contiene en todos sus componentes 0 excepto en la diagonal principal, la cual tendrá únicamente el valor de 1. Se denota esta matriz como I_n o I si no hay necesidad de especificar la n. Matricialmente se ve de la siguiente forma:

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Operaciones Básicas de las Matrices

Dos operaciones muy comunes en las matrices son la suma y resta de matrices. Sin embargo, cabe destacar que únicamente se pueden hacer las dos operaciones mencionadas siempre y cuando las matrices involucradas en la operación tengan el mismo número de filas y columnas (mismo orden)

Sea $A = (a_{ij})$ y $B = (b_{ij})$ dos matrices m x n. Se define $A + B$ la matriz en donde el i-ésimo renglón y la j-esima columna es $a_{ij} + b_{ij}$, en palabras coloquiales se sumará o restará entrada por entrada. Véase el siguiente ejemplo:

Se tiene la matriz A y la matriz B, calcular la matriz A+B:

$$A = \begin{pmatrix} 2 & 7 & 1 \\ 3 & 4 & 9 \end{pmatrix} \qquad B = \begin{pmatrix} 5 & 4 & 3 \\ 2 & 9 & 1 \end{pmatrix}$$

$$\rightarrow A + B = \begin{pmatrix} 2+5 & 7+4 & 1+3 \\ 3+2 & 4+9 & 9+1 \end{pmatrix}$$

$$\rightarrow A + B = \begin{pmatrix} 7 & 11 & 4 \\ 5 & 13 & 10 \end{pmatrix}$$

Existen algunas propiedades específicas de las operaciones suma y resta, las cuales se presentan a continuación:

Sean *A, B, C* matrices *n* x *m* en *K,* entonces se cumplen las siguientes propiedades:

- Asociatividad:*(A+B) +C=A+(B+C)*
- Conmutatividad:*(A+B) =(B+A)*
- Neutro aditivo: *A+0=0+A=A*
- Inverso aditivo: Sea *D* una matriz *n* x *m* que pertenece a *K* tal que *A+D=0*, a esta matriz se le denota por -*A*

Otra operación sumamente común es la operación de la multiplicación, sin embargo, es necesario puntualizar que la operación es diferente cuando se utiliza un escalar a cuando se utiliza otra matriz. Cuando se utiliza un escalar por una matriz se multiplicará cada entrada de la matriz por el escalar. Véase el siguiente ejemplo:

$$A = \begin{pmatrix} 2 & 7 & 1 \\ 3 & 4 & 9 \end{pmatrix} \quad \text{y} \quad \beta = 3$$

$$\rightarrow A^*\beta = \begin{pmatrix} 3x2 & 3x7 & 3x1 \\ 3x3 & 3x4 & 3x9 \end{pmatrix}$$

$$\rightarrow A^*\beta = \begin{pmatrix} 6 & 21 & 3 \\ 9 & 12 & 27 \end{pmatrix}$$

Así como en la suma el producto de una matriz por un escalar también tiene ciertas propiedades, las cuales son las siguientes:

Sean *A, B* matrices *n* x *m* en *K* y sean los escalares *α, β* en *K* entonces se cumplen las siguientes propiedades:

- Asociatividad: *(αβ)A= α(βA)*
- Distibutividad respecto a la suma de matrices: *α(A+B) = αA+ αB*
- Producto por el neutro multiplicativo: 1*A=A*

Para el segundo caso de multiplicación se define *A* = (a_{ij}) con *i* = 1, …, *m* y *j* = 1, …, *n* una matriz *m* x *n* y sea *B* = (b_{jk}) con *j* = 1, …, *n* y *k* = 1, …, *s* una matriz de *n* x *s* como se sigue:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \cdots & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & \cdots & b_{1s} \\ & \cdots & \\ b_{n1} & \cdots & b_{ns} \end{pmatrix}.$$

Se define la multiplicación o producto *AB* una matriz de *m* x *s* en donde cada elemento es de la forma:

$$\sum_{j=1}^{n} a_{ij} b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}$$

Si $A_1, \dots, A_m$ son los renglones de la matriz *A* que también puede ser conocido como un vector horizontal de 1 x *n*, y si $B^1, \dots, B^s$ son las columnas de la matriz B que al igual que en el caso anterior puede ser conocido como un vector, sin embargo el vector es de la forma *n* **x 1**, entonces las *ik*-ésimas entradas del producto *AB* es igual a $A_i B^k$. Dicho producto es de la siguiente forma:

$$AB = \begin{pmatrix} A_1 \cdot B^1 & \cdots & A_1 \cdot B^s \\ \vdots & & \vdots \\ A_m \cdot B^1 & \cdots & A_m \cdot B^s \end{pmatrix}$$

Un ejemplo de lo anteriormente mencionado se muestra a continuación:

Sean $A = \begin{pmatrix} 2 & 7 & 1 \\ 3 & 4 & 9 \end{pmatrix}$ y $B = \begin{pmatrix} 4 & 9 \\ 1 & 3 \\ 7 & 1 \end{pmatrix}$ dos matrices

Se observa que la matriz *A* es una matriz de 2x3 y la matriz *B* es una matriz de 3x2 por lo que la matriz generada por el producto *AB* será una matriz de 2x2.

$$AB = \begin{pmatrix} 2x4 + 7x1 + 1x7 & 2x9 + 7x3 + 1x1 \\ 3x4 + 4x1 + 9x7 & 3x9 + 4x3 + 9x1 \end{pmatrix}$$

$$\rightarrow AB = \begin{pmatrix} 8 + 7 + 7 & 18 + 21 + 1 \\ 12 + 4 + 63 & 27 + 12 + 9 \end{pmatrix}$$

$$\rightarrow AB = \begin{pmatrix} 22 & 40 \\ 79 & 48 \end{pmatrix}$$

Vectores y Valores Propios o Característicos

Sea *V* un espacio vectorial y sea:

$$A: V \rightarrow V$$

Un mapeo lineal de *V* en sí mismo. Un elemento $v \varepsilon V$ es llamado un **vector propio** de *A* si existe un número λ tal que $Av = \lambda v$. Si $v \neq 0$ entonces λ es determinado, de forma única porque $\lambda_1 v = \lambda_2 v$ lo cual implica que $\lambda_1 = \lambda_2$. En este caso se dice que λ es un **valor propio** de *A* que corresponde al vector propio *v*. También se dice que *v* es un vector propio con el valor propio λ. Otra forma de llamar al vector propio y valor propio es **vector característico** y **valor característico** respectivamente.

Si *A* es una matriz cuadrada *n* x *n* entonces el vector propio de *A* es por definición un vector propio del mapeo lineal de K^n en sí mismo representado por una matriz. Este vector propio *X* en *A* es un vector de K^n en donde existe $\lambda \varepsilon K$ tal que $AX = \lambda X$

Con la finalidad de encontrar el valor de λ, se analiza la expresión $AX = \lambda X$ de la cual la expresión λX se escribe como $(\lambda I)X$, por lo tanto:

$$\lambda X = AX$$

$$\rightarrow \lambda X - AX = 0$$

$$\to (\lambda I)X - AX = 0$$

$$\to (\lambda I - A)X = 0$$

Esta última ecuación representa un sistema homogéneo de *n* ecuaciones, como X es diferente de cero entonces $\lambda I - A$ es singular y su determinante es cero:

$$\det(\lambda I - A) = 0$$

Esta ecuación anterior recibe el nombre de **ecuación o polinomio característico** de *A*. Se presentan a continuación, una serie de ejemplos para calcular los vectores propios de una matriz.

- Sea:

$$A = \begin{pmatrix} 4 & 1 \\ -6 & -3 \end{pmatrix}$$

  Determinar sus vectores y valores propios.

Como primer paso, se nota que es una matriz cuadrada de 2x2 por lo que la matriz identidad a utilizar es una matriz identidad de 2x2.

$$\lambda I = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$$

La diferencia se presenta de la siguiente forma:

$$(\lambda I - A) = \begin{pmatrix} \lambda - 4 & -1 \\ 6 & \lambda + 3 \end{pmatrix}$$

Por lo que la ecuación del sistema homogéneo se ve de la siguiente forma:

$$\begin{pmatrix} \lambda - 4 & -1 \\ 6 & \lambda + 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

El siguiente paso es realizar la ecuación o polinomio característico mediante la determinante:

$$\det(\lambda I - A) = 0$$

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 4 & -1 \\ 6 & \lambda + 3 \end{vmatrix}$$

$$= (\lambda - 4)(\lambda + 3) - ((-1)(6))$$

$$= \lambda^2 - 4\lambda - 3\lambda - 6$$

$$= \lambda^2 - \lambda - 6$$

$$= (\lambda - 3)(\lambda + 2)$$

Igualando la ecuación anterior a cero y encontrando las raíces correspondientes se tiene que:

$$\lambda_1 = 3$$

$$\lambda_2 = 2$$

Los valores anteriores son los denominados **valores propios**, por lo que lo único que falta por encontrar son los vectores propios y eso se realiza de la siguiente forma partiendo del mismo sistema homogéneo:

$$\begin{pmatrix} \lambda - 4 & -1 \\ 6 & \lambda + 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Sustituyendo el valor de $\lambda_1 = 3$ en el sistema de ecuaciones se tiene:

$$\begin{pmatrix} 3 - 4 & -1 \\ -6 & 3 + 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -1 & -1 \\ -6 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$-x_1 - x_2 = 0$$
$$-6x_1 + 6x_2 = 0$$

Resolviendo el sistema de ecuaciones anterior, la respuesta es $x_1 = -x_2$, es decir x = (1,-1) por ejemplo. De forma análoga se procede con el valor de $\lambda_2 = -2$ obteniendo $x = (-\frac{1}{6}, 1)$ también como ejemplo. Los vectores anteriores son lo que llaman vectores propios.

- Sea ahora la matriz:

$$A = \begin{pmatrix} 2 & -2 & 3 \\ 10 & -4 & 5 \\ 5 & -4 & 6 \end{pmatrix}$$

Determinar sus vectores y valores propios.

Al igual que el ejercicio anterior se nota que es una matriz cuadrada de 3 x 3, utilizando la matriz Identidad y multiplicarla por λ:

$$\lambda I = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

Posteriormente se realiza la diferencia de la siguiente forma:

$$(\lambda I - A) = \begin{pmatrix} \lambda - 2 & 2 & -3 \\ -10 & \lambda + 4 & -5 \\ -5 & 4 & \lambda - 6 \end{pmatrix}$$

Una vez hecho lo anterior, la ecuación homogénea queda de la siguiente forma:

$$\begin{pmatrix} \lambda - 2 & 2 & -3 \\ -10 & \lambda + 4 & -5 \\ -5 & 4 & \lambda - 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

El siguiente paso es realizar la ecuación o polinomio característico mediante el determinante:

$$det(\lambda I - A) = \begin{vmatrix} \lambda - 2 & 2 & -3 \\ -10 & \lambda + 4 & -5 \\ -5 & 4 & \lambda - 6 \end{vmatrix}$$

$$= (\lambda - 2)\begin{vmatrix} \lambda + 4 & -5 \\ 4 & \lambda - 6 \end{vmatrix} - 2\begin{vmatrix} -10 & -5 \\ -5 & \lambda - 6 \end{vmatrix} + (-3)\begin{vmatrix} -10 & \lambda + 4 \\ -5 & 4 \end{vmatrix}$$

$$= (\lambda - 2)((\lambda + 4)(\lambda - 6) - (-5)(4)) - 2(-10(\lambda - 6) - (-5(-5))) + (-3)(-10(4) - (\lambda + 4)(-5))$$

$$= (\lambda - 2)(\lambda^2 - 6\lambda + 4\lambda - 24 + 20) - 2(-10\lambda + 60 - 25) + (-3)(-40 - (-5\lambda - 20)$$

$$= (\lambda - 2)(\lambda^2 - 2\lambda - 4) - 2(-10\lambda + 35) + (-3)(-40 + 5\lambda + 20)$$

$$= \lambda^3 - 2\lambda^2 - 4\lambda - 2\lambda^2 + 4\lambda + 8 + 20\lambda - 70 + 120 - 15\lambda - 60$$

$$= \lambda^3 - 4\lambda^2 + 5\lambda - 2 = 0$$

$$= (\lambda - 2)(\lambda - 1)^2$$

Por lo tanto, las raíces son las siguientes:

$$\lambda_1 = \lambda_2 = 1 \quad y \quad \lambda_3 = 2$$

Mismos que son los valores propios de la matriz A.

Para encontrar los vectores propios de la matriz A se utiliza $(\lambda I - A)X = 0$ donde X es el vector propio:

$$\begin{pmatrix} \lambda - 2 & 2 & -3 \\ -10 & \lambda + 4 & -5 \\ -5 & 4 & \lambda - 6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Se sustituyen los valores $\lambda_1 = \lambda_2 = 1$, y se resuelve el sistema homogéneo de ecuaciones lineales:

$$\begin{pmatrix} 1-2 & 2 & -3 \\ -10 & 1+4 & -5 \\ -5 & 4 & 1-6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 2 & -3 \\ -10 & 5 & -5 \\ -5 & 4 & -5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Para resolver este sistema, se utilizará el procedimiento de eliminación Gauss-Jordan:

- Se multiplica el primer renglón por -1, posteriormente a la segunda columna se le suma la primera columna multiplicando por -10 y por último a la tercera columna se le suma la primera columna multiplicándola por -5, quedando de la siguiente forma:

$$\begin{pmatrix} -1 & 2 & -3 \\ -10 & 5 & -5 \\ -5 & 4 & -5 \end{pmatrix} \xrightarrow[\substack{-1R1 \\ R2+(-10R1) \\ R3+(-5R1)}]{} \begin{pmatrix} 1 & -2 & 3 \\ 0 & -15 & 25 \\ 0 & -6 & 10 \end{pmatrix}$$

Ahora el pivote es -15, por lo que se procederá a realizar 0 arriba y abajo del pivote dividiendo en primera instancia el segundo renglón entre -15 simplificando el resultado. El siguiente paso es que al primer renglón se le resta el segundo renglón multiplicado por -2

$$\begin{pmatrix} 1 & -2 & 3 \\ 0 & 1 & -\frac{5}{3} \\ 0 & -6 & 10 \end{pmatrix} \xrightarrow[R1-2R2]{} \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & -\frac{5}{3} \\ 0 & -6 & 10 \end{pmatrix} \xrightarrow[R3+6R2]{} \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$x = -\frac{1}{3}z\,, \qquad y = -\frac{5}{3}z, \qquad z = z$$

Por lo tanto, un vector propio puede ser:

x = -1 y = -5 z = 3

Matriz de varianzas y covarianzas

La matriz de covarianza muestral $S = (S_{jk})$ es la matriz de varianzas y covarianzas muestrales de las variables p definida de la siguiente forma matricial:

$$\mathbf{S} = (s_{jk}) = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix}$$

En la matriz anterior, se tiene que las varianzas muestrales de p se encuentran en la diagonal de la matriz y todos los posibles pares de covarianzas muestrales se encuentran fuera de la diagonal.

La j-ésima variable, $S_{jj}=S^2_j$ es calculada de la siguiente manera (varianza muestral):

$$S_{jj} = S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_{ij} - \overline{y_j})^2$$

Donde $\overline{y_J}$ es la media de la j-ésima variable. De forma análoga la covarianza muestral de la j-ésima y k-ésima columna de Y es definida de la siguiente forma:

$$S_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(y_{ij} - \overline{y_j})(y_{ik} - \overline{y_k})$$

BIBLIOGRAFIA

1. Afifi, A., May S. & Clark V. (2012) Practical Multivariate analysis. CRC Press. USA.
2. Beltrán Aguilar, Pedro. Tasas de Caducidad: Guía de Apoyo para la Construcción y Aplicación Asociación Mexicana para las Instituciones de Seguros.
3. Beltrán Aguilar Pedro, (2008). Solvencia II: Los conceptos básicos. Asociación Mexicana de Instituciones de Seguros.
4. Carrasco Arjona Jagoba, (2016). Solvencia II: Análisis Teórico y Práctico. Universidad del País Vasco.
5. Cox D.R. and Oakes D, (1984). Analysis of Survival Data. Chapman and Hall: London, New York.
6. Cuadras, Carles M, (2007). Nuevos Métodos de Análisis Multivariante. España.
7. Herrera Contreras Fernando y Pérez Márquez Fernando, (2018). Modelo Mexicano de Supervisión Basado en Riesgos Tipo Solvencia II. Comisión Nacional de Seguros y Fianzas.
8. Lindsay I. Smith, (2002) A tutorial on Principal Components Analysis, Lindsay I Smith.
9. Mardia, Kantilal Varichand; J. T. Kent y J. M. Bibby (1995), Multivariate Analysis, USA, Academic Press
10. Serge Lang (2000). Linear Algebra. Departament of Mathematics Yale University
11. Tapia Mejía, Pedro (1993). Conservación de la Cartera de Seguro de Vida Individual en México. Comisión Nacional de Seguros y Fianzas