



Modelització Predictiva i Assegurances de Vida: un enfocament pràctic

Jornada

9 de maig de 2017

Barcelona, maig de 2017

MEMBRE DEL INTERNATIONAL ACTUARIAL ASSOCIATION



International Actuarial Association
Association Actuarielle Internationale

www.actuaris.org

MEMBRE DEL ACTUARIAL ASSOCIATION OF EUROPE



actuarial association of europe



RGA re International
Ibérica

Modelización Predictiva y Seguros de Vida: un enfoque práctico

Víctor Barriga – Chief Actuary

Fernando Ballesteros – Senior Pricing Actuary

Juan Carlos Ibarra – Director Business Development

Barcelona, 9 Mayo 2017

Nace el ciclismo de datos, la novedosa manera de correr y de verlo por televisión

En la Tirreno-Adriático se estrenó un sistema por el cuál el seguidor tiene acceso a una gran cantidad de datos de rendimiento de cada uno de los ciclistas en tiempo real

El desarrollo de gadgets para el ciclismo ha evolucionado en sintonía con el de otros aspectos de la vida. Del pulsómetro y cuentakilómetros de principio de los noventa al potenciómetro de hoy. Del contador de pedaladas, al seguimiento por GPS y todo eso, disponible en el ordenador de a bordo que los ciclistas llevan acoplado al manillar. Desde hace años en el ciclismo profesional -y amateur- los datos importan. Y mucho. Las pájaras en las etapas alpinas del Tour se cuentan con los dedos de una mano porque todo está medido al detalle. Chris Froome ha llevado al paroxismo eso de atacar mirando sus vatios y era cuestión de tiempo que esa información estuviera disponible para el público.

El sector Big Data será clave para el empleo a nivel mundial

Con un movimiento claro hacia la microsegmentación, el sector asegurador y financiero son capaces de ofrecer productos específicos que atiendan a las características propias del cliente gracias al análisis de datos. Los expertos aseguran que Big Data se convertirá en los próximos años en un sector clave para el empleo a nivel mundial. "Los profesionales de data science de Europa se están yendo a EE.UU", lamenta Queralt.

Esta tarta no se repartirá solo para los bancos y las aseguradoras, matizan ambos expertos, otros competidores como Google ya han cogido su plato. "Los jugadores globales quieren entrar en el juego", puntualiza Zafra. Otro ejemplo es la red social Facebook que cuenta desde diciembre con ficha bancaria en España.

Fuente: BDS 30 de Marzo



"Nuevas tecnologías y Big Data convertirán al actuario en un profesional clave en la transformación digital"

Tendencias en temas de búsqueda en Google

Fuente: Google Trends



RGA re International
Ibérica

Modelización Predictiva

¿Qué es Modelización Predictiva?

Definición

Modelización predictiva es un **proceso**, no un producto

Buenos Datos (**hechos**) es fundamental para crear un modelo significativo

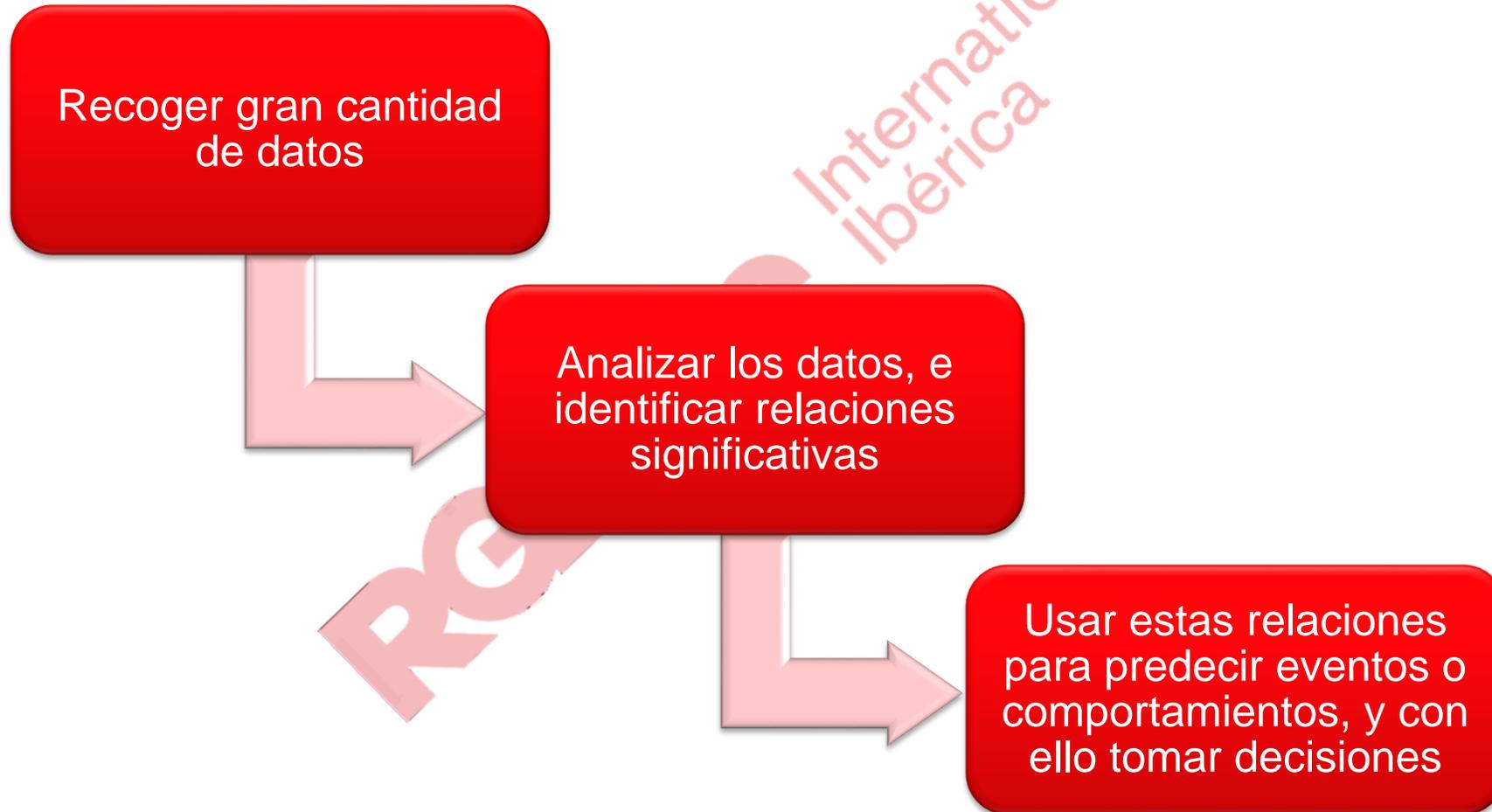
- Modelización Predictiva es el proceso por el cual se usan hechos históricos para crear predicciones sobre eventos o comportamientos futuros

Predicciones son creadas a través de Modelos estadísticos sofisticados

Los Modelos tratan de predecir la probabilidad de un **determinado resultado**

¿Qué es Modelización Predictiva?

En definitiva...



¿Por qué se está volviendo tan popular?

Causas

Innovación con Modelización Predictiva

Big Data

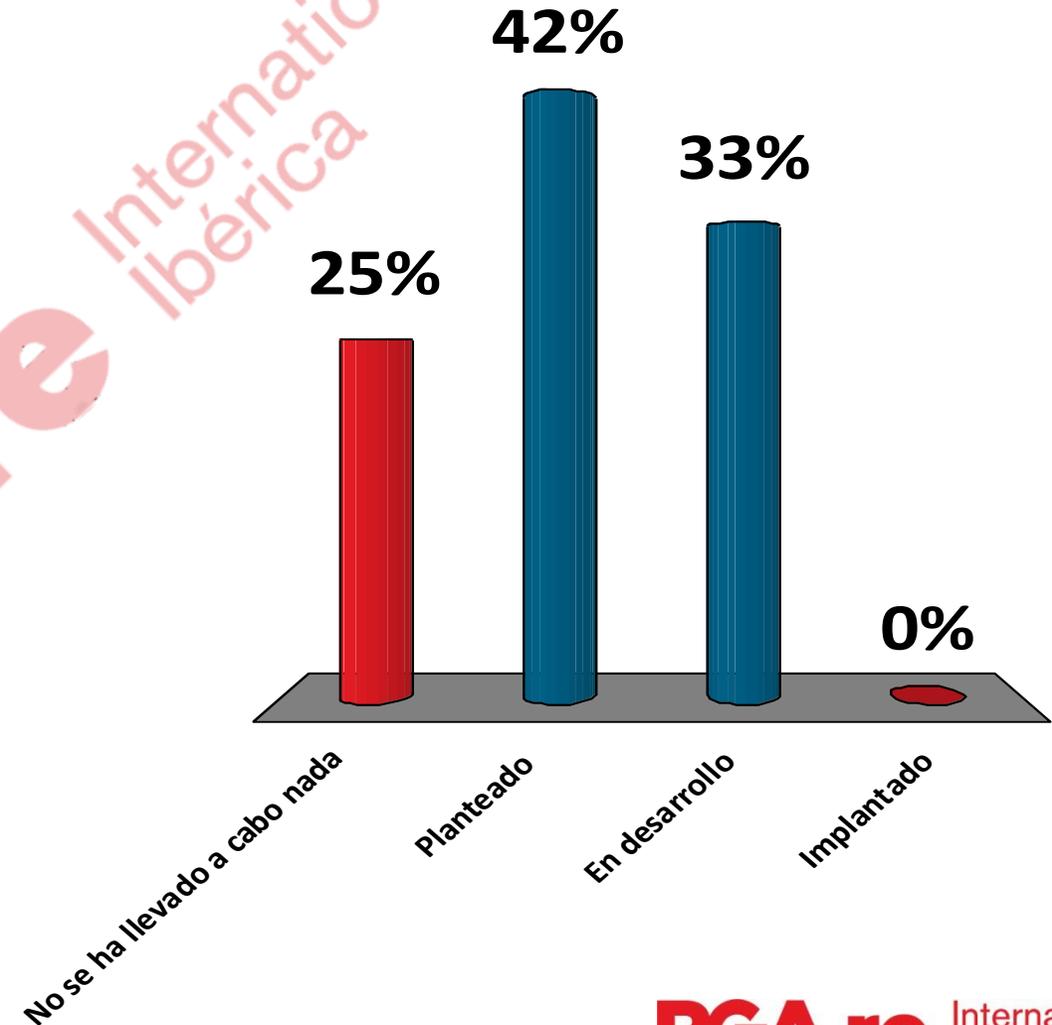
Tecnología

Ciencia de Datos

Entorno

¿Qué grado de desarrollo tiene el Big Data dentro de su compañía?

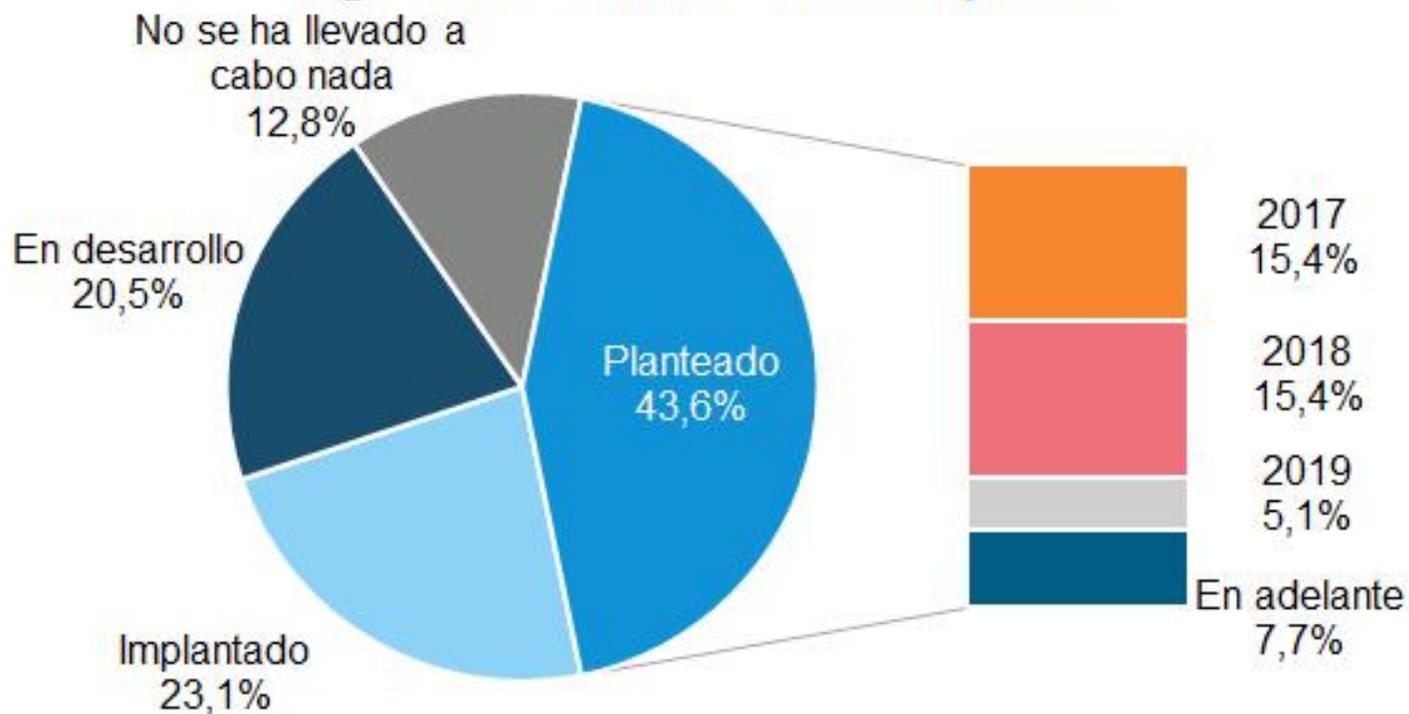
- A. No se ha llevado a cabo nada
- B. Planteado
- C. En desarrollo
- D. Implantado



Qué responde el Sector

¿Qué grado de desarrollo tiene el Big Data dentro de su compañía?

¿Qué grado de desarrollo tiene el Big Data dentro de su compañía?



Fuente: ICEA (Twitter: @icea_es)

El 44% de las aseguradoras españolas han comenzado a trabajar con Big Data

La implantación de la tecnología del Big Data en el sector asegurador español tiene todavía un largo camino por recorrer para alcanzar el desarrollo que existe en otros países en los que está proporcionando mucha utilidad. En 2017, casi el 44% de las entidades del sector asegurador español están trabajando o han comenzado a trabajar en proyectos de implantación de Big Data.

La falta de personal especializado pasa a ser el principal problema del sector para el desarrollo del Big Data añadido a problemas tecnológicos, en las entidades que ya están trabajando en dichos proyectos, y problemas con los datos, en las entidades que aún no están trabajando.

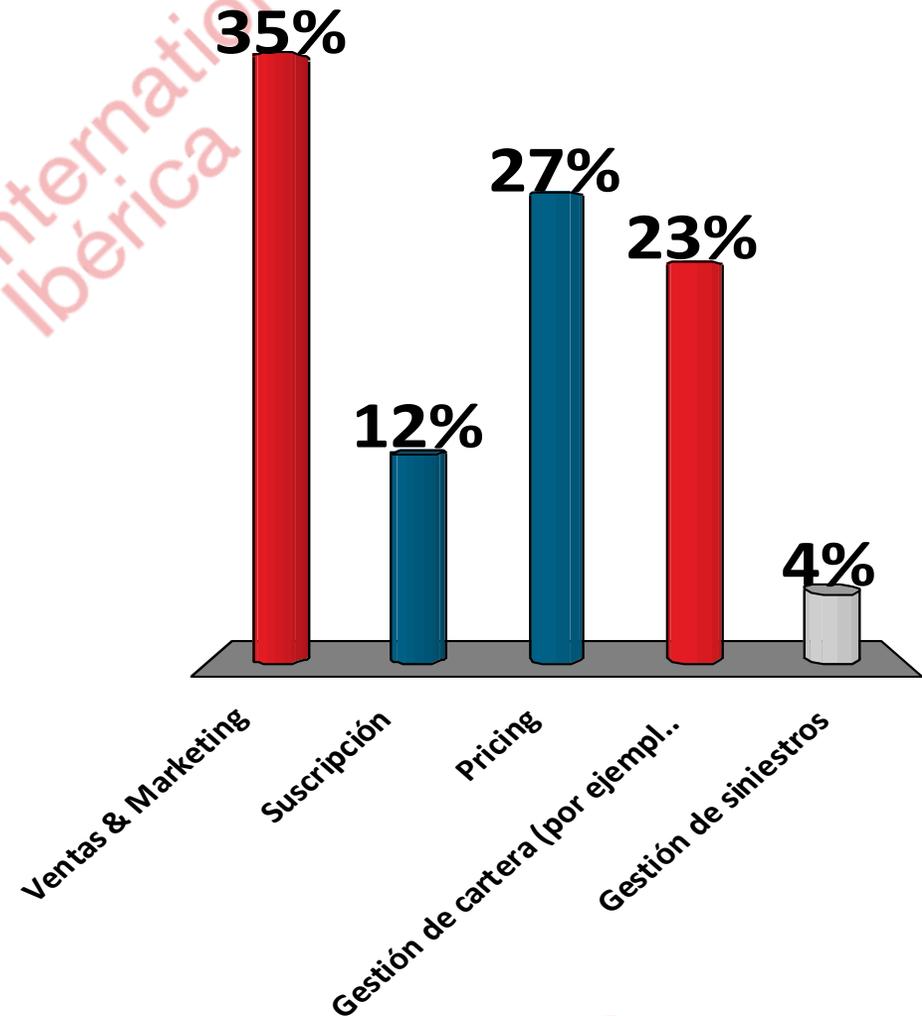
Proceso de Modelización Predictiva

Etapas en el proyecto



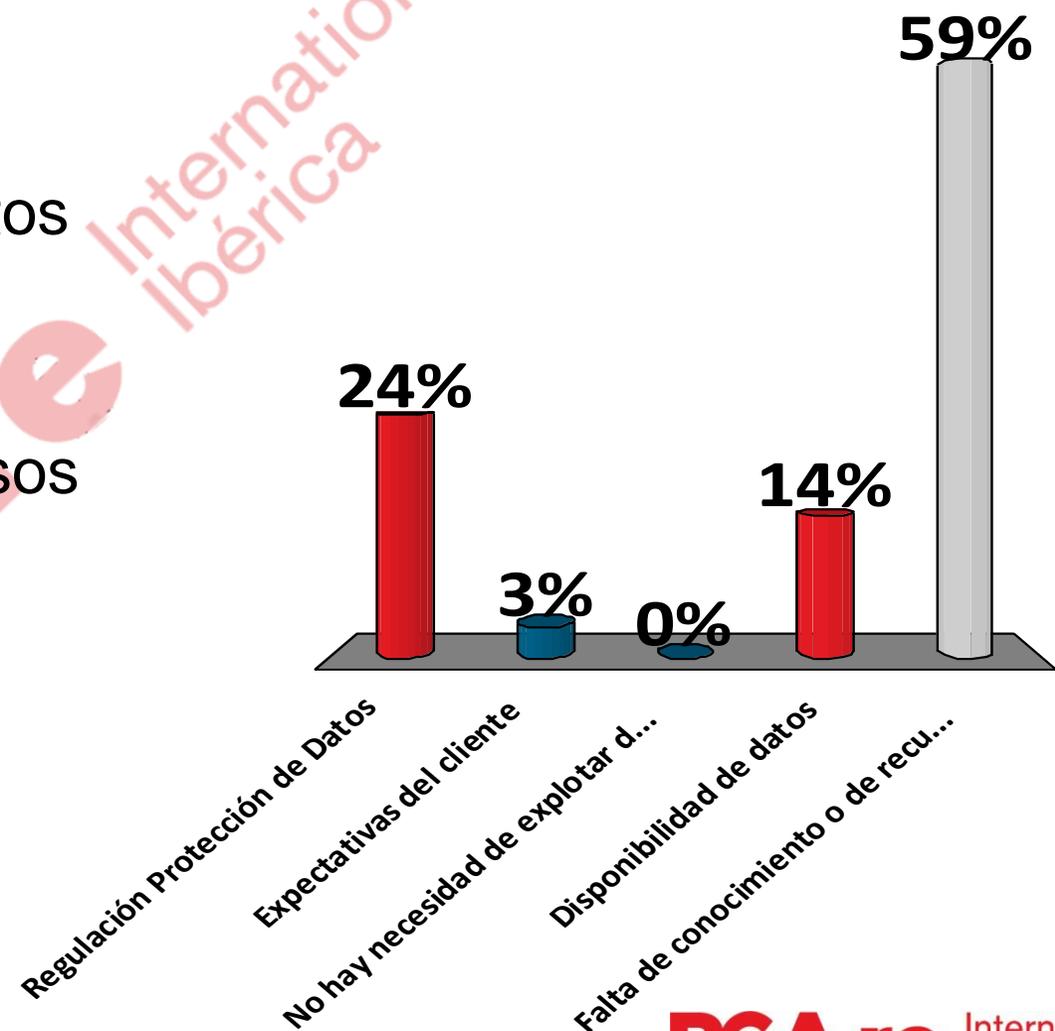
¿Dónde creéis que el uso de datos añade más valor?

- A. Ventas & Marketing
- B. Suscripción
- C. Pricing
- D. Gestión de cartera (por ejemplo: gestión de rescates, cross-selling...)
- E. Gestión de siniestros



¿Cuál es la principal barrera en el uso de datos?

- A. Regulación Protección de Datos
- B. Expectativas del cliente
- C. No hay necesidad de explotar datos para competir
- D. Disponibilidad de datos
- E. Falta de conocimiento o de recursos con habilidades para explotarlos



Retos en Regulación Protección Datos

¿Cuál es la principal barrera en el uso de Datos?

Regulación de las insurtech: equilibrio entre protección a los asegurados y apoyo a la innovación

IAIS analiza los impactos potenciales de las innovaciones financieras en el mercado asegurador y en su enfoque de la supervisión

“Las insurtech tendrán un impacto significativo en los modelos de negocio de las aseguradoras”, afirma la Asociación Internacional de Supervisores de Seguros (IAIS) en su reciente informe ‘FinTech Developments in the Insurance Industry’. “La regulación, junto con un firme nivel de supervisión, tendrá que evolucionar para garantizar el equilibrio adecuado entre un mantenimiento de la protección de los asegurados sin menoscabar indebidamente la innovación”, señala el informe.

El futuro de los Seguros de Vida

 UK:

- Desarrollo de Bases técnicas (mortalidad, invalidez y rescates)
- Modelo de tarificación utilizando Códigos Postales
- Análisis de experiencia
- Predicción de Fumador
- Suscripción predictiva con datos de agencia de crédito
- Suscripción predictiva con datos bancasegurador
- Evaluación calidad Agente
- Programa de retención de cartera

 Iberia:

- Suscripción predictiva con datos bancasegurador

 South África:

- Análisis de experiencia GLTD

Aplicaciones del análisis de Datos en el mundo



 India:

- Predicción Fraude Siniestros

 Australia:

- Suscripción predictiva / Cross sell con datos bancasegurador

 USA:

- Modelo de tarificación para GLTD
- Segmentación de riesgos para cross-sell
- Segmentación de clientes
- Segmentación de riesgos geográfica
- Estudio Rescates
- Suscripción predictiva / Cross-sell utilizando datos productos No-Vida
- Term Tail Lapses
- Evaluación calidad Agente
- Programa de retención de cartera

Asia:

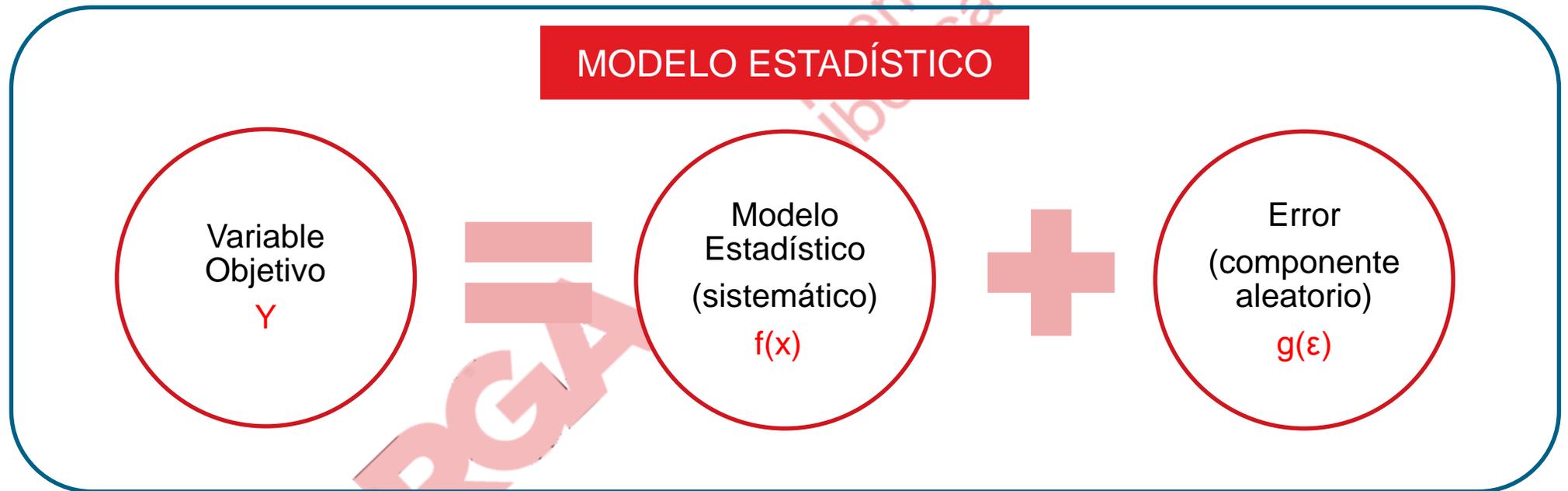
- Segmentación de riesgos para cross-sell & up-sell
- Suscripción predictiva con datos bancasegurador
- Mejoras en Guías de Suscripción
- Segmentación de precios
- Estudios de experiencia actuarial
- Modelos propensión de compra



RGA re International
Ibérica

Modelos Lineales Generalizados (GLM) en Seguros

Introducción – Modelo estadístico



Modelo Lineal

Modelo Lineal – Fórmula

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i$$
$$y_i, i = 1, \dots, n ; x_j, j = 1, \dots, p$$

Modelo Lineal – Supuestos

- a) Normalidad: $\varepsilon_i \sim N(0, \sigma^2)$; $y_i \sim N(0, \sigma^2)$
- b) Varianza constante: y_i es representativa de la población
- c) Observaciones independientes e idénticamente distribuidas
- d) Relación Lineal entre Variable Y y Variable(s) explicativas.

Modelo Lineal

Modelo Lineal – Componentes

1. La distribución de probabilidad para $Y_i \sim \text{Normal}$
2. El **regresor lineal**: $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$
3. La **relación**: $E(Y_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \rightarrow \text{Función Identidad (I)}$

Modelo Lineal – Ventajas

1. Procesos de variable respuesta son lineales o fácilmente transformables.
2. Estimación completa y efectiva de parámetros por Mínimos Cuadrados Ordinarios.
3. Resultados fáciles de entender y explicar. Permite establecer intervalos de confianza.

Modelo Lineal

GENIAL!!!! PERO...

Nos encontramos varias limitaciones cuando aplicamos un Modelo Lineal:

- a) Relaciones no lineales
- b) Outliers
- c) Distribuciones de datos siempre positivas

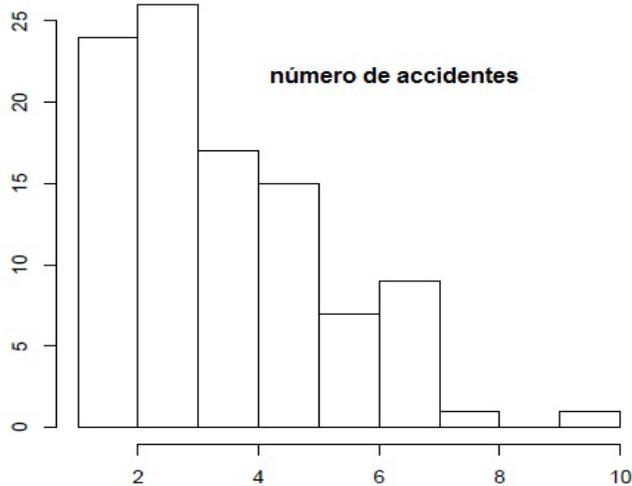
Y EN EL SECTOR SEGUROS??

- a) Procesos restringidos a binarios (mortalidad, caídas, Suscripción) o de conteo → No normales.
- b) Procesos donde la varianza de Y depende de la media (Poisson, Gamma...)
- c) No siempre podemos validar los supuestos del Modelo Lineal

Modelo Lineal – Limitaciones en aplicación a Sector Seguros

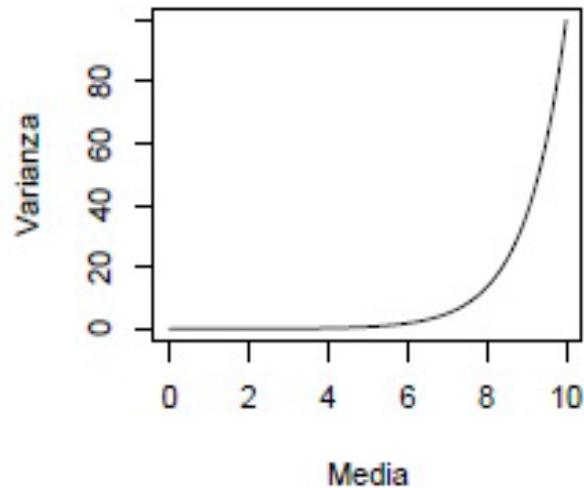
¿Normalidad?

- Número de siniestros (Poisson)
- Caídas/Propensión a compra/Fumador (Binomial)
- Supervivencia (Exponencial)
- [...]



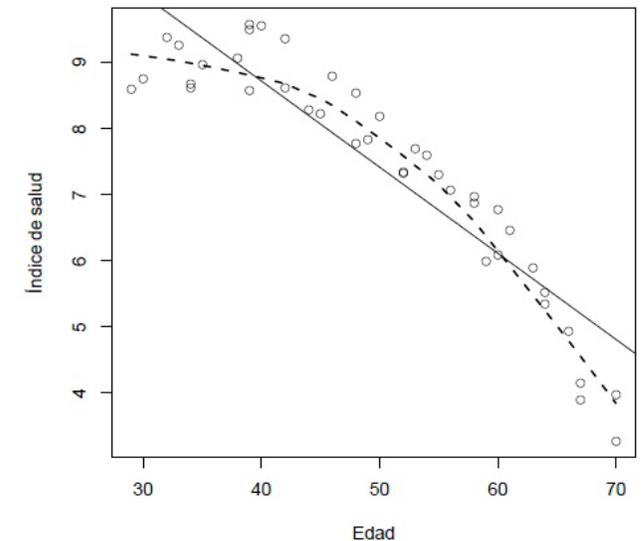
¿Varianza Constante?

- Poisson: Δ Media \rightarrow Δ varianza
- Gamma: Δ Media \rightarrow Δ varianza
- [...]



¿Linealidad?

- ¿Es lineal la relación entre Edad y Estado de salud?
- La diferencia de estado de salud es igual entre los 30 y los 40 que entre los 60 y 70?



Modelo Lineal Generalizado

Modelo Lineal Generalizado – Fórmula

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon = \sum_i \beta_i X_i + \varepsilon = \bar{X} \bar{\beta} + \varepsilon$$

- Extensión del Modelo Lineal Clásico
- Permite extender el marco del modelo lineal a variables que no sean normalmente distribuidas.
- Muy útil en Sector Seguros (muy utilizados en variables binarias o de conteo)
- Requiere juicio experto en la definición de sus componentes
- Mantiene el predictor (regresor) lineal (pero también podremos tener funciones no lineales de los predictores).
- Y, flexibiliza:
 - Yi puede seguir cualquier distribución de la familia exponencial
 - La relación entre Yi y el predictor no tiene que ser la Identidad

Modelo Lineal Generalizado

Modelo Lineal Generalizado – Componentes

1. **Componente aleatorio (Y)** → La variable respuesta ahora sigue una distribución de la familia exponencial de media μ y parámetro de dispersión θ (Normal, Log-Normal, Poisson, Gamma, Binomial...)

$$Y_i \sim \text{Exp}(\mu_i, \theta)$$

2. **Predictor Lineal (η)** → combinación lineal de predictores

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

3. **La Función de enlace, ligadura o Link $g(\cdot)$** →

$$E(Y) = \mu_i$$

$$g(\mu_i) = \eta_i ; \mu_i = g^{-1}(\eta_i)$$

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

Modelo Lineal Clásico:

$$g(\mu_i) = \mu_i$$

$$\mu_i = E(Y_i) = \eta_i$$

Modelo Lineal Generalizado - ¿Qué Componente Aleatorio (Y_i)?

	Poisson	Binomial	Normal	Gamma	Inverse Gaussian
Notación	$P(\mu)$	$B(m, \pi)/m$	$N(\mu, \sigma^2)$	$G(\mu, \nu)$	$IG(\mu, \sigma^2)$
Dominio	$(0, +\infty)$	$(0, 1)$	$(-\infty, +\infty)$	$(0, +\infty)$	$(0, +\infty)$
Tipo	Discreta	Discreta	Continua	Continua	Continua

- Cada Distribución es específica en términos de media y varianza
- La varianza es función de la media

Modelo Lineal Generalizado - ¿Qué función vínculo?

FUNCIONES VÍNCULO CANÓNICAS O ALTERNATIVAS		
DISTRIBUCIÓN	FUNCIÓN VÍNCULO CANÓNICA	POSIBLES FUNCIONES VÍNCULO ALTERNATIVAS
NORMAL	Identidad	Logarítmica
POISSON	Logarítmica	Identidad, Raíz cuadrada
BINOMIAL	Logit	Probit
GAMMA	Recíproca	Identidad, Logarítmica
NORMAL INVERSA	Inversa al cuadrado	

Modelo Lineal Generalizado – Elección de Predictores



Modelo 1 \rightarrow AIC 1
Modelo 2 \rightarrow AIC 2
Modelo 3 \rightarrow AIC 3

.....
.....

MODELO ÓPTIMO \rightarrow Menor AIC
(mejor ajuste con menor complejidad)

Aplicaciones prácticas en el Sector Asegurador

GLM: APLICACIONES SECTOR ASEGURADOR		
DISTRIBUCIÓN	FUNCIÓN VÍNCULO	APLICACIONES
NORMAL	Identidad	Aplicación General
POISSON	Logarítmica	Frecuencia siniestros
BINOMIAL	Logit	Retención, Cross-sell, UW
GAMMA	Logarítmica	Severidad / Siniestralidad
NORMAL INVERSA	Logarítmica	Coste Siniestro
COMPUESTA (*)	Logarítmica	Prima Pura (o Loss Ratio)

(*) Distribución Tweedie permite un modelo compuesto de Poisson y Gamma en el que tanto frecuencia como severidad se mueven en la misma dirección ante cambios en los predictores.

Construcción y Evaluación (Ejemplo)

**Ya tenemos todo para
empezar a construir y
evaluar nuestros modelo
GLM...**



RGA re International
Ibérica

Ejemplo de Modelización Predictiva utilizando R

Modelo Predictivo

Objetivo y Selección del Modelo

- El objetivo del modelo es construir un modelo predictivo que haga uso de variables socio-económicas para segmentar a los clientes por clases de riesgo permitiendo ahorro de costes y agilización del proceso de suscripción.
- Se trata por tanto de un problema de clasificación.
- El modelo logístico es una solución sencilla y eficiente para este tipo de situaciones.

Datos Disponibles

Vista General



Datos Disponibles

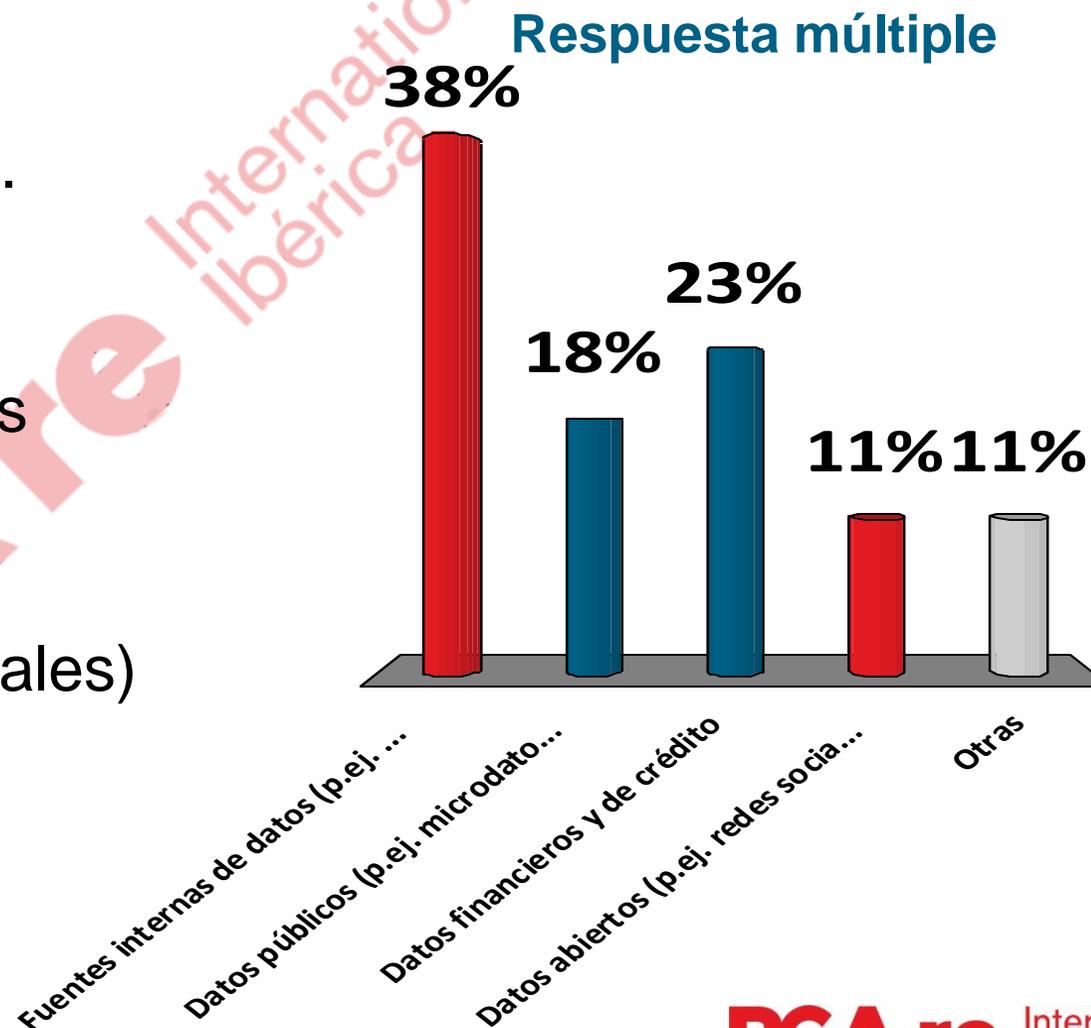
Vista General

Id	Región	Estrato	Género	Edad	Estado Civil	Relación Laboral	Industria	Nivel Educativo	Tipo Unidad Familiar	Superficie Vivienda	Ingresos Mensuales	Decisión Suscripción
1011-1	1	6	1	25	1	NA	NA	5	3	90	3	0
1031-2	1	6	1	46	2	NA	NA	5	7	120	4	1
1041-4	1	6	2	19	1	NA	NA	5	3	90	3	0
1051-1	1	6	1	22	1	3	832	5	2	90	2	0
1071-4	1	6	1	22	1	NA	NA	5	3	80	4	0

*La mayoría de las variables son categóricas.
La base de datos contiene ~15k registros.*

¿Cuáles de las siguientes fuentes de datos consideran que pueden ser potencialmente utilizadas en la modelización predictiva aplicada a los seguros de vida?

- A. Fuentes internas de datos (p.ej. siniestralidad, procesos de suscripción, etc...)
- B. Datos públicos (p.ej. microdatos publicados por INE)
- C. Datos financieros y de crédito
- D. Datos abiertos (p.ej. redes sociales)
- E. Otras



Modelo Predictivo

Pasos a Seguir

- 1 Importación y exploración visual de los datos
- 2 Construcción del modelo predictivo y selección de predictores
- 3 Interpretación y validación del modelo predictivo
- 4 Aplicación del modelo predictivo



Software

R (lenguaje de programación)

- R es un lenguaje de programación enfocado al análisis estadístico.
- Se trata de uno de los lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy popular en campos como la minería de datos o la investigación biomédica.
- A esto contribuye la posibilidad de cargar diferentes 'bibliotecas' o 'paquetes' con funcionalidades adicionales de cálculo y visualización.
- R es un software de uso libre que se distribuye a través de una licencia GNU GPL (Licencia Pública General GNU).

<https://www.r-project.org/>

[https://es.wikipedia.org/wiki/R_\(lenguaje_de_programaci%C3%B3n\)](https://es.wikipedia.org/wiki/R_(lenguaje_de_programaci%C3%B3n))

Logo R: De Hadley Wickham and others at RStudio - <https://www.r-project.org/logo/>, CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=35599651>

Modelo Predictivo

Importación de los Datos

```
data <- read.csv(file = "Predictive UW database.csv",  
                 header = TRUE)
```

```
summary(data)
```

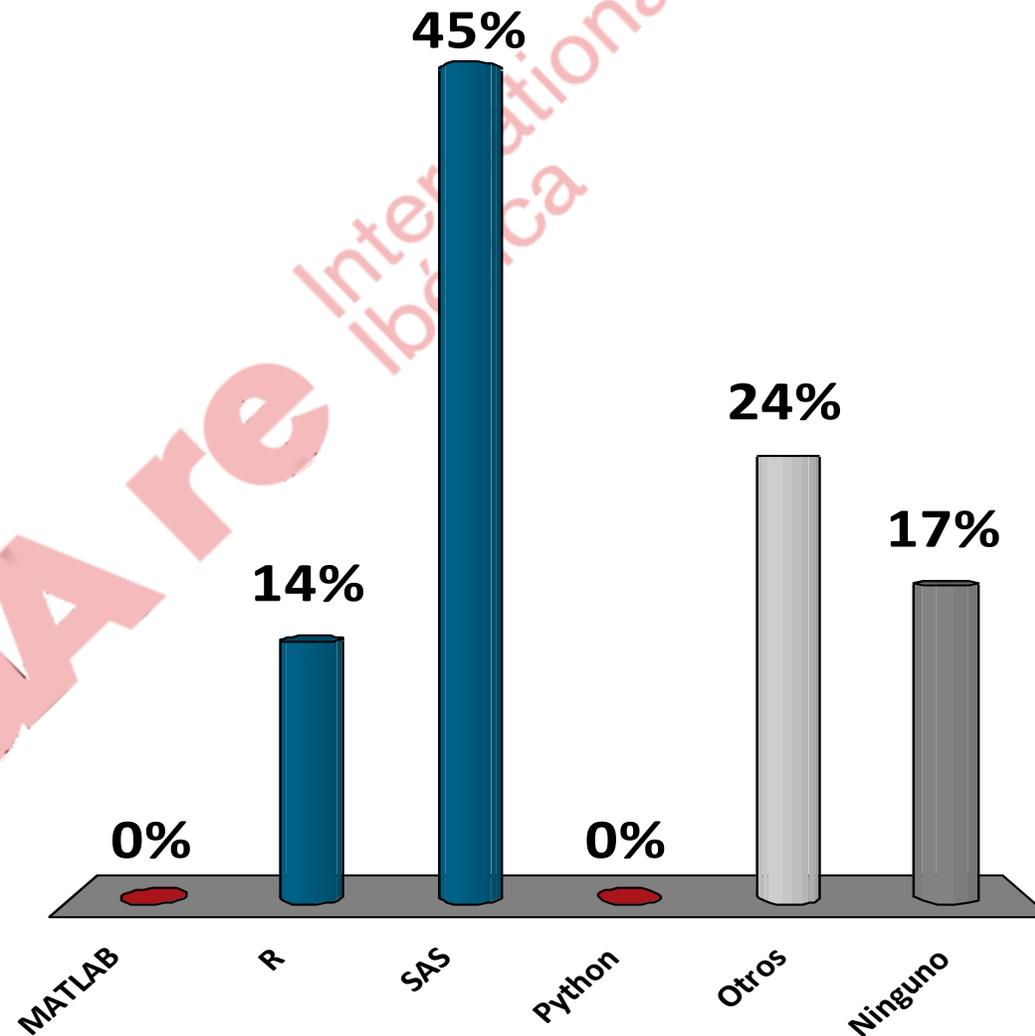
```
str(data)
```

```
head(data)
```

```
correlation_matrix <- cor(data[, -1])
```

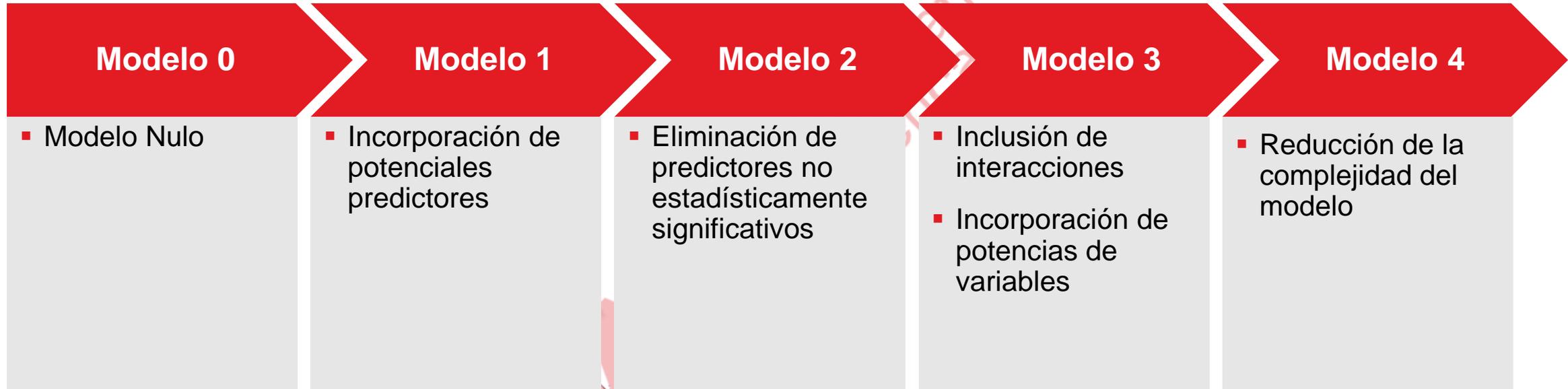
¿Qué software principalmente utiliza su compañía para hacer este tipo de análisis?

- A. MATLAB
- B. R
- C. SAS
- D. Python
- E. Otros
- F. Ninguno



Modelo Predictivo

Construcción del Modelo



Modelo Predictivo

Comparación de Modelos

Modelo	Grados de Libertad	AIC	Notas
Model0	1	14,476	Modelo nulo
Model1	35	13,071	Incorporación de potenciales predictores
Model2	15	13,067	Eliminación de predictores no estadísticamente significativos
Model3	21	13,000	Inclusión de interacciones y potencias de variables
Model4	17	13,016	Reducción de la complejidad del modelo

Modelo Predictivo

Modelo Final

```
Model4 <- glm(Y ~ X1 + X2 + X3..., family = binomial(), data =  
PredictiveUwData)
```

```
summary(Model4)
```

RGA re International
Ibérica

Modelo Predictivo

Proceso de Implementación del Modelo

Predicción

Validación

Aplicacion

Definicion Producto

Validación del modelo con datos originales

Modelo Predictivo

Validación del Modelo

```
PredictiveUwData$pred_score <- predict(Model14,  
  PredictiveUwData, type="response")
```

```
PredictiveUwData <- PredictiveUwData %>%  
  arrange(pred_score) %>%  
  mutate(grpid = factor(ntile(pred_score, 10),  
    labels = 1:10))
```

```
write.csv(PredictiveUwData, "PredictiveUwOutput.CSV")
```

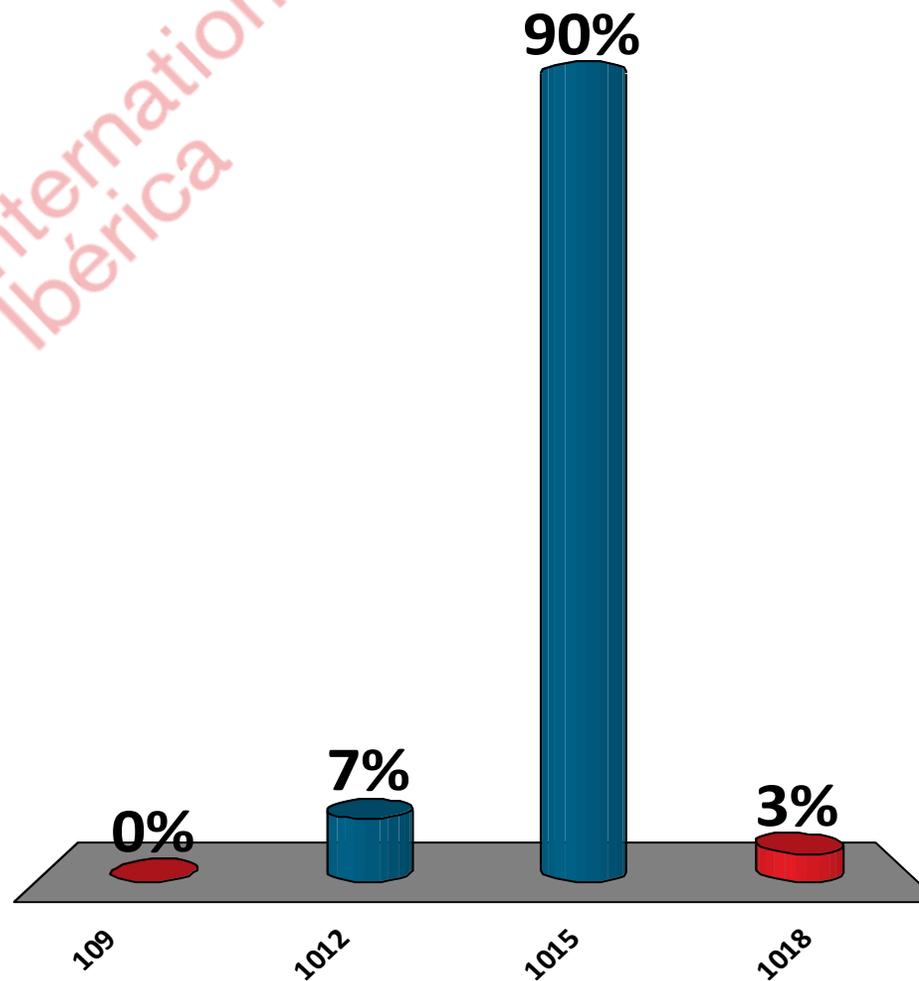


RGA re International
Ibérica

Aplicaciones prácticas al Seguro de Vida

¿Qué era un PETABYTE?

- A. 10^9
- B. 10^{12}
- C. 10^{15}
- D. 10^{18}

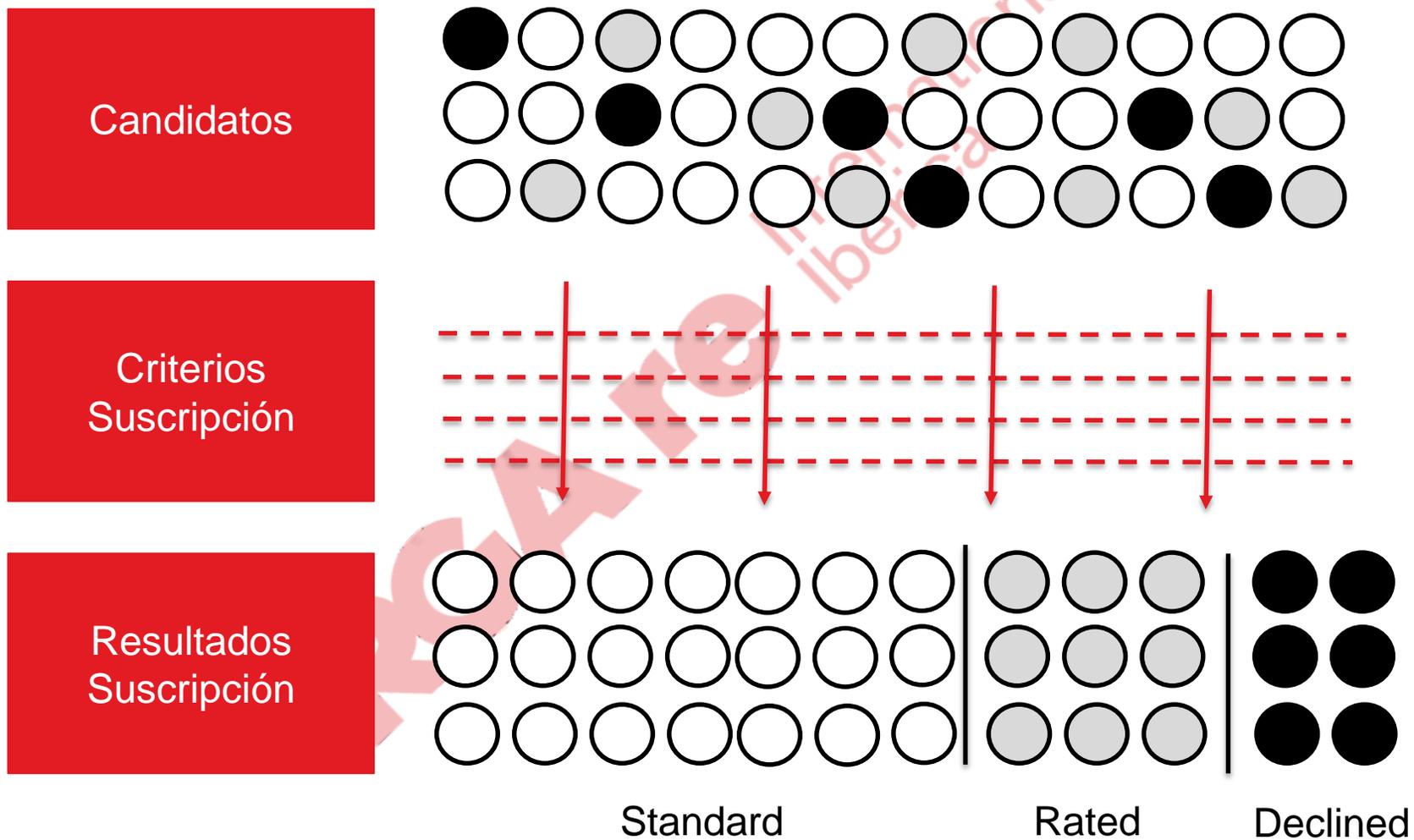


Aplicaciones prácticas en los Seguros Personales

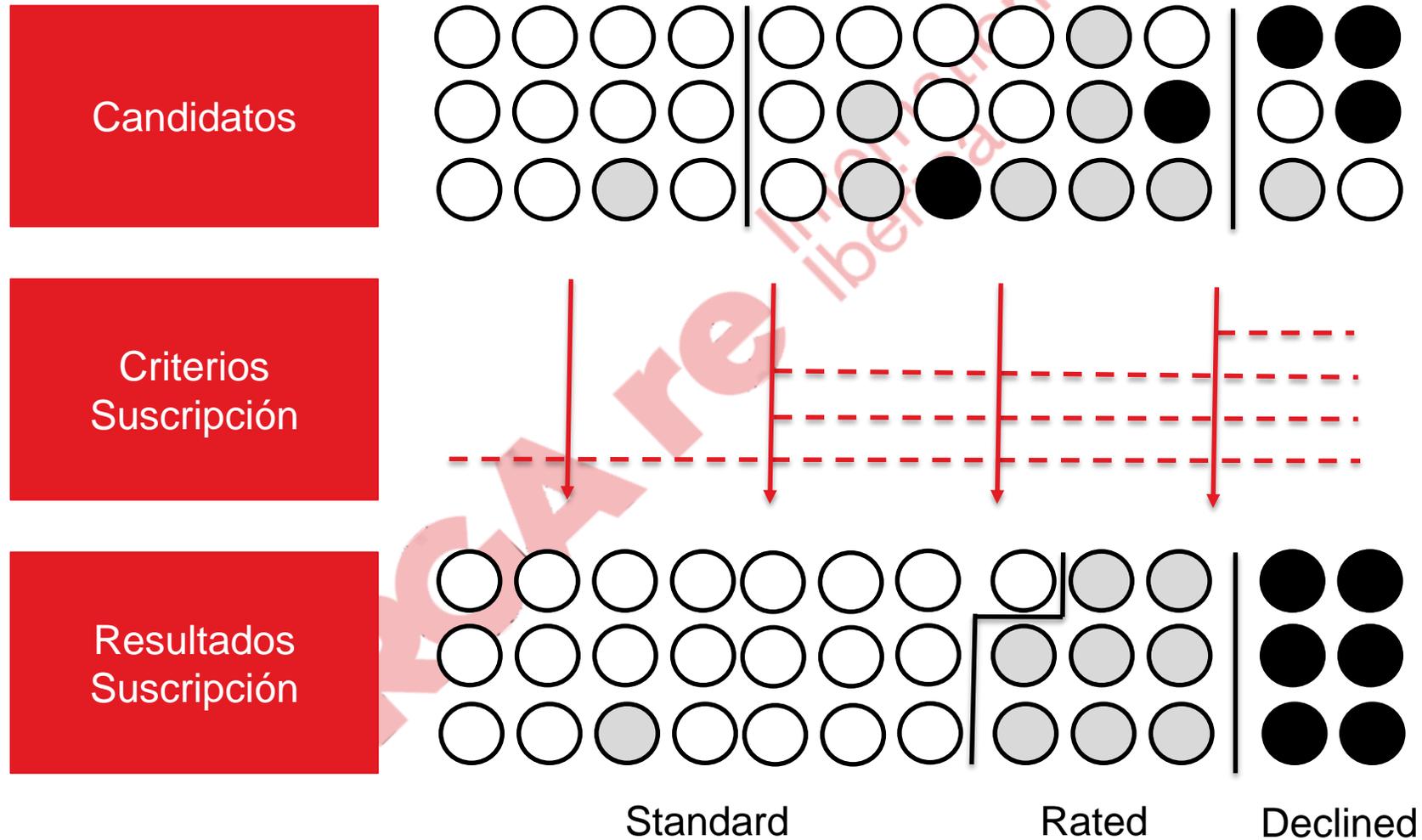
Algunos ejemplos más significativos

- Simplificación de los procesos de Suscripción
- Detección/Prevención del fraude
- Fidelización/Propensión a Rescatar
- Propensión a Comprar
- Detección Fumador

El proceso de Suscripción convencional



Predicción + Proceso simplificado de Suscripción

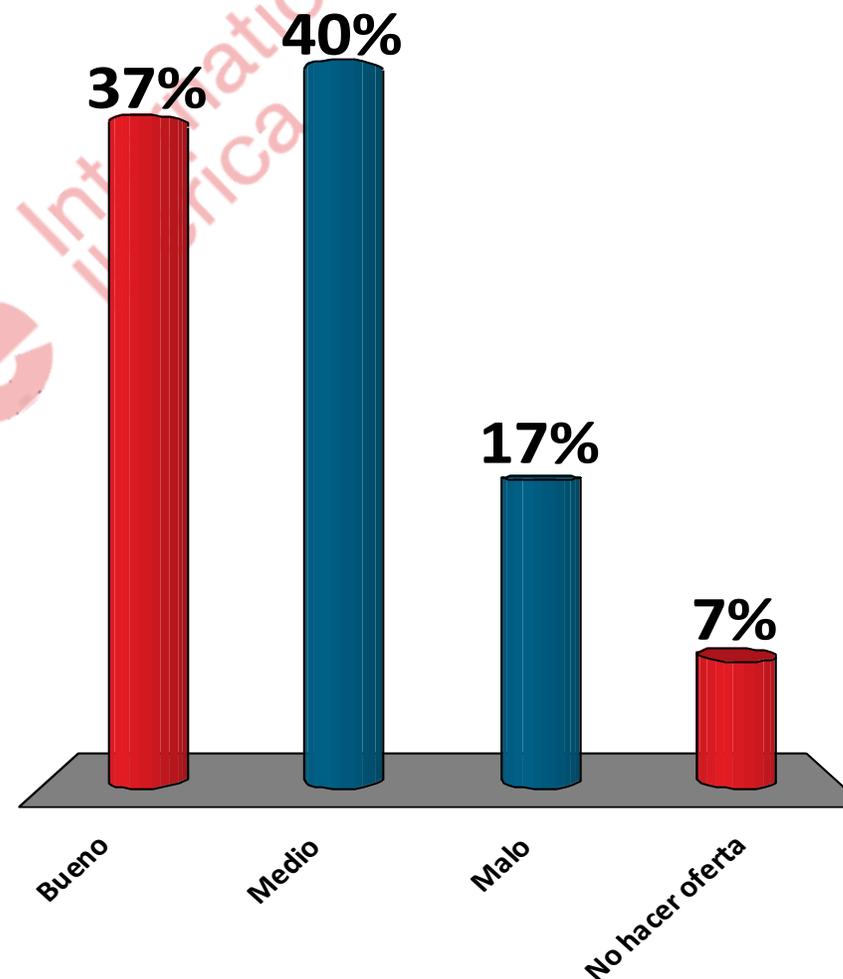


Ejemplo 1

1. **Edad:** 54
2. **Género:** Hombre
3. **Estado Civil:** Casado
4. **Ingresos:** 5.000 €/mes
5. **Ocupación:** Trabaja
6. **Educativo:** Estudios Universitarios
7. **Región:** Melilla

Ejemplo 1: ¿qué tipo de riesgo es?

- A. Bueno
- B. Medio
- C. Malo
- D. No hacer oferta



Ejemplo 1

1. **Edad:** 54
2. **Género:** Hombre
3. **Estado Civil:** Casado
4. **Ingresos:** 5.000 €/mes
5. **Ocupación:** Trabaja
6. **Educativo:** Estudios Universitarios
7. **Región:** Melilla

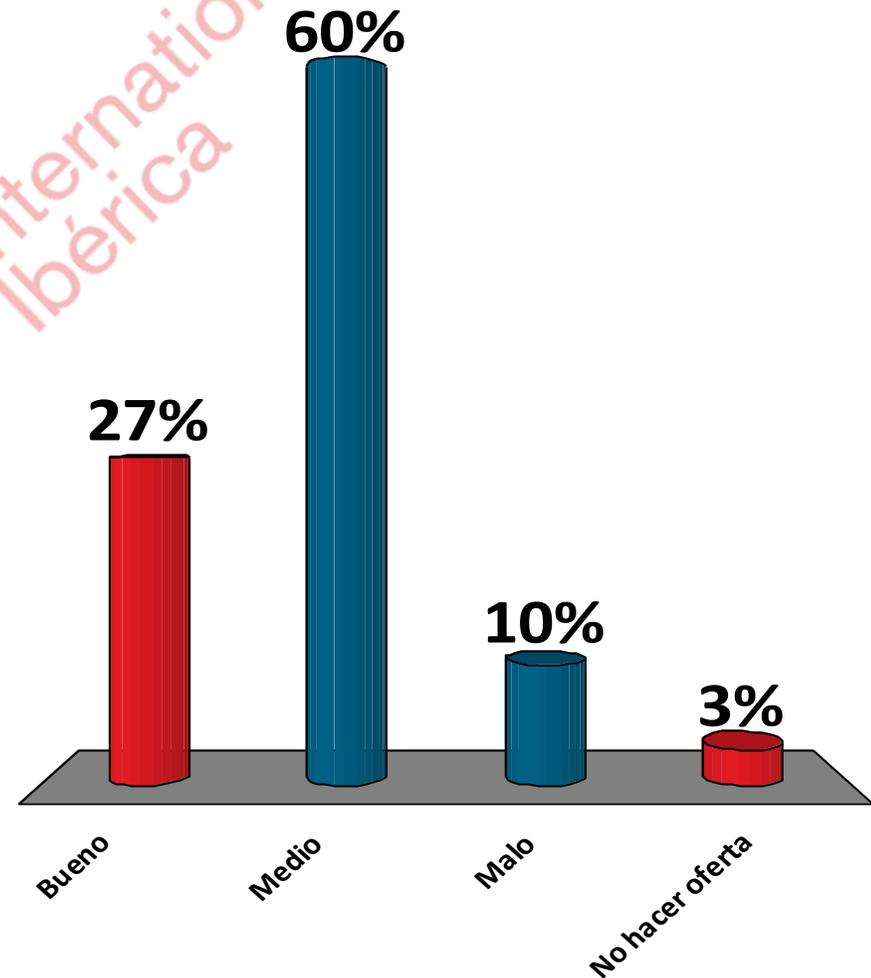
Scoring: 0,041
Tipo de Riesgo: BUENO

Ejemplo 2

1. **Edad:** 45
2. **Género:** Mujer
3. **Estado Civil:** Separada
4. **Ingresos:** 1.175 €/mes
5. **Ocupación:** Trabaja
6. **Educativo:** Bachillerato
7. **Región:** La Rioja

Ejemplo 2: ¿qué tipo de riesgo es?

- A. Bueno
- B. Medio
- C. Malo
- D. No hacer oferta



Ejemplo 2

1. **Edad:** 45
2. **Género:** Mujer
3. **Estado Civil:** Separada
4. **Ingresos:** 1.175 €/mes
5. **Ocupación:** Trabaja
6. **Educativo:** Bachillerato
7. **Región:** La Rioja

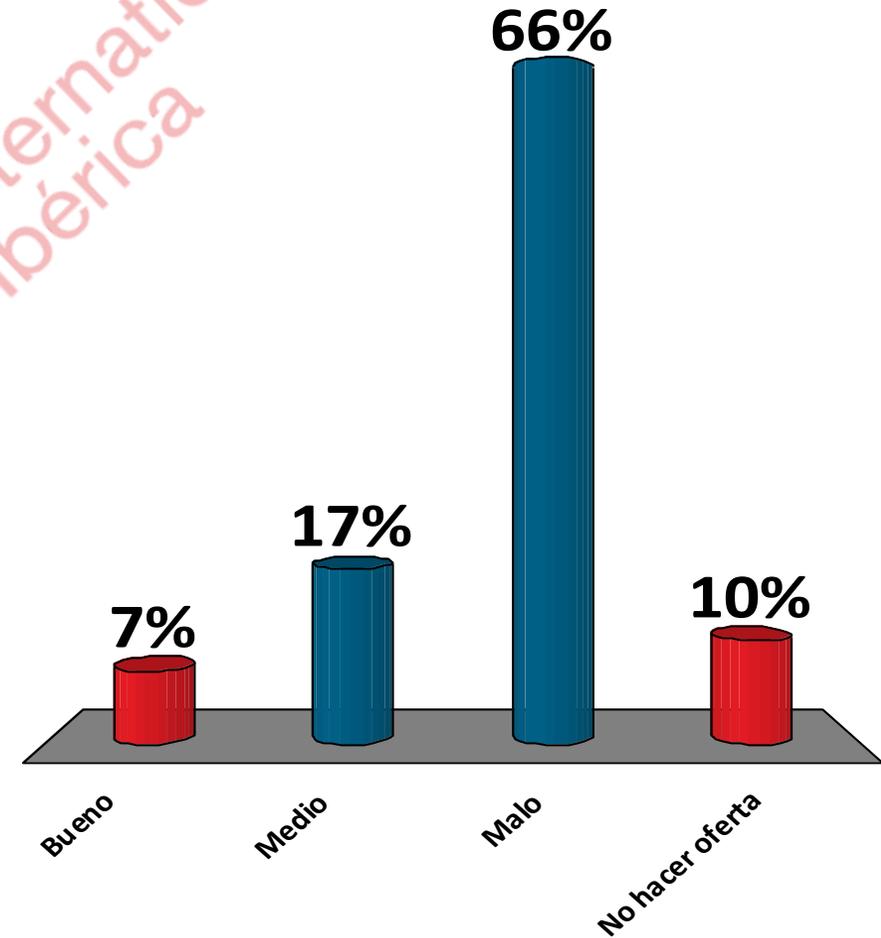
Scoring: 0,104
Tipo de Riesgo: MEDIO

Ejemplo 3

1. **Edad:** 49
2. **Género:** Mujer
3. **Estado Civil:** Soltera
4. **Ingresos:** 925 €/mes
5. **Ocupación:** Trabaja
6. **Educativo:** Ed. Primaria incompleta
7. **Región:** Galicia

Ejemplo 3: ¿qué tipo de riesgo es?

- A. Bueno
- B. Medio
- C. Malo
- D. No hacer oferta

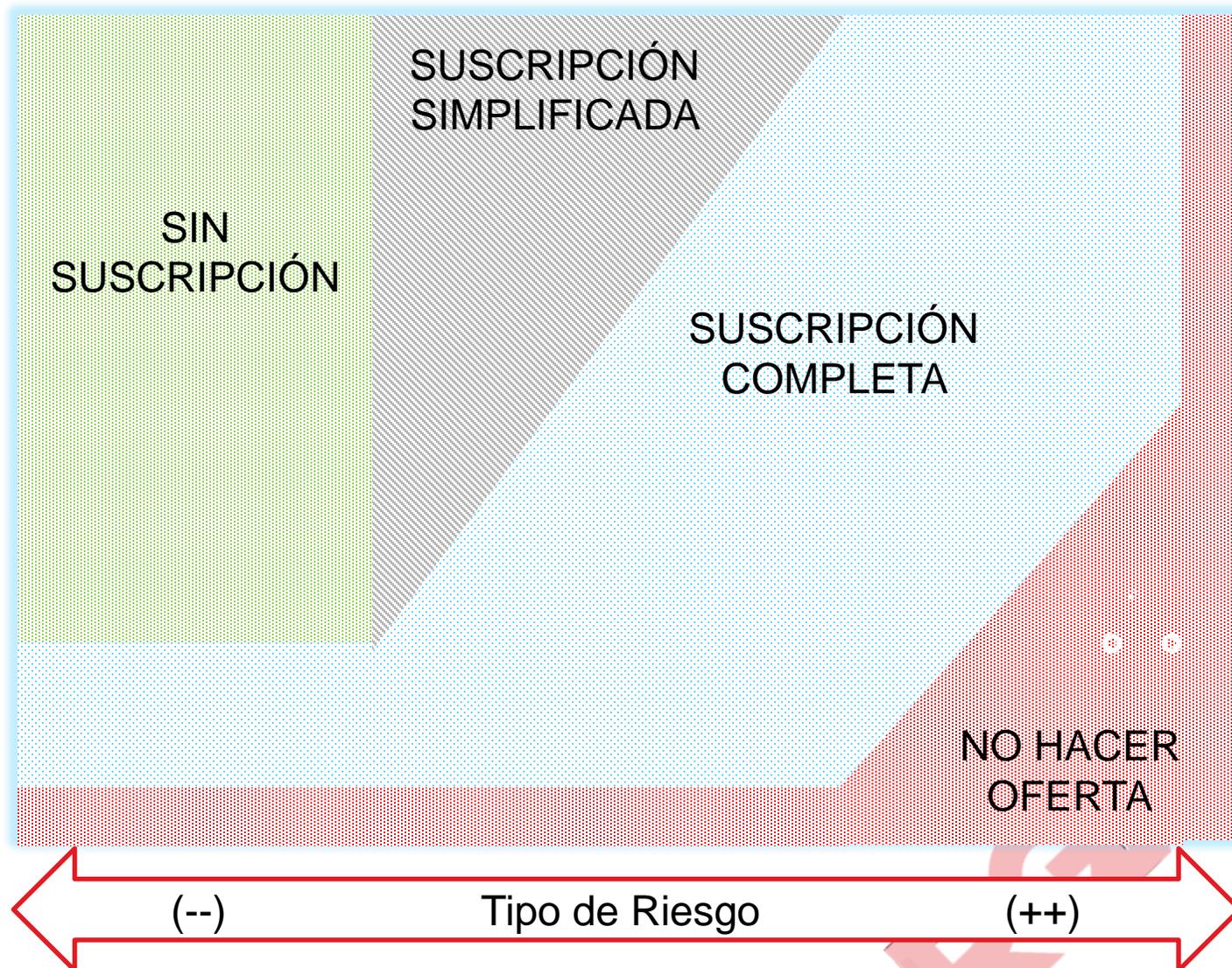


Ejemplo 3

1. **Edad:** 49
2. **Género:** Mujer
3. **Estado Civil:** Soltera
4. **Ingresos:** 925 €/mes
5. **Ocupación:** Trabaja
6. **Educativo:** Ed. Primaria incompleta
7. **Región:** Galicia

Scoring: 0,400
Tipo de Riesgo: MALO

Resultado de nuestro modelo



Detección / Prevención del Fraude



Modelo efectivo contra el fraude - Optimización de Recursos (Caso Real)

Objetivos

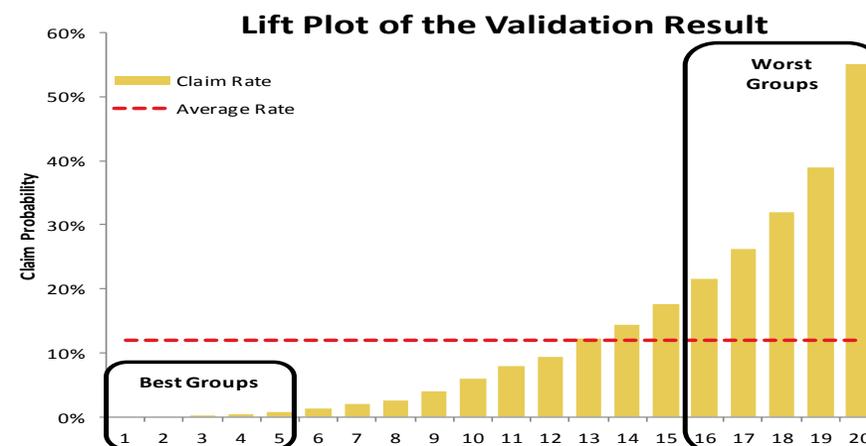
- Determinar la probabilidad de que un siniestro sea un caso de fraude y tomar las medidas apropiadas.
- Optimizar los recursos destinados a la investigación de siniestros

Proceso de Modelización

- Uso de experiencia de siniestros que se han investigado con resultados conocidos.
- Se combina información demográfica, información sobre pólizas y antecedentes de siniestros.

Resultados

- No es preciso investigar los primeros 25%.
- Especial énfasis en los 25% peores.
- Se consigue información sobre los factores generadores de los casos de fraude para incorporarlos en la definición de precios.



Conclusiones

1

El análisis predictivo es un proceso basado en datos históricos con una amplia gama de aplicaciones potenciales en el seguro.

2

La utilización de estas técnicas para modelos de suscripción simplificados en el ramo de Vida está siendo muy popular, pero hay muchas más aplicaciones.

3

Es esencial para obtener un buen resultado que los datos sean de buena calidad, pertinentes y con un volumen material.

4

Son soluciones personalizadas en base a datos específicos y necesidades específicas. La clave es la flexibilidad del enfoque.

5

El futuro nos deparará todavía más oportunidades. Sólo necesitamos una contestación positiva a estas dos preguntas: ¿Existen datos fiables? ¿Son datos predictivos?

RGA re International Ibérica

©2015, RGA International Reinsurance Company dac Sucursal en España. Todos los derechos reservados.

Ninguna parte de esta publicación puede ser reproducida en cualquier forma sin el previo permiso de RGA.

La información contenida en esta publicación es para el exclusivo uso interno del destinatario y no podrá ser divulgada por persona distinta del destinatario y asimismo, tampoco se podrá publicar o modificar la información sin el previo consentimiento expreso y por escrito de RGA.