

Del dato a la decisión

Daniel Arteta Salazar

2025-11-13

Fuente: https://rpubs.com/profe_ferro/1359822

Librerías

```
library(dplyr)
library(tibble)
library(ggplot2)
```

Parte 1 - Identificación de estructuras

Actividad 1: Detectives de estructuras

A continuación se crean objetos de distinto tipo. Debe:

1. Identificar la estructura de cada uno.
2. Proponer un uso actuarial posible.
3. Anotar operaciones útiles.

```
# vector
edades <- c(25, 31, 47)
# matriz
transicion <- matrix(c(0.98, 0.02, 0.10, 0.90),
                     nrow = 2, byrow = TRUE)
# data frame
colnames(transicion) <- c("Vida", "Siniestro")
rownames(transicion) <- c("Vida", "Siniestro")

polizas <- data.frame(ID = 1:3,
                      Edad = edades,
                      Riesgo = c("Bajo", "Medio", "Alto"))

# lista
resumen <- list(MediaEdad = mean(edades),
                Conteo = nrow(polizas))
```

1.1 Identificación de estructuras

```
# Estructura
str(edades)
```

```
##  num [1:3] 25 31 47
```

```
str(transicion)
```

```
## num [1:2, 1:2] 0.98 0.1 0.02 0.9
## - attr(*, "dimnames")=List of 2
## ..$ : chr [1:2] "Vida" "Siniestro"
## ..$ : chr [1:2] "Vida" "Siniestro"

str(polizas)

## 'data.frame': 3 obs. of 3 variables:
## $ ID : int 1 2 3
## $ Edad : num 25 31 47
## $ Riesgo: chr "Bajo" "Medio" "Alto"

str(resumen)

## List of 2
## $ MediaEdad: num 34.3
## $ Conteo : int 3

# Clase
class(edades)

## [1] "numeric"

class(transicion)

## [1] "matrix" "array"

class(polizas)

## [1] "data.frame"

class(resumen)

## [1] "list"
```

Objeto	Estructura	Posible uso actuarial	Operaciones útiles
edades	Vector numérico	Crear tramos de edad, calcular frecuencias de siniestros, identificar valores atípicos en montos de reclamaciones	<code>mean()</code> , <code>median()</code> , <code>summary()</code>
transicion	Matrix	Matrices de transición para seguros de salud, cálculos de primas colectivas	<code>matrix()</code> , <code>solve()</code> , <code>rowSums()</code>
polizas	Data Frame	Base de datos de pólizas, consolidación de siniestros por período, análisis de persistencia	<code>group_by()</code> , <code>summarise()</code> , <code>mutate()</code>
resumen	Lista	Almacenar múltiples modelos de tarificación por ramo, resultados de simulaciones, listas de parámetros por año	<code>lapply()</code> , <code>names()</code> , <code>unlist()</code>

Parte 2 - Transformación de estructuras

Actividad 2: Del caos al orden

Se entrega información en estructuras distintas. Organizar para que sea utilizable y:

1. Convertir una matriz en data.frame con nombres de columnas.
2. Agregar una columna con el número total de siniestros por cliente (a partir de una lista).
3. Convertir el resultado a tibble.
4. Ordenar por número de siniestros (descendente).
5. Crear un objeto lista llamado resultado que contenga al menos:
 - resultado\$base (la tabla final)
 - resultado\$metadatos (lo que consideres útil: n filas, n cols, etc.)

```
# Datos iniciales
mat <- matrix(c(1, 2, 3, 0.2, 0.5, 0.8),
              ncol = 2)
colnames(mat) <- c("ID", "Indicador")

# Lista de siniestros por cliente
siniestros <- list(
  c(0, 1, 0),
  c(1, 0, 0),
  c(0, 0, 0)
)
```

Desarrollo

```
# Punto 1
matriz <- as.data.frame(mat) # convertir en matriz

# Punto 2
matriz$Siniestros <- sapply(siniestros, sum) # crear columna 'Siniestros'

# Punto 3
matriz <- as_tibble(matriz) # convertir a tibble

# Punto 4
matriz %>%
  arrange(desc(Siniestros)) # ordenar por 'Siniestros'
```

```
## # A tibble: 3 x 3
##   ID Indicador Siniestros
##   <dbl>   <dbl>   <dbl>
## 1     1     0.2     1
## 2     2     0.5     1
## 3     3     0.8     0
```

```
# Punto 5
resultado <- list(
  base = matriz,
  metadatos = list(
    filas = nrow(matriz),
    columnas = ncol(matriz),
    nombres_columnas = colnames(matriz)
  )
) # crear lista 'resultado'

str(resultado)
```

```
## List of 2
## $ base      : tibble [3 x 3] (S3: tbl_df/tbl/data.frame)
## ..$ ID      : num [1:3] 1 2 3
## ..$ Indicador : num [1:3] 0.2 0.5 0.8
## ..$ Siniestros: num [1:3] 1 1 0
## $ metadatos:List of 3
## ..$ filas      : int 3
## ..$ columnas   : int 3
## ..$ nombres_columnas: chr [1:3] "ID" "Indicador" "Siniestros"
```

Parte 3 - Aplicación integradora

Actividad 3: Riesgo y edad

Una aseguradora desea explorar si la frecuencia de siniestros (conteo) varía con la edad.

1. Partiendo de la tabla final de la Parte 2, agregar una columna **Edad** con los valores simulados entre 20 y 60 (definir semilla para reproducibilidad).
2. Crear los grupos etarios <30, 31-45 y >45.
3. Calcular el promedio de siniestros por grupo.
4. Construir un gráfico de barras con el promedio por grupo (títulos y ejes claros).
5. Crear un objeto lista llamado **analisis_final** que contenga:
 - analisis_final\$datos (tu tabla con edades y grupos),
 - analisis_final\$resumen (promedios por grupo),
 - analisis_final\$grafico (el objeto ggplot),
 - analisis_final\$interpretacion (texto corto con 2-4 líneas).

```
set.seed(123)
# Punto 1
n <- nrow(matriz)
matriz$Edad <- sample(20:60, size = n, replace = TRUE)
print(matriz)
```

```
## # A tibble: 3 x 4
##   ID Indicador Siniestros Edad
##   <dbl>     <dbl>     <dbl> <int>
## 1     1       0.2         1    50
## 2     2       0.5         1    34
## 3     3       0.8         0    33
```

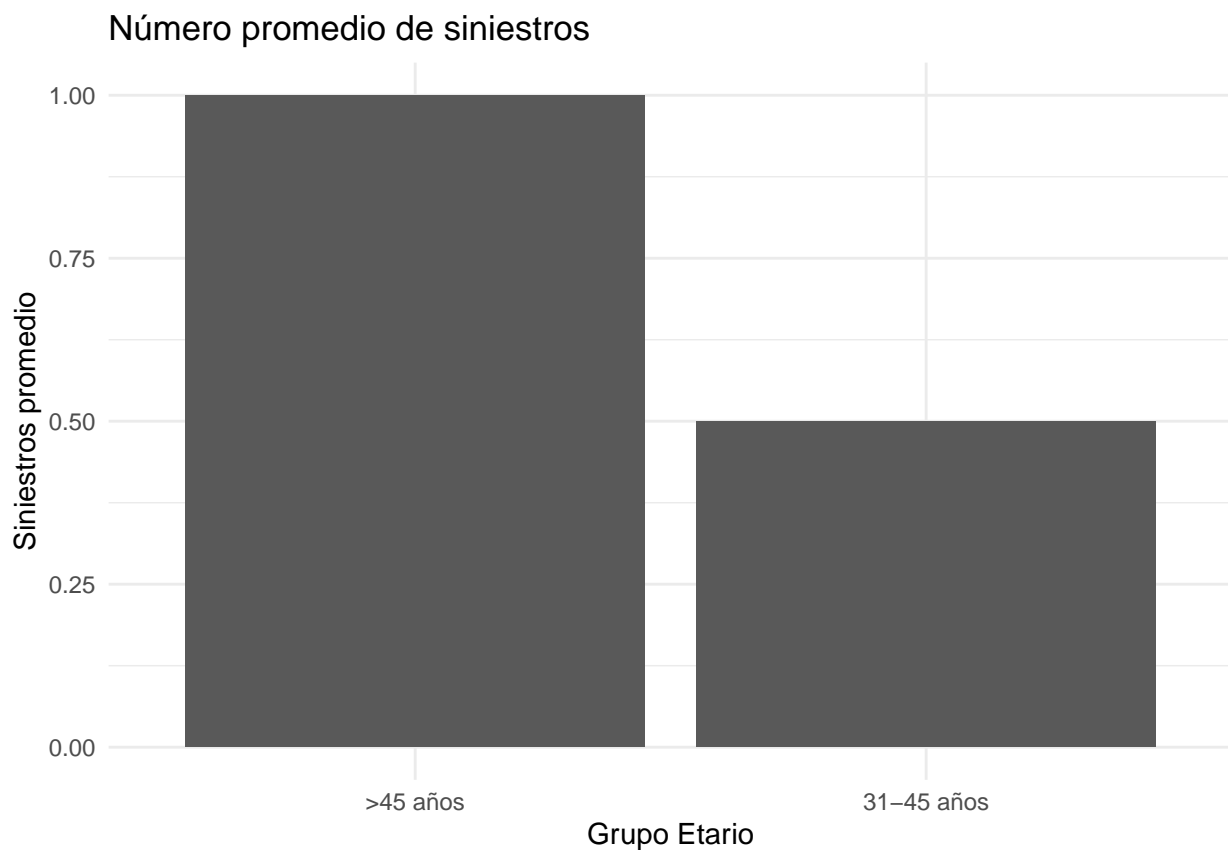
```
# Punto 2
matriz <- matriz %>%
  mutate(GrupoEtario = case_when(
    Edad <= 30 ~ "<30 años",
    Edad <= 45 ~ "31-45 años",
    TRUE ~ ">45 años"
  ))
print(matriz)
```

```
## # A tibble: 3 x 5
##   ID Indicador Siniestros Edad GrupoEtario
##   <dbl>     <dbl>     <dbl> <int> <chr>
## 1     1       0.2         1    50 >45 años
## 2     2       0.5         1    34 31-45 años
## 3     3       0.8         0    33 31-45 años
```

```
# Punto 3
promedio <- matriz %>%
  group_by(GrupoEtario) %>%
  summarise(
    Promedio = mean(Siniestros)
  )
print(promedio)
```

```
## # A tibble: 2 x 2
##   GrupoEtario Promedio
##   <chr>         <dbl>
## 1 31-45 años     0.5
## 2 >45 años      1
```

```
# Punto 4
grafico <- ggplot(data = promedio, aes(x = GrupoEtario, y = Promedio)) +
  geom_col() +
  labs(
    title = "Número promedio de siniestros",
    x = "Grupo Etario",
    y = "Siniestros promedio"
  ) +
  theme_minimal()
print(grafico)
```



```
# Punto 5
 analisis_final <- list(
  datos = matriz,
```

```
resumen = promedio,  
grafico = grafico,  
interpretacion = "Los usuarios mayores de 45 años tienen un promedio de siniestros mayor frente a los  
)  
print( analisis_final$interpretacion)
```

```
## [1] "Los usuarios mayores de 45 años tienen un promedio de siniestros mayor frente a los otros grupo"
```

Cierre reflexivo

¿Qué estructura fue más conveniente para consumir y para reportar resultados?

Para el consumo y análisis exploratorio de los datos, el data.frame es la estructura más conveniente. Esto se debe a que permite visualizar de manera tabular y organizada la información de cada póliza, con columnas definidas que facilitan la comprensión de las relaciones entre variables.

¿Qué decisión hubieras tomado con estos hallazgos si fueses analista de riesgo?

Con solo 3 observaciones, no es posible extraer conclusiones estadísticamente válidas. Los datos parecen veraces, pero la muestra es demasiado pequeña para tomar decisiones de tarificación o ajustes de riesgo. Se recomienda recolectar más datos antes de evaluar cambios en primas o perfiles de riesgo.

¿Qué transformación habría sido imposible si hubieras mantenido la estructura original?

Cruzar variables directamente (edad + riesgo + probabilidad de siniestro) para calcular métricas segmentadas, como la prima específica por edad y nivel de riesgo, sin reorganizar manualmente los datos.