

# Entrega: curso de datos extremos

Laura Montaldo, CI: 3.512.962-7

2024-02-29



UNIVERSIDAD  
DE LA REPÚBLICA  
URUGUAY



# Índice

Resumen	3
Motivación y objetivo del estudio	4
El enfoque de conteo de eventos y los modelos de base Poissoniana.	7
POT (Peaks Over Treshold) y variantes	12

## Resumen

Your abstract goes here.

## Motivación y objetivo del estudio

Los índices de *S&P* son una familia de índices de renta variable<sup>1</sup> diseñados para medir el rendimiento del mercado de acciones en Estados Unidos que cotizan en bolsas estadounidenses. Ésta familia de índices está compuesta por una amplia variedad de índices basados en tamaño, sector y estilo. Los índices están ponderados por el criterio *float-adjusted market capitalization* (FMC). Además, se disponen de índices ponderados de manera equitativa y con límite de capitalización de mercado, como es el caso del *S&P 500*. En este sentido, el *S&P500* entraría en el conjunto de índices ponderados por capitalización bursátil ajustada a la flotación (ver [S&P Dow Jones Indices](#)). El mismo mide el rendimiento del segmento de gran capitalización del mercado estadounidense. Es considerado como un indicador representativo del mercado de renta variable de los Estados Unidos, y está compuesto por 500 empresas constituyentes.

Se busca crear un indicador de una posible crisis bursátil. Como variable de referencia se toma la relación de precios al cierre de ayer sobre la de hoy

$$Indicador_t = \frac{Precio_{t-1}}{Precio_t}, \quad \text{para } t = 1, \dots, T \quad (1)$$

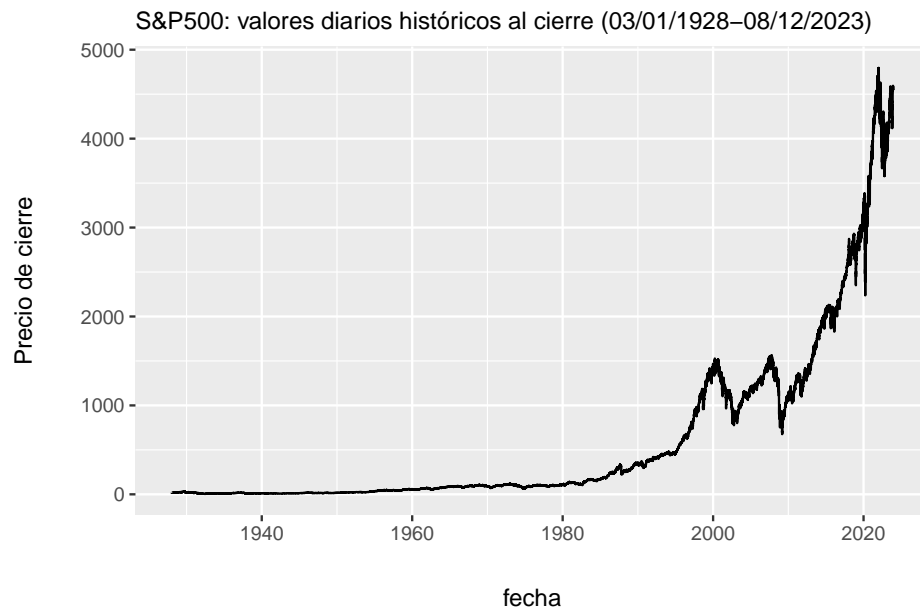
Interpretación del Indicador:

- Si el  $Indicador_t \leq 1$ , el precio de cierre de hoy es mayor o igual que el de ayer, lo cual podría ser considerado una señal positiva.
- Si el  $Indicador_t > 1$ , el precio de cierre de hoy es menor que el de ayer, lo cual podría considerarse una señal de alerta.

En la siguiente figura @ref(fig:plot1) se muestra la evolución histórica desde la fecha 03/01/1928 hasta 08/12/2023 del precio al cierre del día del índice S&P 500.

---

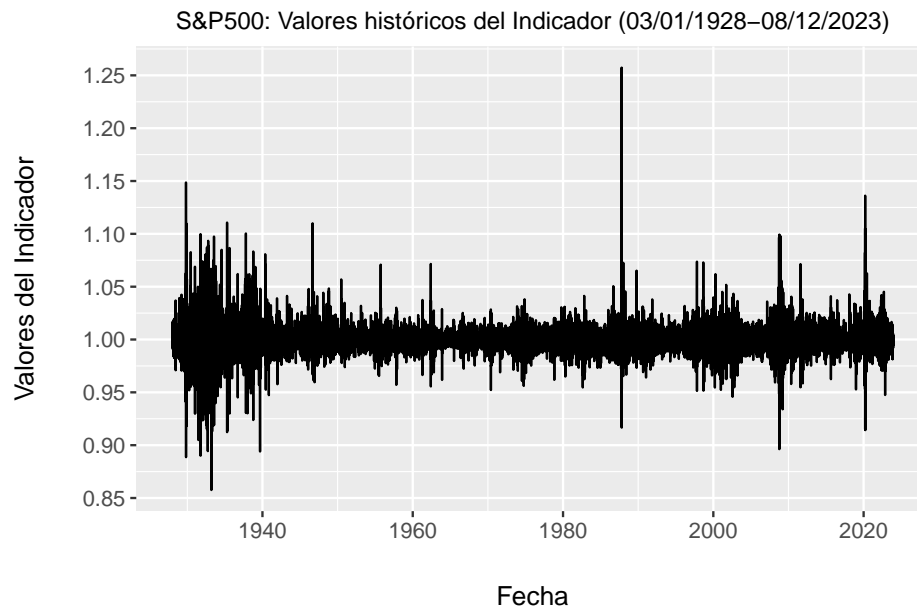
<sup>1</sup>En inglés se llaman equity indices



```
dim(df)
```

```
## [1] 24100      9
```

```
ggplot(df, aes(x = Date, y = relacion)) +
  geom_line() +
  ggtitle("S&P500: Valores históricos del Indicador (03/01/1928-08/12/2023)") +
  xlab("Fecha") +
  ylab("Valores del Indicador") +
  scale_x_date(limits = date_range) +
  scale_y_continuous(breaks = seq(0, ceiling(max(df$relacion)), by = 0.05)) +
  theme(
    axis.title.x = element_text(margin = margin(t = 20, b = 40)),
    axis.title.y = element_text(margin = margin(r = 20, l = 40)),
    plot.title = element_text(size = 11, hjust = 0.5), # Center the plot title
    axis.text = element_text(size = 10), # Adjust the size of axis text
    axis.title = element_text(size = 12), # Adjust the size of axis titles
    legend.title = element_text(size = 10), # Adjust the size of legend title
    legend.text = element_text(size = 8) # Adjust the size of legend text
  )
```



A la columna relativa a la relacion de precios se la resta por 1 para tener centrados los valores de la relacion de precios en cero. Y posteriormente analizar si las series, fijando distintos umbrales son estacionarias.

```
df$rel_cero=df$relacion-1
head(df)
```

```
##      Date  Open  High   Low Close Volume Dividends Stock.Splits  relacion
## 2 1928-01-03 17.76 17.76 17.76 17.76      0         0           0 0.9943693
## 3 1928-01-04 17.72 17.72 17.72 17.72      0         0           0 1.0022574
## 4 1928-01-05 17.55 17.55 17.55 17.55      0         0           0 1.0096866
## 5 1928-01-06 17.66 17.66 17.66 17.66      0         0           0 0.9937712
## 6 1928-01-09 17.50 17.50 17.50 17.50      0         0           0 1.0091428
## 7 1928-01-10 17.37 17.37 17.37 17.37      0         0           0 1.0074841
##      rel_cero
## 2 -0.005630652
## 3  0.002257388
## 4  0.009686614
## 5 -0.006228800
## 6  0.009142848
## 7  0.007484119
```

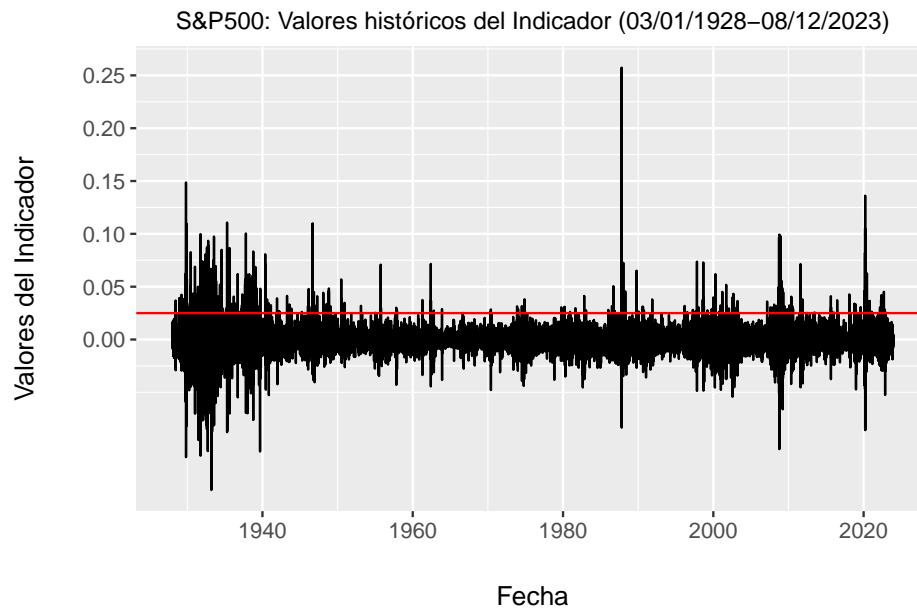
## El enfoque de conteo de eventos y los modelos de base Poissoniana.

Fijaremos un cierto umbral, llamaremos “evento” a cuando la variable observada supera ese umbral y, dado un cierto intervalo del tiempo  $J$ , contaremos

$N(J)$  = número de eventos en el intervalo  $J$ .

$N(\text{mes})$  = cantidad de períodos de 3 días durante un mes, en que se registró un aumento mayor o igual a 2.5% del índice de un día para el otro.

```
ggplot(df, aes(x = Date, y = rel_cero)) +  
  geom_line() +  
  geom_hline(yintercept = 0.025, linetype = "solid", color = "red") + # Add horizontal line  
  ggtitle("S&P500: Valores históricos del Indicador (03/01/1928-08/12/2023)") +  
  xlab("Fecha") +  
  ylab("Valores del Indicador") +  
  scale_x_date(limits = date_range) +  
  scale_y_continuous(breaks = seq(0, ceiling(max(df$relacion)), by = 0.05)) +  
  theme(  
    axis.title.x = element_text(margin = margin(t = 20, b = 40)),  
    axis.title.y = element_text(margin = margin(r = 20, l = 40)),  
    plot.title = element_text(size = 11, hjust = 0.5), # Center the plot title  
    axis.text = element_text(size = 10), # Adjust the size of axis text  
    axis.title = element_text(size = 12), # Adjust the size of axis titles  
    legend.title = element_text(size = 10), # Adjust the size of legend title  
    legend.text = element_text(size = 8) # Adjust the size of legend text  
  )
```



```
filtered_df_0_025 <- df %>%
  filter(rel_cero>= 0.025)
```

```
head(filtered_df_0_025)
```

```
##           Date  Open  High   Low Close Volume Dividends Stock.Splits relacion
## 1 1928-06-11 18.68 18.68 18.68 18.68      0          0           0 1.036938
## 2 1928-07-11 18.95 18.95 18.95 18.95      0          0           0 1.025330
## 3 1928-12-06 22.91 22.91 22.91 22.91      0          0           0 1.039284
## 4 1929-02-07 24.71 24.71 24.71 24.71      0          0           0 1.031566
## 5 1929-03-25 24.51 24.51 24.51 24.51      0          0           0 1.042432
## 6 1929-04-01 24.88 24.88 24.88 24.88      0          0           0 1.026125
##      rel_cero
## 1 0.03693793
## 2 0.02532979
## 3 0.03928414
## 4 0.03156620
## 5 0.04243162
## 6 0.02612546
```

```
data=filtered_df_0_025[,c('Date', 'rel_cero')]
n=dim(data)[1]
```

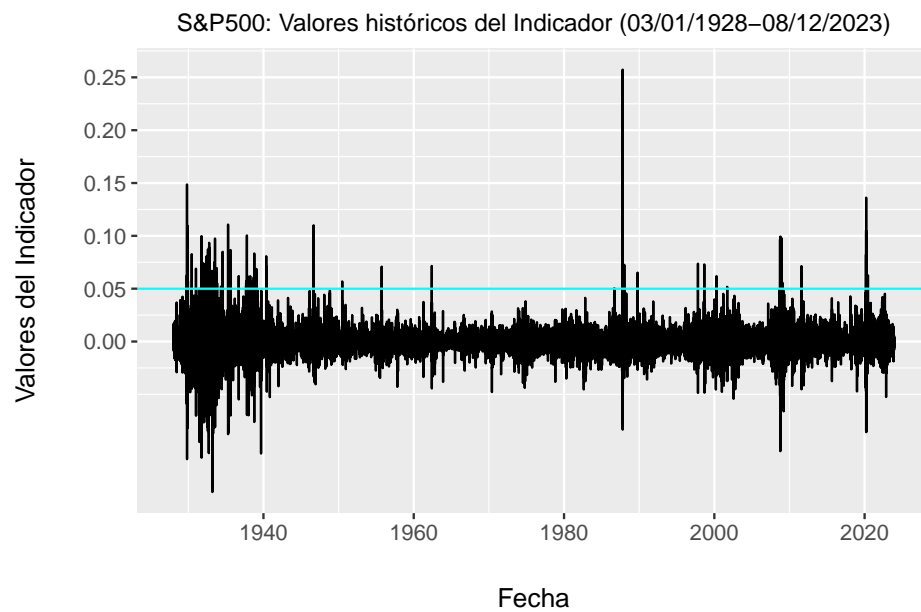


```
fecha_maxima <- max(data$Date)
```

```
# Reescalar el tiempo dividiendo cada fecha por la fecha máxima
data$tiempo_reescalado <- as.numeric(data$Date - min(data$Date)) / as.numeric(fecha_maxima - min(data$Date))
head(data)
```

```
##           Date    rel_cero tiempo_reescalado
## 1 1928-06-11 0.03693793      0.00000000000
## 2 1928-07-11 0.02532979      0.0008690614
## 3 1928-12-06 0.03928414      0.0051564311
## 4 1929-02-07 0.03156620      0.0069814600
## 5 1929-03-25 0.04243162      0.0083140209
## 6 1929-04-01 0.02612546      0.0085168019
```

```
ggplot(df, aes(x = Date, y = rel_cero)) +
  geom_line() +
  geom_hline(yintercept = 0.05, linetype = "solid", color = "cyan") + # Add horizontal line
  ggtitle("S&P500: Valores históricos del Indicador (03/01/1928-08/12/2023)") +
  xlab("Fecha") +
  ylab("Valores del Indicador") +
  scale_x_date(limits = date_range) +
  scale_y_continuous(breaks = seq(0, ceiling(max(df$relacion)), by = 0.05)) +
  theme(
    axis.title.x = element_text(margin = margin(t = 20, b = 40)),
    axis.title.y = element_text(margin = margin(r = 20, l = 40)),
    plot.title = element_text(size = 11, hjust = 0.5), # Center the plot title
    axis.text = element_text(size = 10), # Adjust the size of axis text
    axis.title = element_text(size = 12), # Adjust the size of axis titles
    legend.title = element_text(size = 10), # Adjust the size of legend title
    legend.text = element_text(size = 8) # Adjust the size of legend text
  )
```

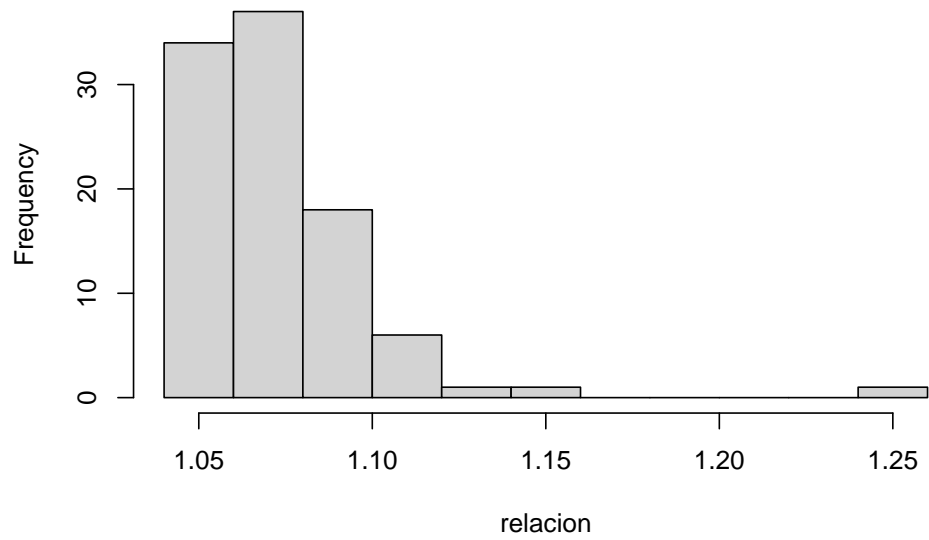


```
filtered_df_0_05 <- df %>%  
  filter(rel_cero >= 0.05)  
dim(filtered_df_0_05)
```

```
## [1] 98 10
```

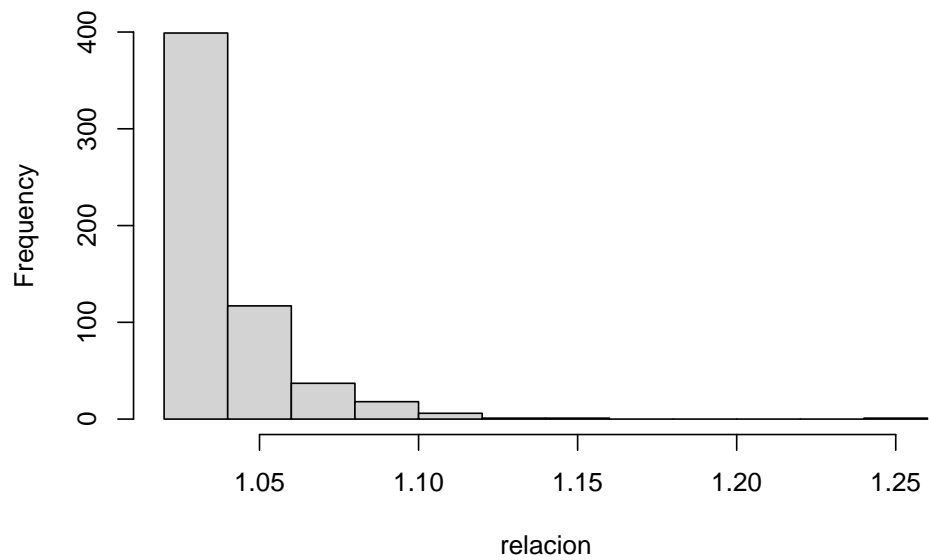
```
hist(filtered_df_0_05$relacion, main = "Histograma relacion con umbral 0.05 ", xlab = "rela
```

**Histograma relacion con umbral 0.05**



```
hist(filtered_df_0_025$relacion, main = "Histograma relacion con umbral 0.025 ", xlab = "relacion")
```

**Histograma relacion con umbral 0.025**



## POT (Peaks Over Treshold) y variantes

- Perera, Gonzalo, Angel Segura, y Carolina Crisci. 2021. *Curso de estadística de datos extremos, cap. 1 a cap. 5*.
- Stephenson, A. G. 2002. «evd: Extreme Value Distributions». *R News* 2 (2): 0.  
<https://CRAN.R-project.org/doc/Rnews/>.