
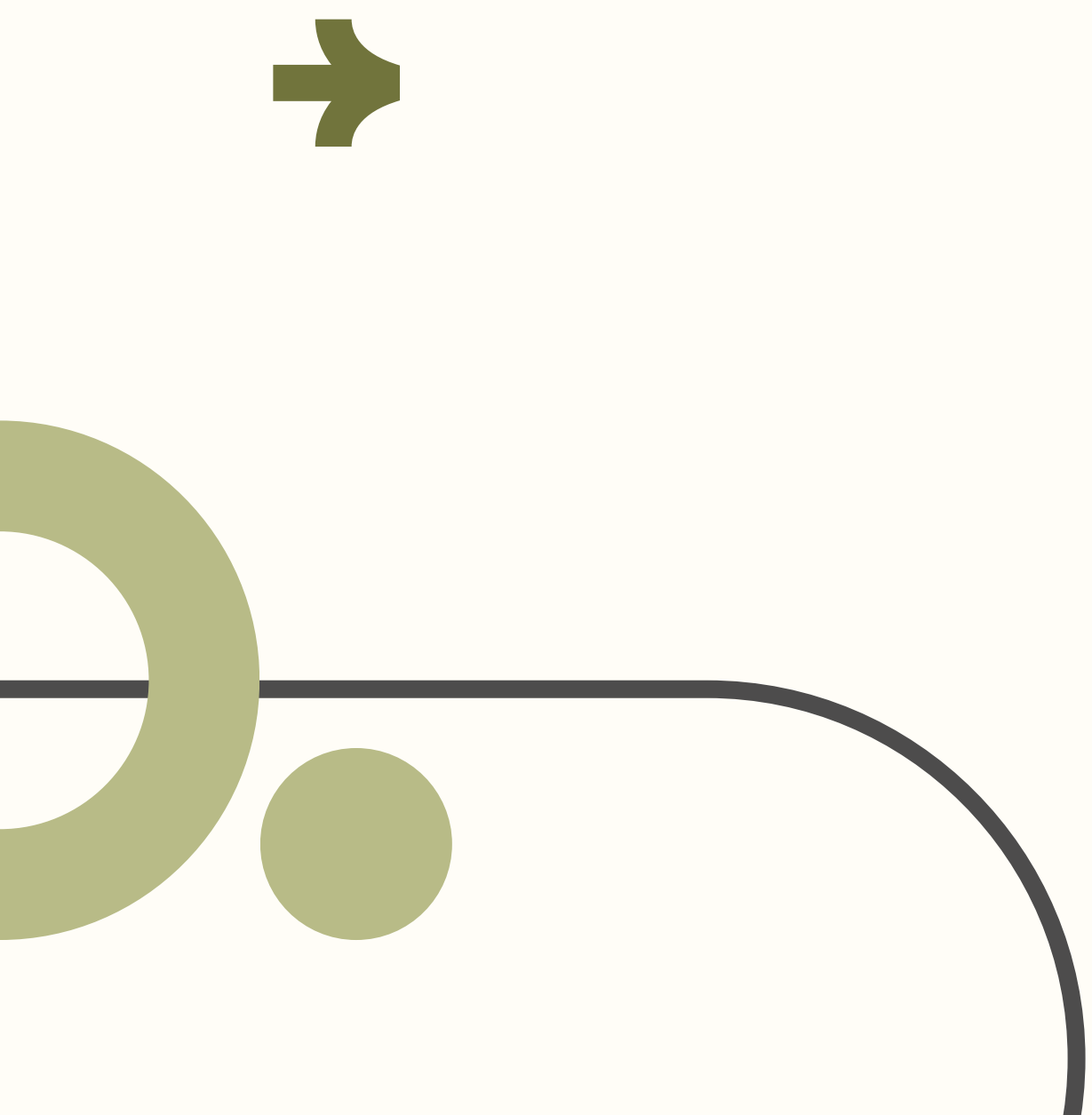

Ciencias Actuariales
Universidad de Costa Rica

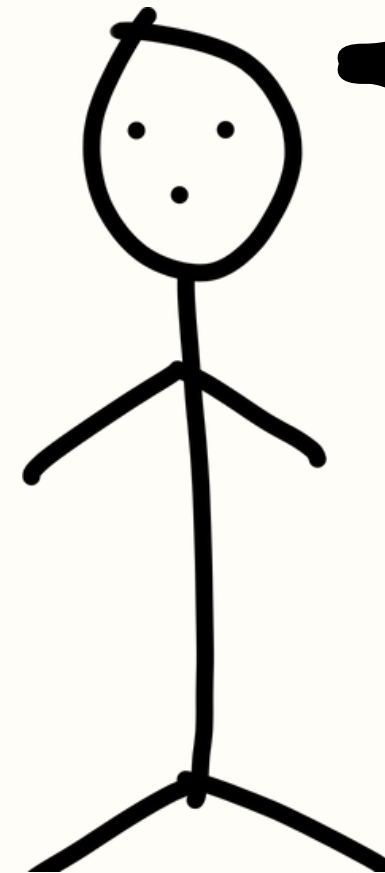


ANÁLISIS DE DATOS: DESCUBRIENDO LA DISTRIBUCIÓN IDEAL CON MÉTRICAS AVANZADAS

Ashley Padilla



Ya me aprendí todas las distribuciones en el curso de Proba, pero si me dan unos datos ¿Cómo saber cual distribución tienen?



01

Criterio de Información de Akaike (AIC)

$$AIC = 2k - 2\ln(L)$$

k = números de parámetros

L = máximo valor de la función de verosimilitud

- Valores más bajos de AIC: Indican un mejor ajuste del modelo a los datos.

Criterio de Información Bayesiano (BIC)

$$BIC = k \ln(n) - 2 \ln(L)$$

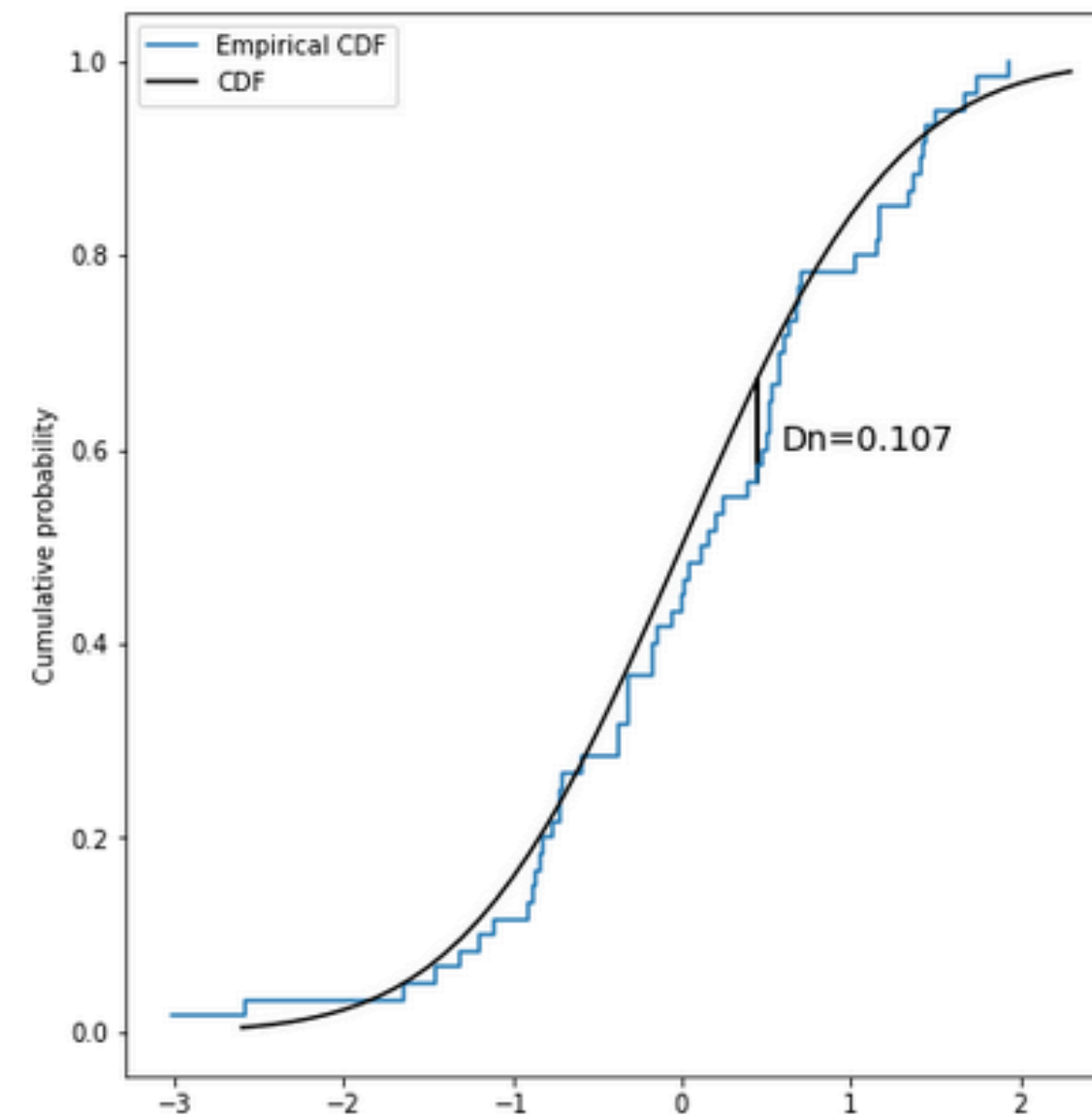
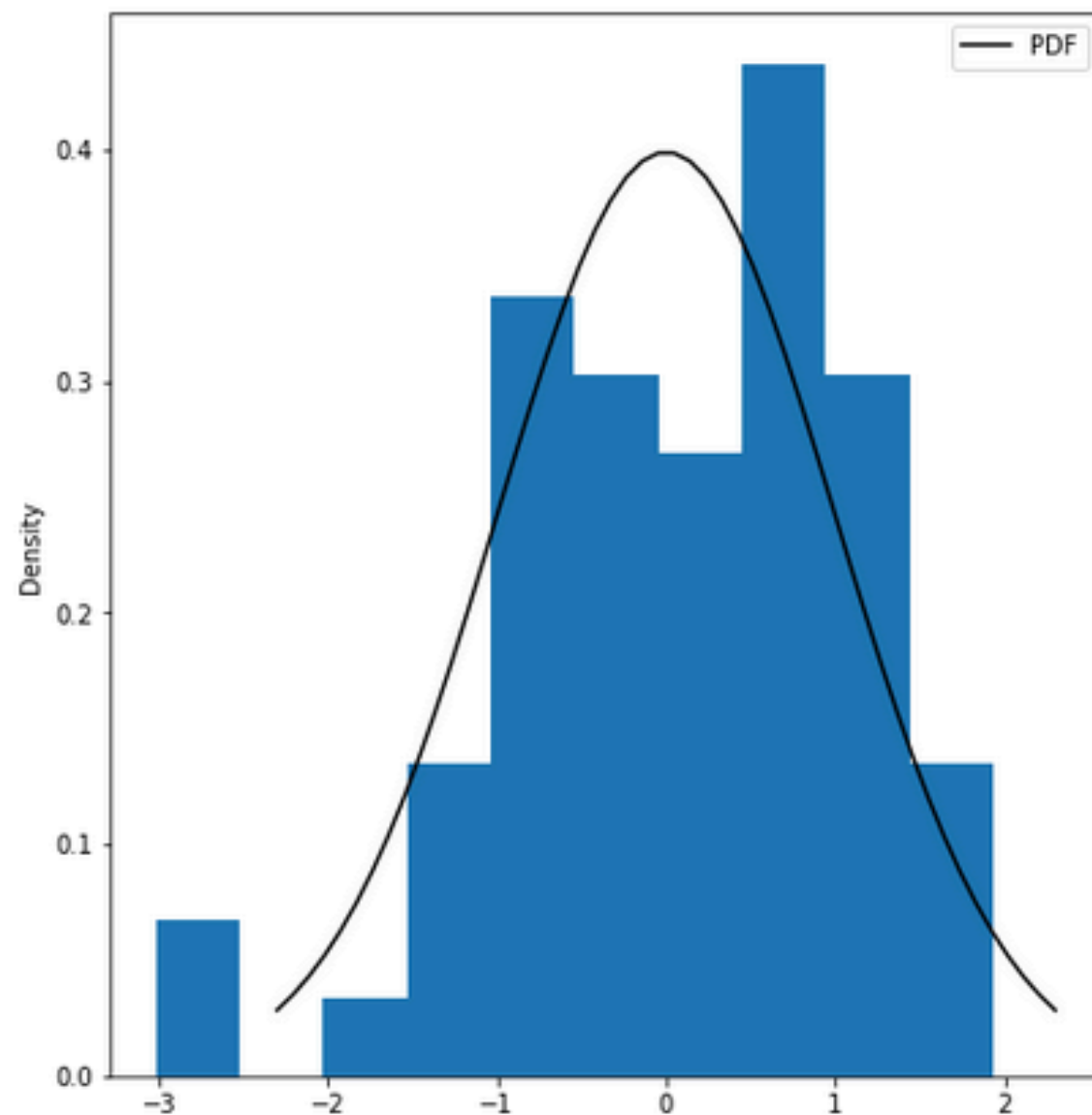
n = tamaño de la muestra

k = números de parámetros

L = máximo valor de la función de verosimilitud

03

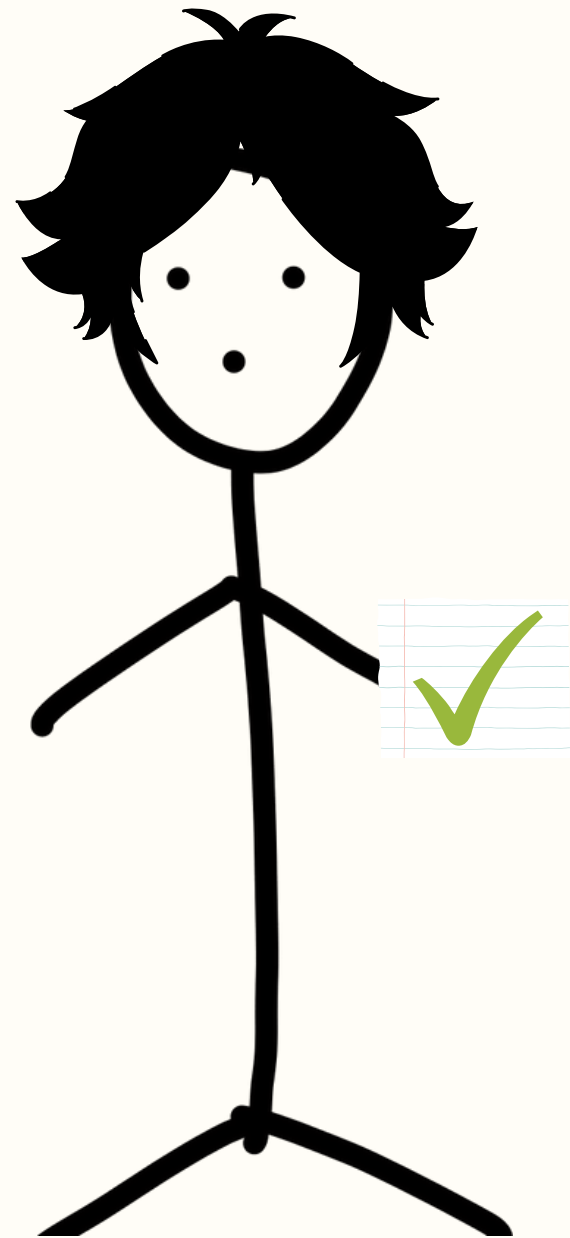
Kolmogorov-Smirnov



Kolmogorov- Smirnov

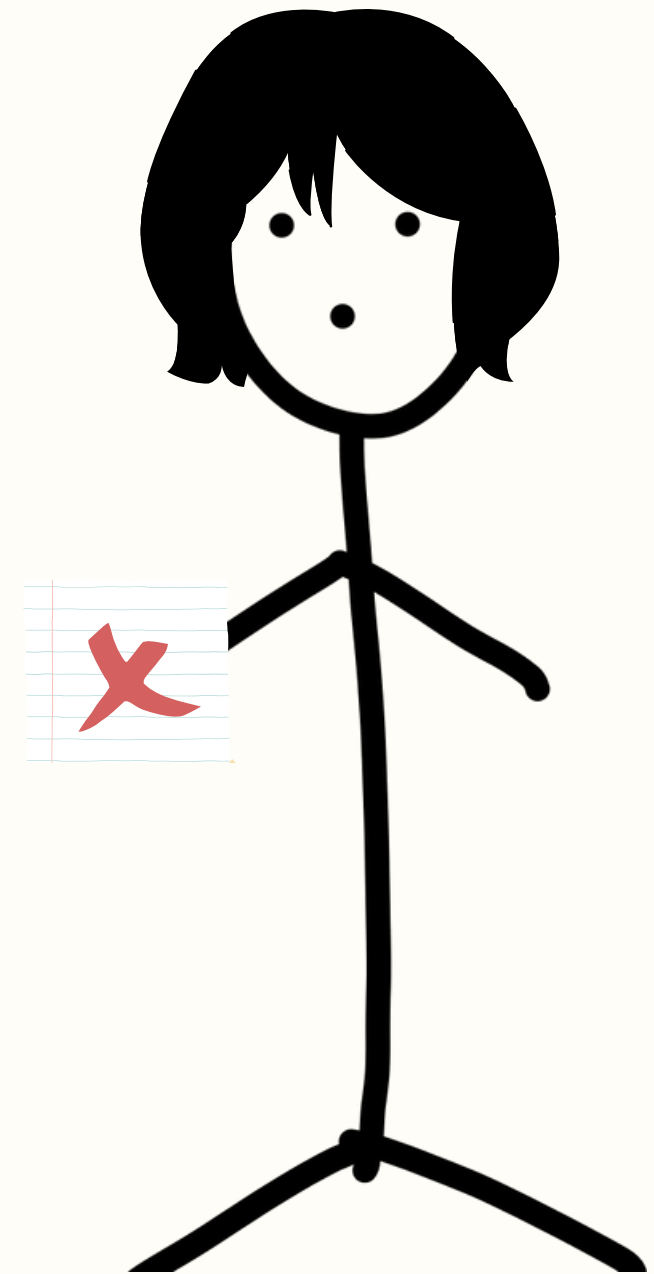
$$H_0 : F(x) = F_0(x)$$

$$H_1 : F(x) \neq F_0(x)$$



No Rechazamos

ó



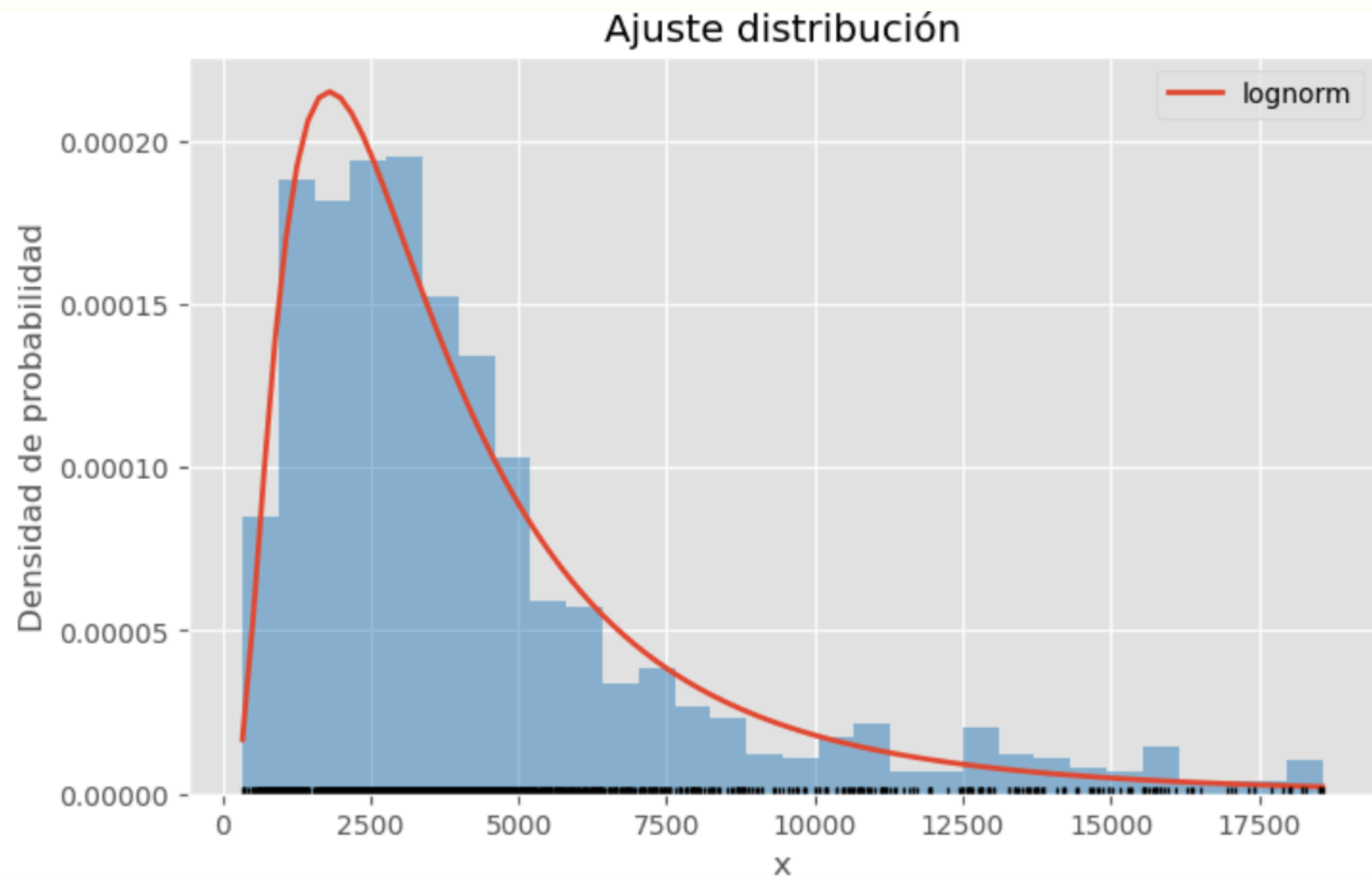
Rechazamos

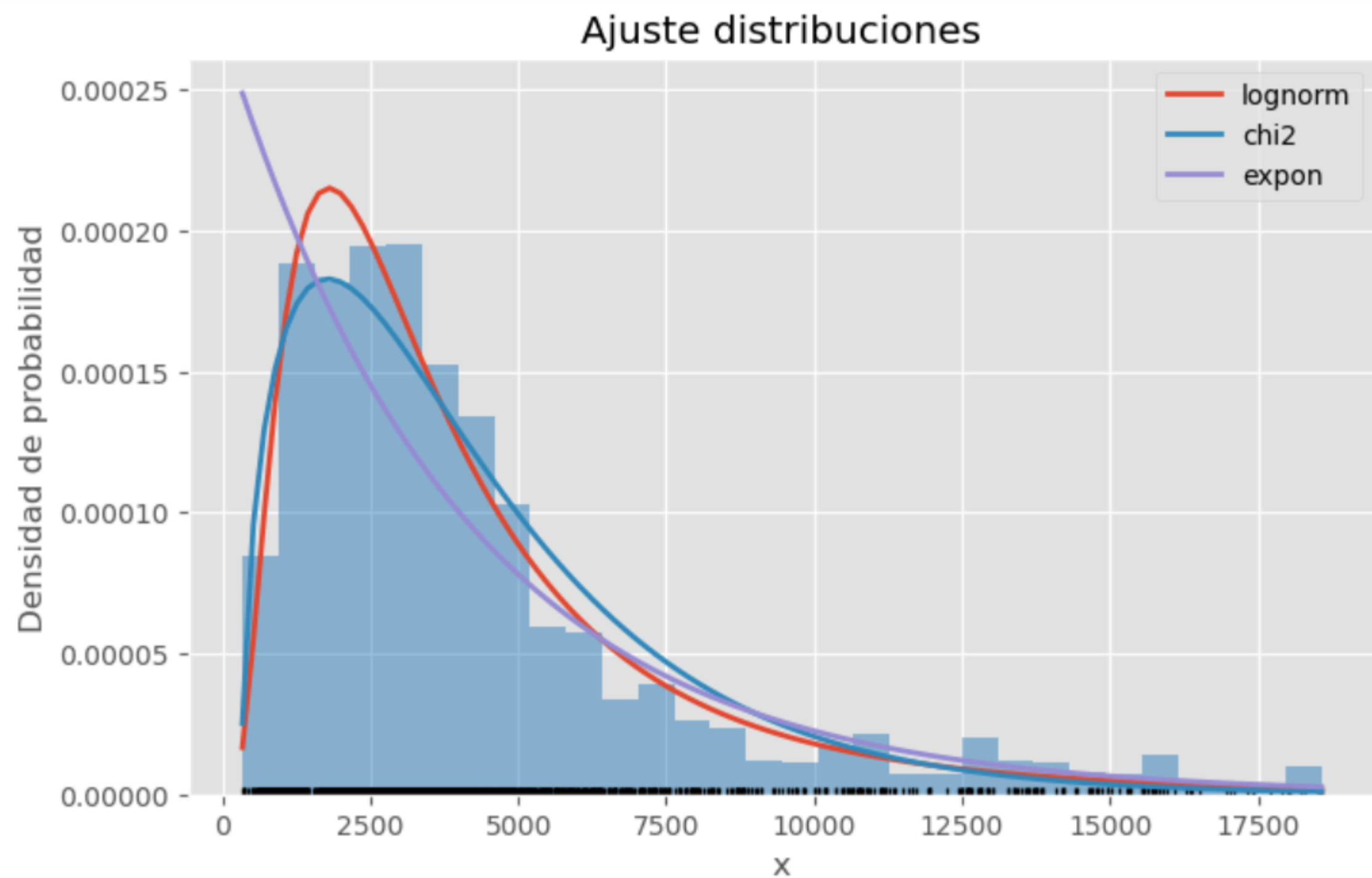


DISTRIBUCIONES A USAR



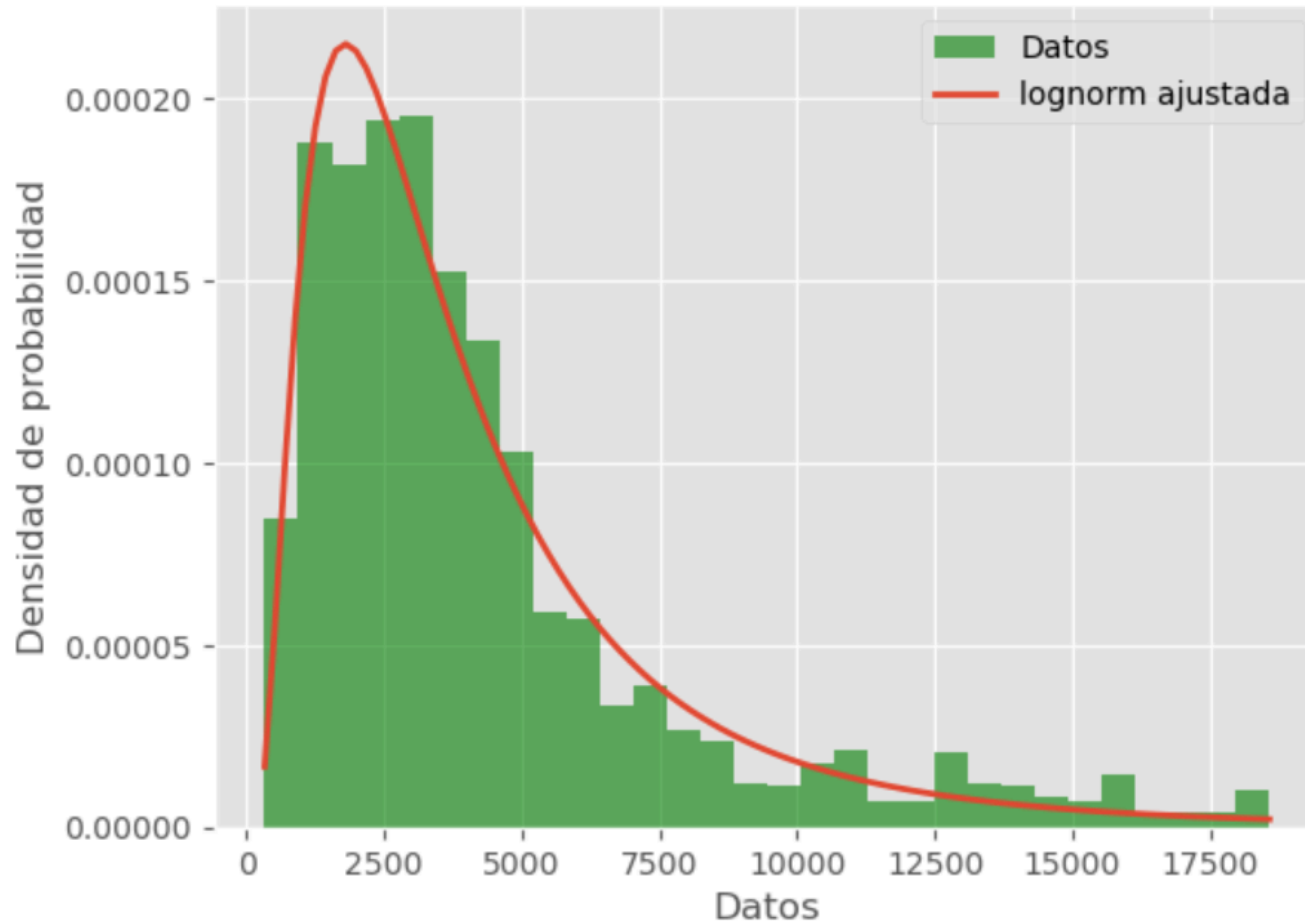
**Normal
Exponencial
Uniforme
Log-normal
Weibull
Gamma
Beta
Chi2
T-Student
Bernoulli
Binomial
Poisson.**





<code>l]</code> :	distribucion	log_likelihood	aic	bic	n_parametros
0	lognorm	-1.488561e+04	2.977722e+04	2.979337e+04	3
1	chi2	-1.490368e+04	2.981336e+04	2.982952e+04	3
2	expon	-1.497215e+04	2.994831e+04	2.995908e+04	2
3	beta	-1.501004e+04	3.002809e+04	3.004962e+04	4
4	t	-1.519599e+04	3.039799e+04	3.041414e+04	3
5	gamma	-1.523564e+04	3.047728e+04	3.049343e+04	3
6	norm	-1.544997e+04	3.090393e+04	3.091470e+04	2
7	uniform	-1.579604e+04	3.159609e+04	3.160686e+04	2
8	weibull_min	-inf	inf	inf	3

Histograma de datos y mejores distribuciones ajustadas



Distribución: norm

KS estadístico: 0.16692744536016302

KS valor p: 1.111506990349673e-39

Distribución: expon

KS estadístico: 0.10883597681271057

KS valor p: 4.603240165410763e-17

Distribución: uniform

KS estadístico: 0.488205399098209

KS valor p: 0.0

Distribución: lognorm

KS estadístico: 0.027434988792896298

KS valor p: 0.17386130955932377

Distribución: weibull_min

KS estadístico: 0.06337762213032727

KS valor p: 4.5847084017769e-06

Distribución: gamma

KS estadístico: 0.10763352538153936

KS valor p: 1.0697455564700814e-16

Distribución: beta

KS estadístico: 0.12101423519016863

KS valor p: 5.2760099278194815e-21

Distribución: chi2

KS estadístico: 0.05286116457477241

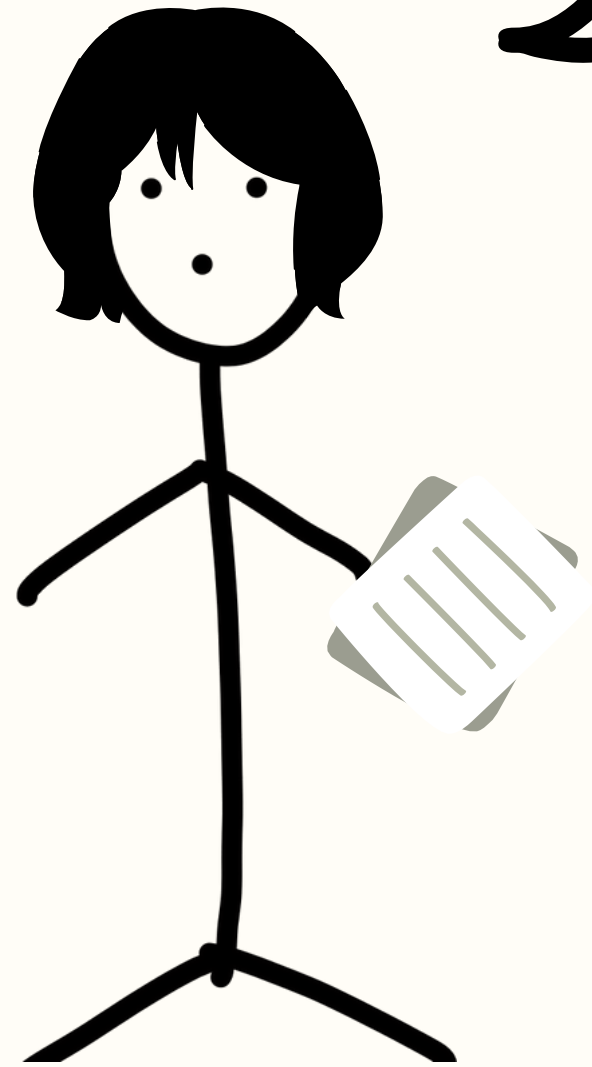
KS valor p: 0.0002377885215068998

Distribución: t

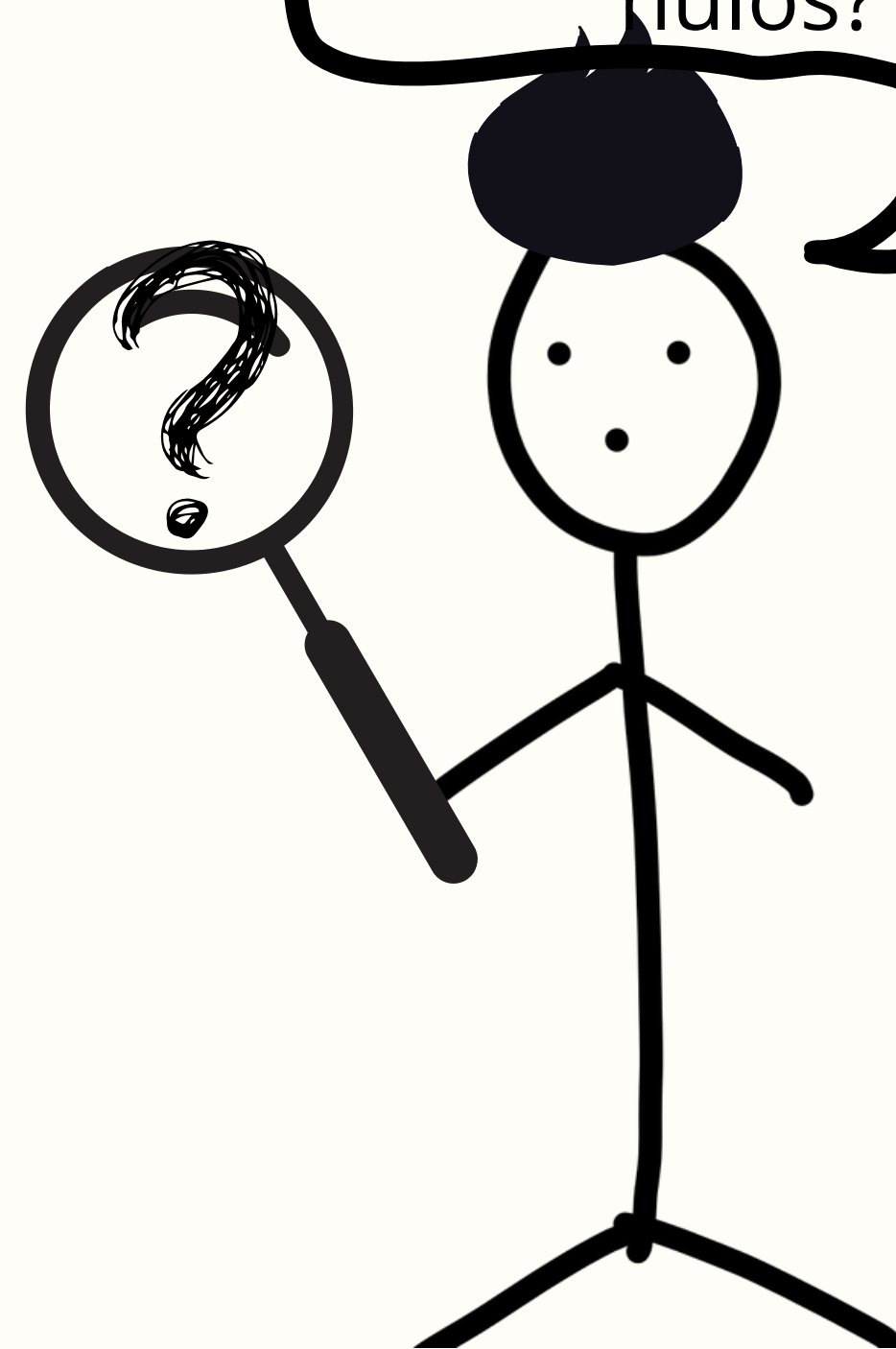
KS estadístico: 0.1302127283524324

KS valor p: 2.920324001677451e-24

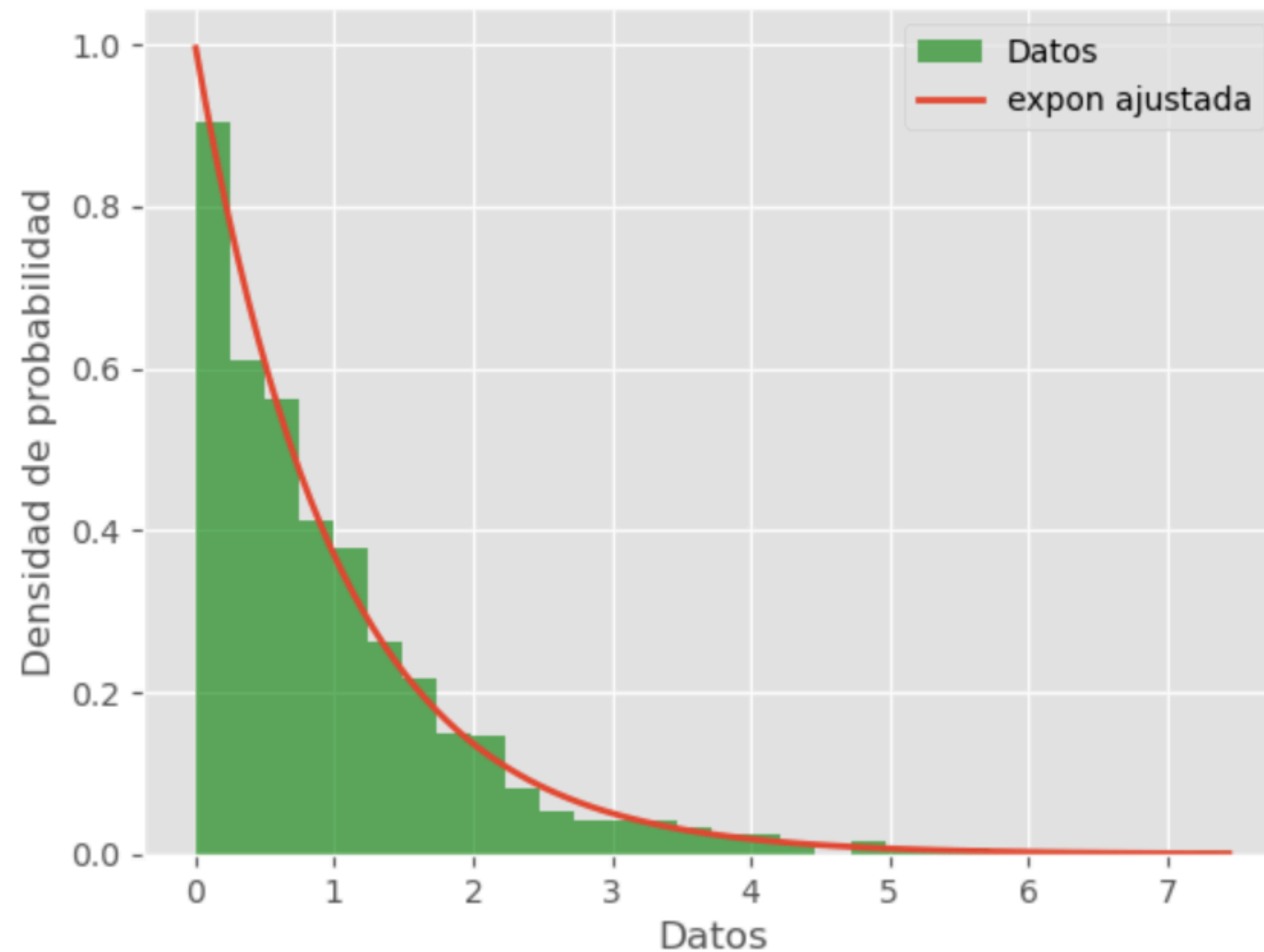
Listo! Ya sabemos
qué distribución
usar



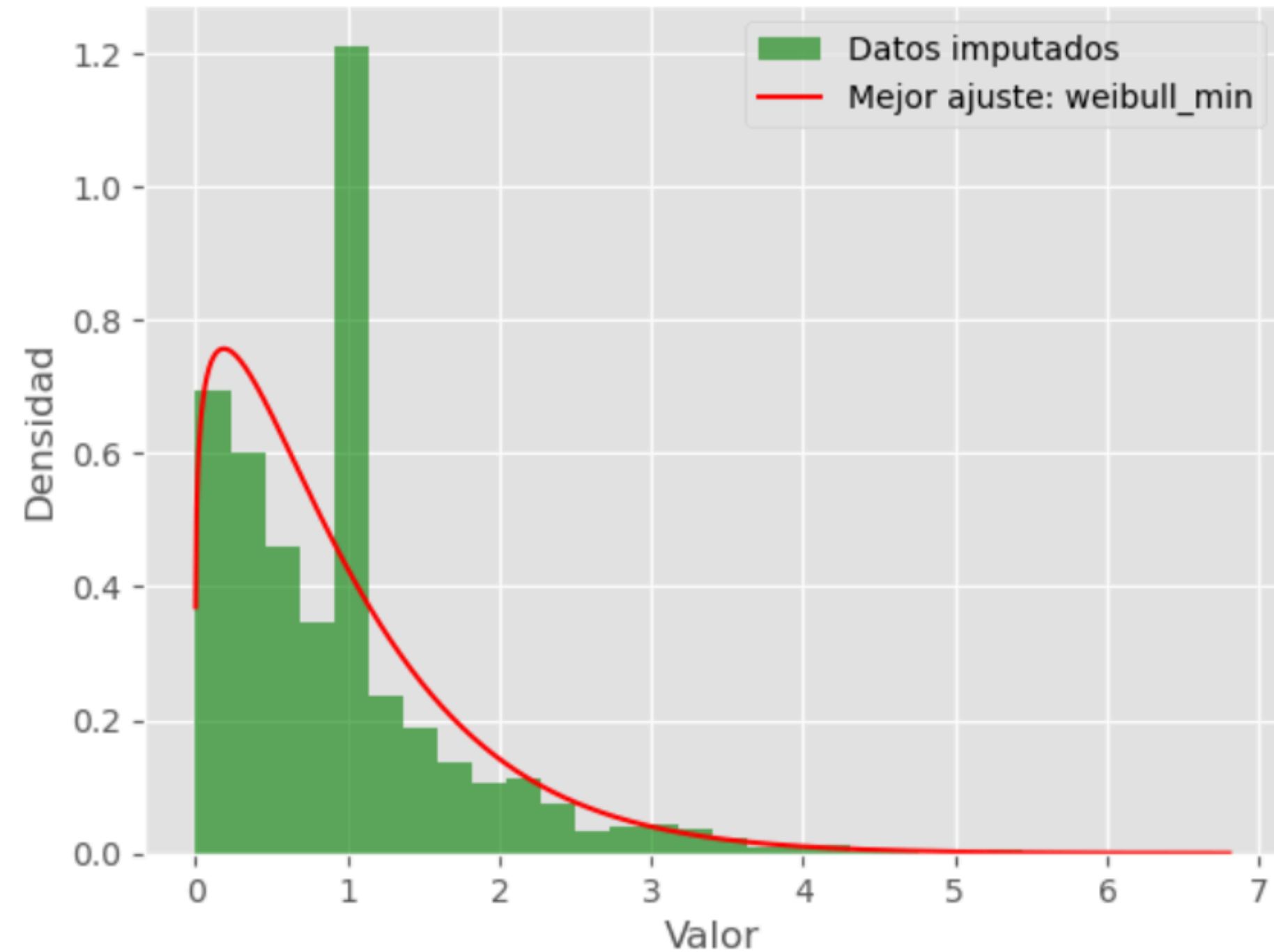
¿Qué pasa si
tenemos valores
nulos?



Histograma de datos y mejores distribuciones ajustadas

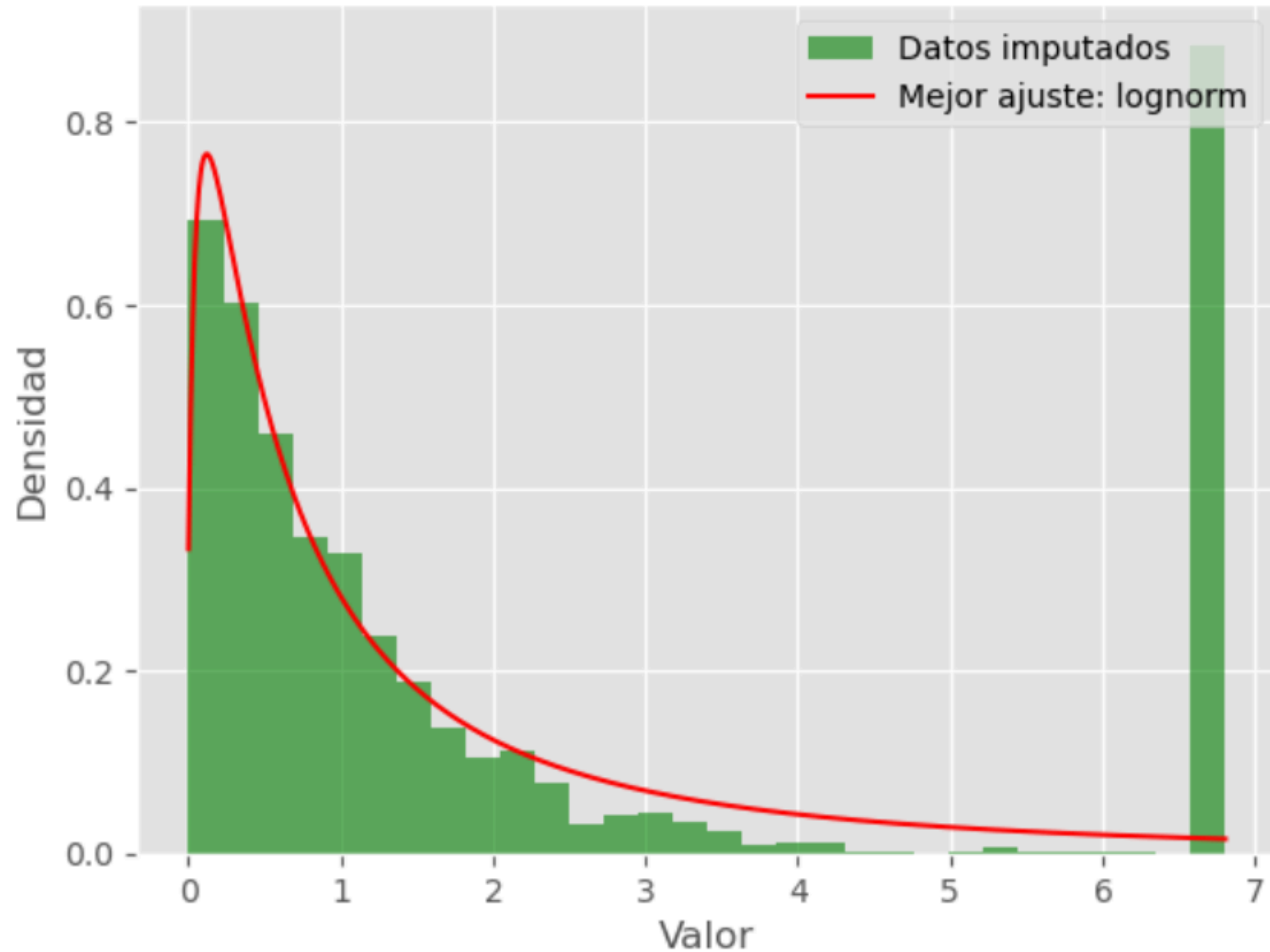


Histograma de datos imputados con la distribución ajustada



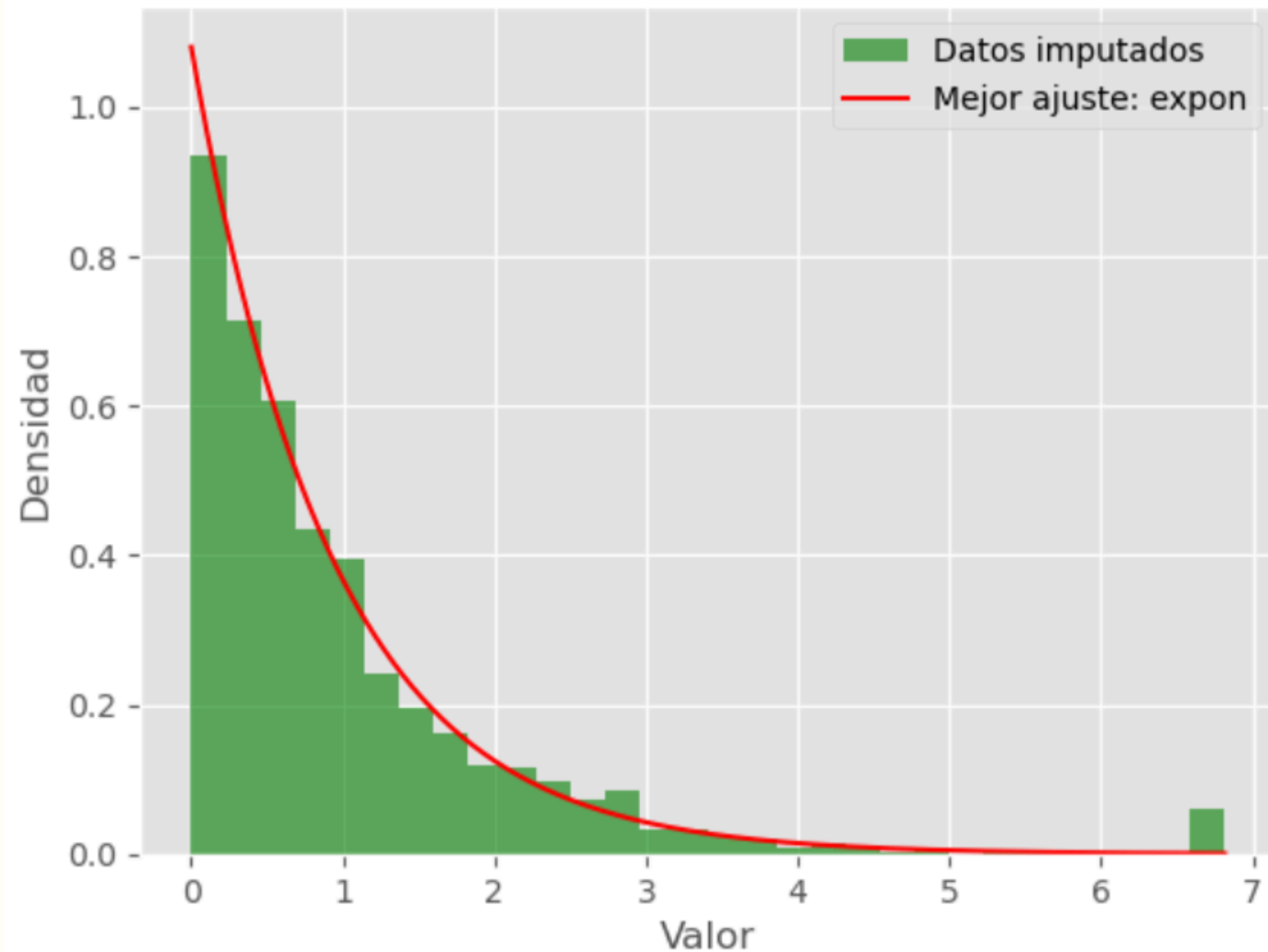
300 datos imputados
con la media

Histograma de datos imputados con la distribución ajustada



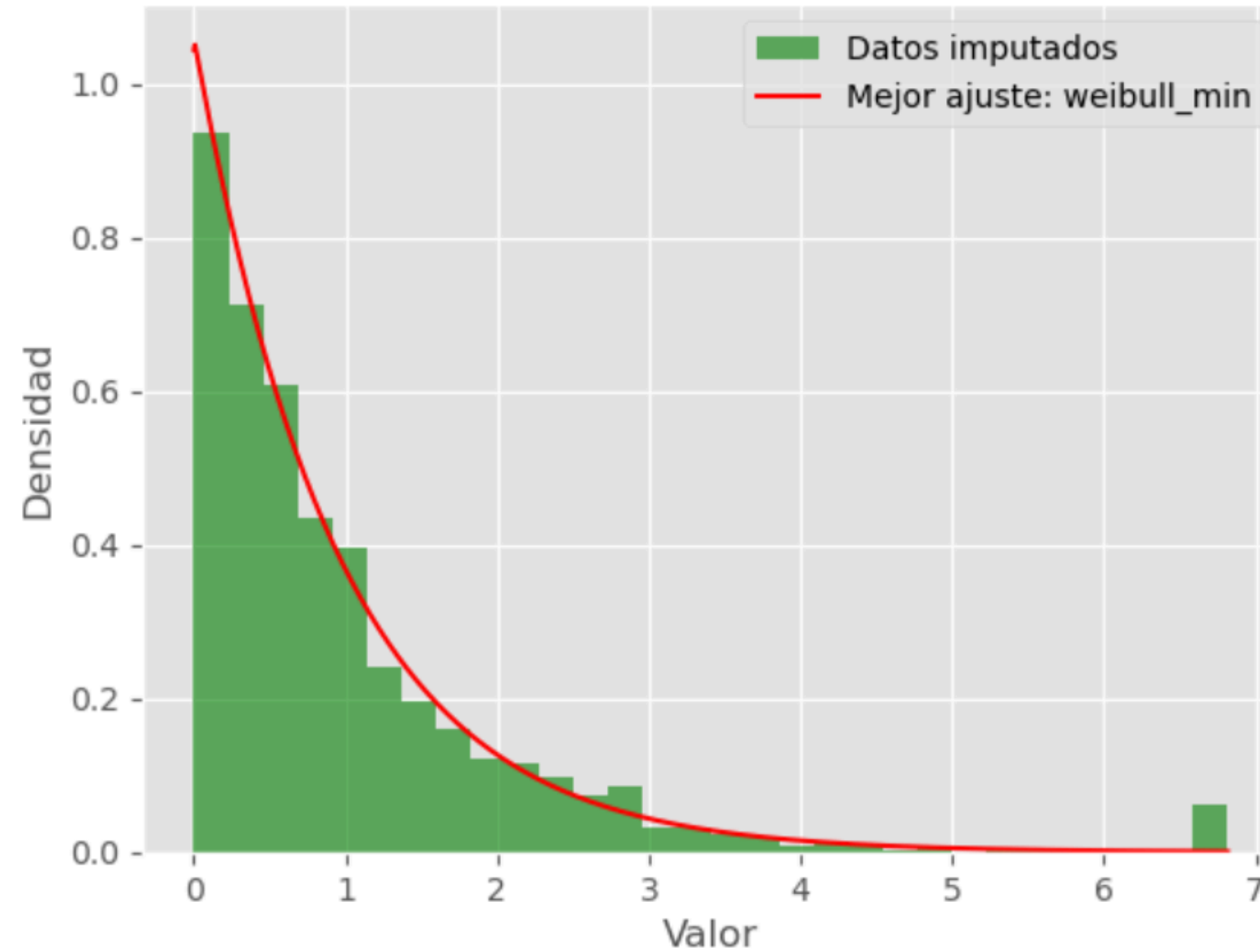
300 datos imputados
con la maximo

Histograma de datos imputados con la distribución ajustada



20 datos imputados
con el minimo

Histograma de datos imputados con la distribución ajustada



20 datos imputados
con la media



Conclusiones

Comprender la Distribución Original

Evaluar el Impacto de la Imputación

Consistencia con los Análisis Futuros



Muchas
GRACIAS

