

## Introdução

Certas amostras de dados apresentam distribuição consideravelmente diferente da distribuição normal. Um dos pressupostos da inferência estatística clássica é o da normalidade dos resíduos. Outro é a homoscedasticidade dos mesmos. Estas duas hipóteses dificilmente são atingidos se a distribuição da variável dependente não for próxima da distribuição normal.

Para contornar este problema, podemos:

- 1 Transformar a variável resposta de maneira que a sua distribuição torne-se aproximadamente normal
- 2 Utilizar o método dos mínimos quadrados ponderados, de forma a remover a heteroscedasticidade do modelo
- 3 Calcular erros heteroscedásticos consistentes, com o método de Eicker-White, por exemplo.

Neste trabalho, focamos na primeira abordagem, qual seja, a da transformação da variável resposta.

## Gênese da distribuição Lognormal

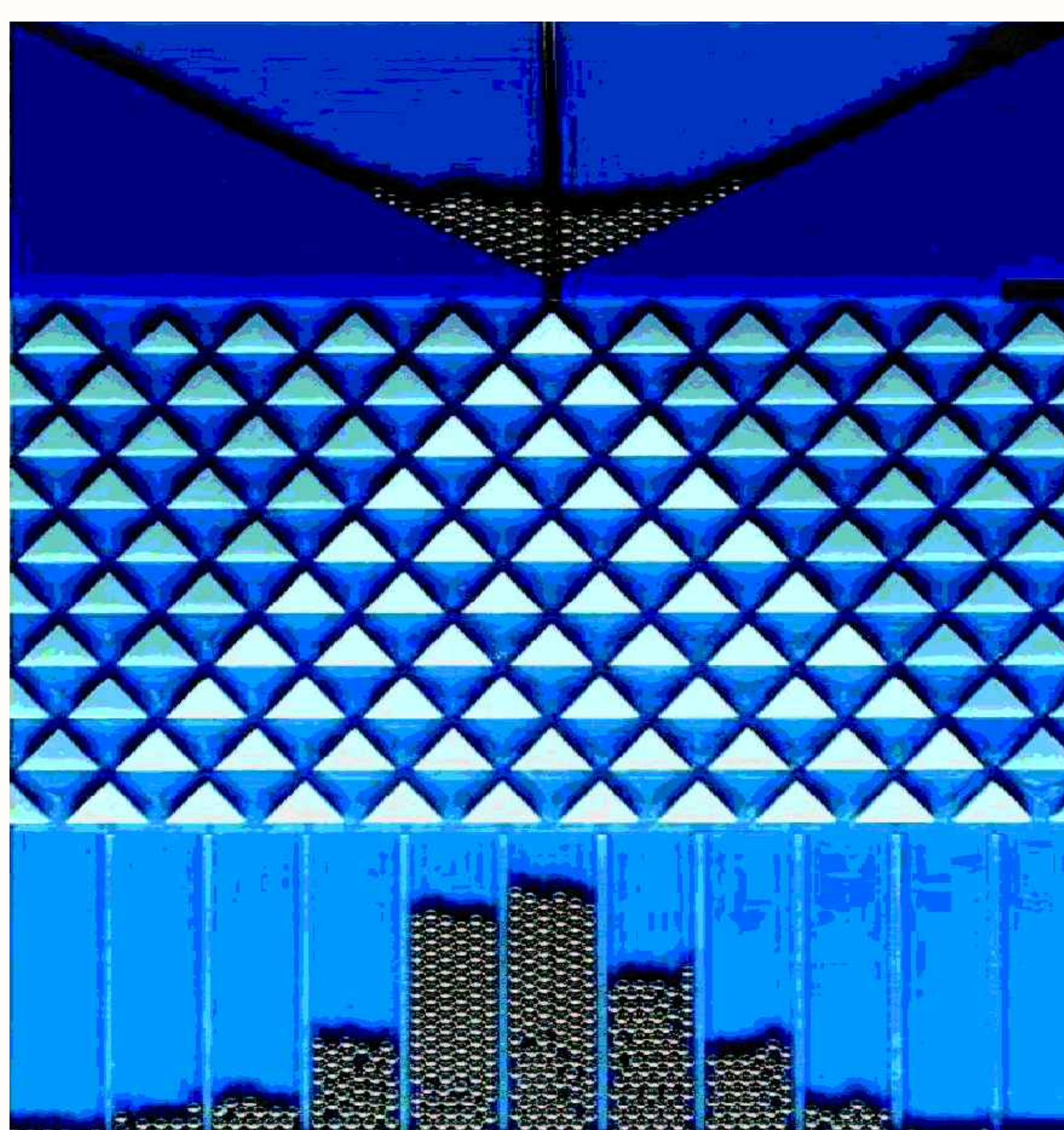


Figura 1: Modelo físico demonstrando a gênese da distribuição normal. FONTE: 1

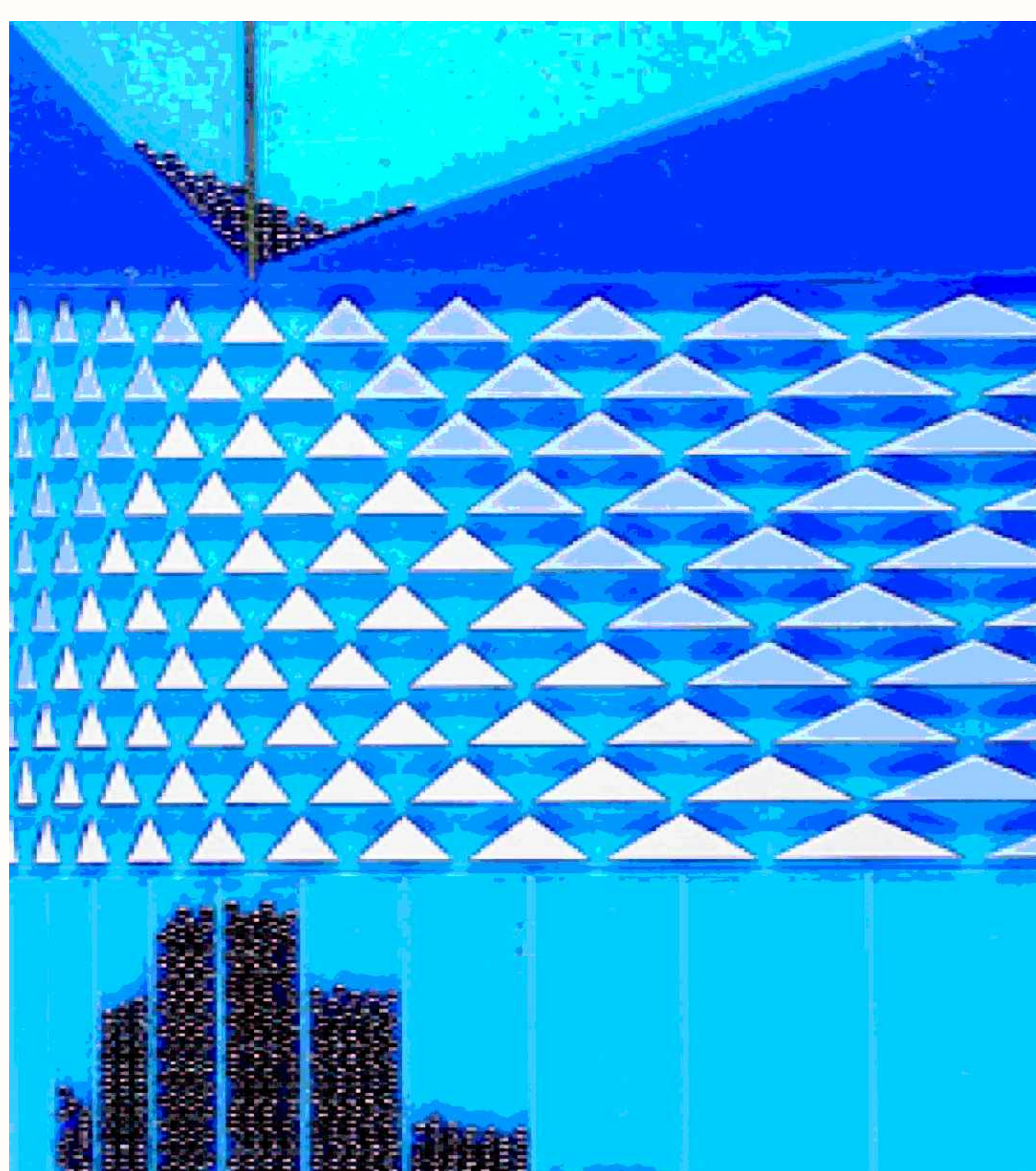


Figura 2: Modelo físico demonstrando a gênese da distribuição lognormal. FONTE: 1

## Características

### Medidas de tendência central

Na distribuição lognormal, diferentemente do que ocorre nas distribuições simétricas, como a distribuição normal, moda, média e mediana tem valores distintos, o que pode ser observado na figura 3.

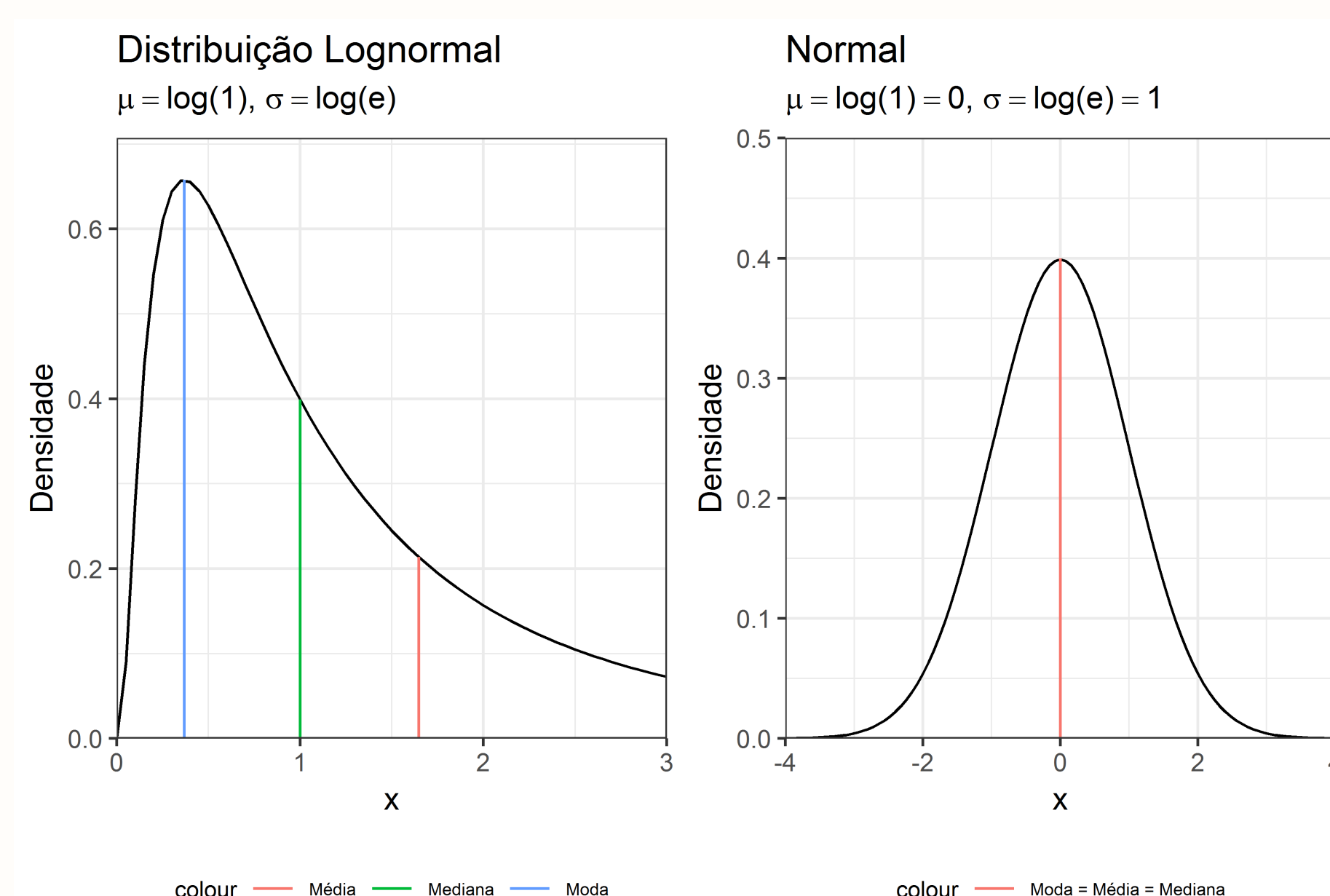


Figura 3: Comparação entre distribuição normal e lognormal. FONTE: Autor

### Formulação

a. Função densidade de probabilidade

$$f(x; \mu, \sigma) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right) & \forall x > 0 \\ 0 & \text{se } x = 0 \end{cases}$$

b. Valor Esperado (Média)

$$\mathbb{E}(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

c. Moda

$$Mo = \exp(\mu - \sigma^2)$$

d. Mediana

$$v = \exp(\mu)$$

e. Variância

$$\text{Var}(X) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$$

$$\text{Var}(X) = \mathbb{E}^2(X)(\exp(\sigma^2) - 1)$$

f. Coeficiente de Variação

$$CV = \sqrt{e^{\sigma^2} - 1}$$

g. Parâmetros na escala anti-logarítmica

$$\mu^* = e^\mu$$

$$\sigma^* = e^\sigma$$

h. Intervalos de Confiança

@68,3% (1σ)

$$[\mu^*/\sigma^*; \mu^* \cdot \sigma^*]$$

@95,5% (2σ)

$$[\mu^*/(\sigma^*)^2; \mu^* \cdot (\sigma^*)^2]$$

## Estimação dos parâmetros

A melhor maneira de estimar os parâmetros de dados log-normais é através do cálculo dos parâmetros na escala logarítmica.

Por exemplo, seja  $X$  uma variável aleatória com distribuição aproximadamente lognormal. Para estimar o parâmetro  $\mu(X)$ , primeiro estima-se o valor da média da variável  $\ln(X)$ ,  $\bar{x}$ , e depois retorna-se para a estimativa do parâmetro  $\mu$  na escala original através da transformação inversa ( $\mu = \exp(\bar{x})$ ).

## Resultados

No que tange à Engenharia de Avaliações, um dos aspectos mais importantes é o de qual seria a melhor medida de tendência central a se adotar para estimar o valor do bem.

Na tabela 1 mostramos que a adoção da média, moda ou mediana é praticamente irrelevante quando o erro-padrão  $\sigma$  da regressão é pequeno. Porém, para altos valores de  $\sigma$ , percebe-se uma discrepância muito grande entre os valores obtidos com a média, moda ou mediana.

Para efeito de comparação, a tabela 1 mostra os valores da moda e da média em função do desvio-padrão  $\sigma$ , quando a mediana da distribuição tem valor fixado em  $\mu^* = 1.000.000$ .

Tabela 1: Moda e Média em função do desvio-padrão, quando  $\mu = 1.000.000$ . FONTE: Autor.

Medida / $\sigma$	0,1	0,25	1,0	2,0
Moda	990.050	939.413	367.879	18.316
%	-1,0%	-6,1%	-63,2%	-98,2%
Média	1.005.013	1.031.743	1.648.721	7.389.056
%	+0,5%	+3,2%	+64,9%	+638,9%
IC @68,3%	904.837	778.800	367.879	135.335
(1 $\sigma$ )	1.105.170	1.284.025	2.718.281	7.389.056
IC @95,5%	818.731	606.531	135.335	18.316
(2 $\sigma$ )	1.221.403	1.648.721	7.389.056	54,6x10 <sup>6</sup>

Em suma, fixado  $\mu^*$ , à medida que o desvio-padrão aumenta, a moda da distribuição tende a zero, enquanto a média da distribuição tende a  $+\infty$ .

## Outros detalhes

Para saber mais sobre a distribuição lognormal, visitar o [site](#).

## Referências

- [1] LIMPert, E.; STAHEL, W.A.; ABBT, M. Log-normal Distributions across the Sciences, *BioScience*
- [2] J. Doe, J. Smith, Other article name, *Phys. Rev. Lett.*
- [3] [http://droubi.me/dist\\_lognormal/](http://droubi.me/dist_lognormal/)