

## **DISTRIBUIÇÃO LOGNORMAL: PROPRIEDADES E APLICAÇÕES NA ENGENHARIA DE AVALIAÇÕES**

**Luiz Fernando Palin Droubi**

**Ministério do Planejamento, Desenvolvimento e Gestão**

MPDG/SPU/SC

Praça XV de Novembro, 336, Centro, Florianópolis/SC, 88010-400

luiz.droubi@planejamento.gov.br

**Willian Zonato**

**Ministério do Planejamento, Desenvolvimento e Gestão**

MPDG/SPU/SC

Praça XV de Novembro, 336, Centro, Florianópolis/SC, 88010-400

willian.zonato@planejamento.gov.br

**Norberto Hochheim**

Universidade Federal de Santa Catarina

Campus Reitor João David Ferreira Lima, s/n - Trindade, Florianópolis/SC, 88040-900

hochheim@gmail.com

### **Resumo:**

Um procedimento muito comum na área de engenharia de avaliações é a adoção de transformações das variáveis para a obtenção de um modelo de regressão “melhor” ajustado. Dentre tais transformações, a mais usual e preferida de muitos avaliadores é a função logaritmo, especialmente para a variável dependente. Muitas vezes esta transformação é adequada e percebe-se uma notória melhora no ajuste do modelo. Outras vezes, esta transformação pode não ser adequada. Apesar do modelo aparentar-se melhor ajustado, problemas podem ocorrer quanto às verificações das hipóteses clássicas da regressão, as quais nem sempre os avaliadores estão tão atentos quanto estão com as verificações dos intervalos de confiança e níveis de significância. No entanto, o avaliador que assim procede estará verificando intervalos de confiança e níveis de significâncias incorretos, haja vista que a hipótese da heteroscedasticidade implica na incorreção destas inferências. Pretende-se apresentar aos avaliadores a importância da adoção de escolhas criteriosas para as transformações da variáveis através da análise do histograma da variável original e da variável transformada. Normalmente, uma boa escolha de transformação leva a uma distribuição aproximadamente normal. Quando a variável dependente apresenta distribuição lognormal, esta transformação é a transformação logarítmica. Desta maneira, descrevemos as características básicas desta distribuição, sua formulação, características além do seu relacionamento com a distribuição normal. Por fim, o presente estudo evidencia as implicações da adoção da transformação da variável dependente e o problema da retransformação da variável dependente à sua escala original.

**Palavras-chave:** distribuição lognormal, regressão-linear, heterocedasticidade, Box-Cox

## Abstract

A very common procedure in this field is the adoption of transformations to obtain a better regression model. The most usual and preferred of many transformations is the logarithm function. Often this transformation is adequate and there is a noticeable improvement in the fit of the model. At other times, this transformation may not be adequate: although the model appears to be better adjusted, problems with the verifications of the classical linear regression hypotheses may appear, which are often neglected by the appraiser. However, if he proceeds without verifying heteroscedasticity, he might be verifying confidence intervals and incorrect significance levels, since this hypothesis implies the incorrectness of these inferences. We understand that the best way to present to the appraisers the importance of careful choices of transformations is by analyzing the histogram of the original and transformed variable. Typically, a good choice of transformation leads to a transformed variable distributed closely to a normal distribution. When the dependent variable presents lognormal distribution, this transformation is the logarithmic transformation. So we describe the basic characteristics of this distribution, its formulation, characteristics and its relationship with the normal distribution. Finally, we describe other ways to deal with heteroscedastic data, the implications of adopting the transformation of the dependent variable and address the problem of the retransformation of the dependent variable to its original scale.

**Keywords:** Lognormal distribution, linear-regression, heteroscedasticity, Box-Cox

## 1. INTRODUÇÃO

A transformação de variáveis é um procedimento comum na Engenharia de Avaliações. No entanto, a transformação dos dados por vezes é realizada sem uma análise profunda do comportamento das variáveis. O conceito da navalha de Occam aplicado a estatística pode ser assim resumido: “modelos mais simples deveriam ser preferidos até que os dados justifiquem modelos mais complexos” (ver Balasubramanian, p.1). De fato, a *Food and Drug Administration* (FDA), órgão federal dos EUA que atua no controle da comercialização de alimentos e medicamentos no país, recomenda:

“A transformação desnecessária de dados deve ser evitada. Caso tenha sido realizada transformação de dados, uma justificativa para a escolha da transformação junto com a interpretação das estimativas dos efeitos do tratamento com base nos dados transformados deve ser fornecida.” (FDA, 1988 apud KEENE (1985))

No entanto, a transformação logarítmica é especial, por uma série de aspectos, como pode ser visto em KEENE (1985). A distribuição lognormal apresenta diversas aplicações práticas. É comum, na área de avaliação de imóveis, mas não apenas, nos depararmos com dados que seguem esta distribuição. Neste artigo pretende-se demonstrar as principais características da distribuição lognormal, sua relação com a distribuição normal de Gauss, assim como debater a melhor maneira de se lidar com dados lognormais.

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1 Definição e Formulação

Diz-se que uma variável aleatória  $x$  tem distribuição lognormal se o seu logaritmo  $\ln(x)$  tem distribuição normal. A formulação da distribuição lognormal para os parâmetros  $\mu$  e  $\sigma$  pode ser vista na equação (1) (LIMPERT; STAHEL; ABBT, 2001, p. 345).

$$\begin{cases} f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2\right] & \text{para } x > 0 \\ 0 & \text{caso contrário} \end{cases} \quad (1)$$

### 2.2 Propriedades

#### 2.2.1 Valor Esperado e Variância

O valor Esperado  $E$  de uma variável aleatória com distribuição lognormal  $X$  pode ser calculado como visto na equação (2), enquanto a sua variância  $Var(X)$  é calculada de acordo com a equação (3) (LIMPERT; STAHEL; ABBT, 2001, p. 345).. Substituindo (2) em (3) chegamos na equação (4).

$$E[X] = e^{\left(\mu + \frac{\sigma^2}{2}\right)} \quad (2)$$

$$Var(X) = e^{(2\mu + \sigma^2)}(e^{\sigma^2} - 1) \quad (3)$$

$$Var(X) = E^2[X] \cdot (e^{\sigma^2} - 1) \quad (4)$$

### 2.2.2 Medidas de Tendência Central

Ao contrário da distribuição normal de Gauss e outras distribuições de probabilidade simétricas, em que as medidas de tendência central convergem para um mesmo valor, na distribuição lognormal, moda, média e mediana apresentam localizações distintas. A figura 1 mostra a posição das medidas de tendência central (moda, média e mediana) para um variável aleatória de distribuição lognormal.

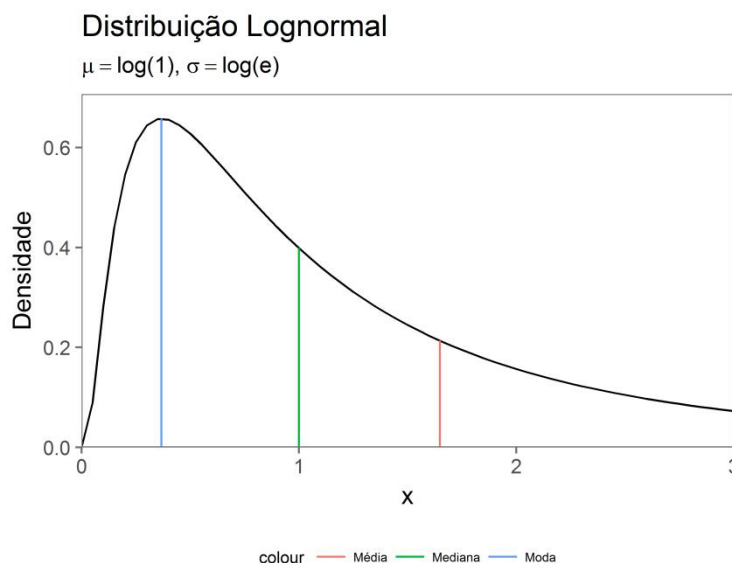


Figura 1 - Posições de medidas de tendência central numa distribuição lognormal  
Fonte: Autor

### 2.2.3 Efeito das variações do desvio-padrão na forma da distribuição

Através da análise da figura 2 percebemos que, apesar das mudanças de forma decorrentes dos diversos fatores de  $\sigma$ , as características da distribuição lognormal são sempre as mesmas: tem início na origem do sistema de coordenadas, ou seja, a distribuição é truncada do seu lado esquerdo. Depois apresenta um pico (moda) que, considerando-se  $\mu$  constante, tem o valor da sua abcissa cada vez menor. Finalmente, a distribuição lognormal sempre apresenta uma cauda à direita, tendendo a zero quando o valor das abcissas tende a  $+\infty$ .

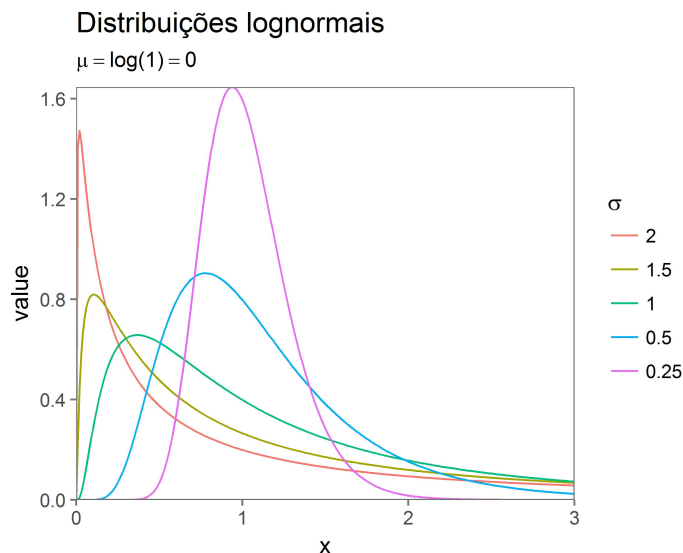


Figura 2 - Distribuição lognormal com  $\mu = 0$  e diversos valores de desvio-padrão  
Fonte: Autor

#### 2.2.4 Relação com a distribuição normal

Lembrando que a função densidade de probabilidade de uma variável aleatória com distribuição normal é dada de acordo com a equação (5) (MATLOFF, 2009, p.159):

$$f_w(t) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(t - \mu)^2}{\sigma^2}\right) \quad (5)$$

É fácil mostrar que a distribuição normal-padrão ( $N(0,1)$ ) apresenta a função densidade de probabilidade descrita na equação (6), obtida com a substituição dos parâmetros  $\mu = 0$  e  $\sigma = 1$ :

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} t^2\right) \quad (6)$$

Seja  $X$  uma variável aleatória de distribuição normal padronizada ( $X \sim N(0,1)$ ),  $f_X$  a função densidade de probabilidade e  $Y = e^X$ . Então  $(F_Y)$  é igual a:

$$\begin{aligned} F_Y(y) &= P(e^X \leq y) \\ F_Y(y) &= P(X \leq \ln(y)) \\ F_Y(y) &= \int_{-\infty}^{\ln(y)} f_X(x) dx \\ F_Y(y) &= \int_{-\infty}^{\ln(y)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \end{aligned} \quad (7)$$

o que equivale a:

$$F_Y(y) = \int_0^y \frac{1}{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\ln(x)^2}{2}\right) dx \quad (8)$$

Ou seja, a distribuição de uma variável  $Y = e^X$ , em que  $X \sim N(0,1)$  é equivalente a distribuição de uma variável lognormal com parâmetros  $\mu = 0$  e  $\sigma = 1$ .

A figura 3 ilustra este fato.

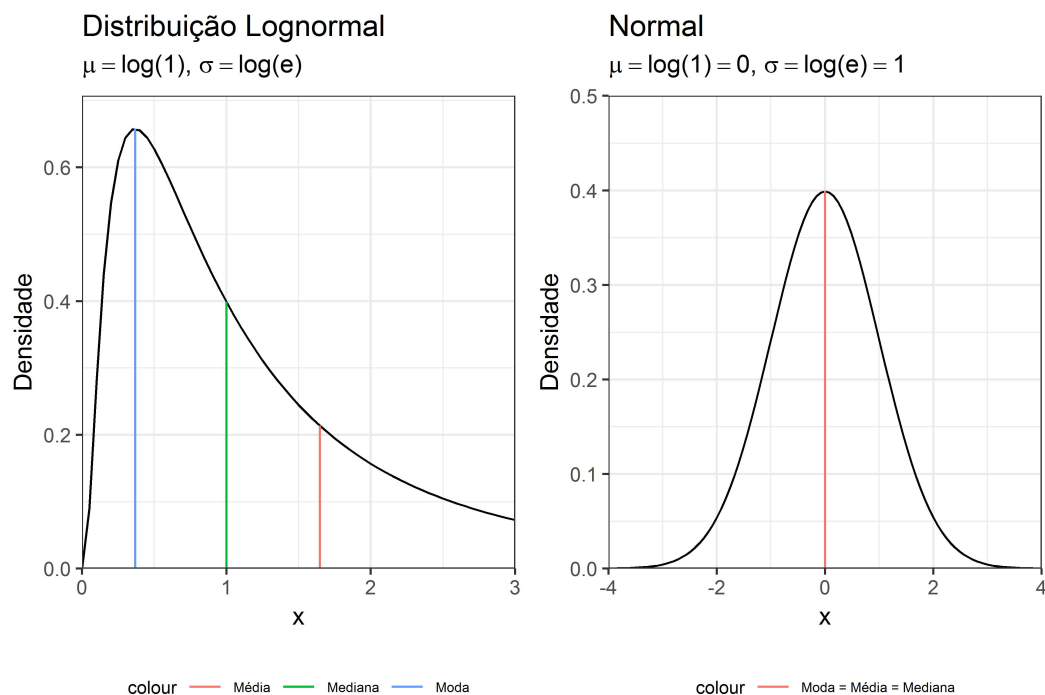


Figura 3 - Comparação entre distribuições normal e lognormal padronizadas.  
Fonte: Autor

## 2.2.5 Analogia com o Teorema do Limite Central

Segundo Matloff (2009, p.158-159), o teorema do limite central diz, resumidamente, que uma variável aleatória que é a soma de outras variáveis aleatórias independentes tem distribuição aproximadamente normal. Quanto maior for o número das variáveis somadas, mais próximo da distribuição normal.

Analogamente, assim como o resultado da soma de diversas variáveis independentes resulta numa variável aleatória de distribuição normal, o *produto* de diversas variáveis aleatórias resulta numa distribuição lognormal.

## 2.2.6 Transformação de variável e Homoscedasticidade

De acordo com Matloff (2017, p. 138), se uma variável aleatória  $W$  é aproximadamente normal, com baixo coeficiente de variação ( $CV = \sigma / \mu$ ), e  $g(W)$  é uma função suave, então a nova variável também será aproximadamente normal, com média  $g(EW)$  e variância:

$$[g'(EW)]^2 \text{Var}(W) \quad (9)$$

Assumindo que os erros de uma função de regressão sejam heteroscedásticos, seguindo uma função conhecida  $\sigma(t) = \mu(t)$ , se aplicarmos a função logaritmo natural à variável dependente, lembrando que a derivada da função logaritmo natural é  $\frac{d}{dt} \ln(t) = \frac{1}{t}$  e que  $\text{Var}(W) = \sigma^2(W)$ , segundo a equação 9, teremos:

$$\frac{1}{\mu^2(t)} \mu^2(t) = 1 \quad (10)$$

Ou seja, o uso da transformação logaritmo natural, para este caso em particular, conduz à homoscedasticidade do modelo.

$$\frac{d}{dt} \sqrt{t} = \frac{0,5}{\sqrt{t}} \rightarrow \text{Var}(\sqrt{W}) = \left( \frac{0,5}{\sqrt{t}} \right)^2 (\sqrt{t})^2 = 0,25 \quad (11)$$

De acordo com Matloff (2017, p. 138), ainda, se  $\sigma(t) = \sqrt{\mu(t)}$ , a transformação raiz-quadrada é que traria de volta a homoscedasticidade.

### 3. EXEMPLO

#### 3.1 Dados

A título de exemplo de aplicação da distribuição lognormal, serão utilizados os dados oriundos de Hochheim (2015, pp. 21–22). A variável VT possui distribuição aproximadamente lognormal, como pode ser visto nas figuras 4 a 6.

Além da distribuição lognormal, foram testadas a aderência de outras distribuições de probabilidade com forma similar à distribuição dos dados, tais como a distribuição de Weibull e a distribuição gama. Para a distribuição lognormal, foram utilizados os estimadores do método dos momentos (MME) e da máxima verossimilhança (MLE).

Foi constatado que a distribuição que melhor se ajusta aos dados foi a distribuição lognormal.

Percebe-se ainda nas figuras 4 a 6 que outras distribuições (gama e Weibull) inicialmente crescem mais rapidamente e tem pico mais achatado que os dados empíricos e as distribuições log-normais ajustadas.

#### 3.2 Ajuste de distribuições aos dados

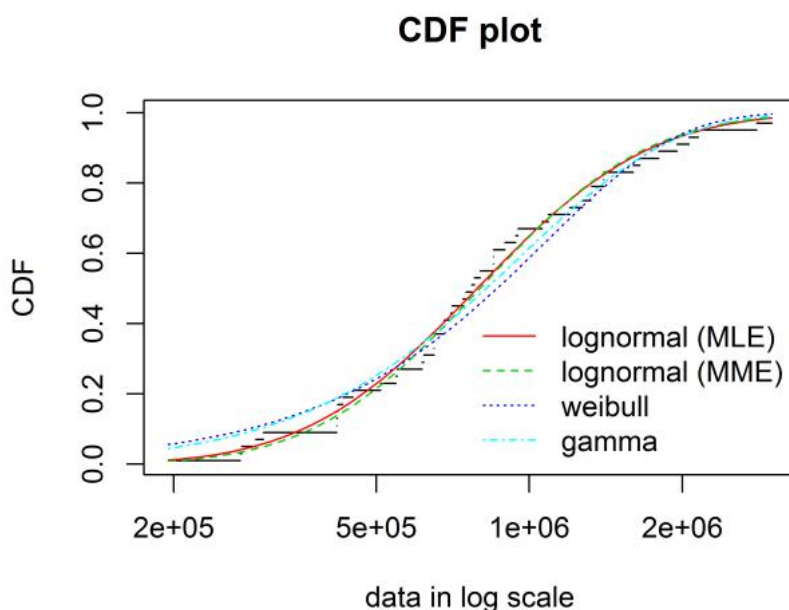


Figura 4 - Função densidade acumulada para os dados e para as distribuições testadas  
Fonte: Autor



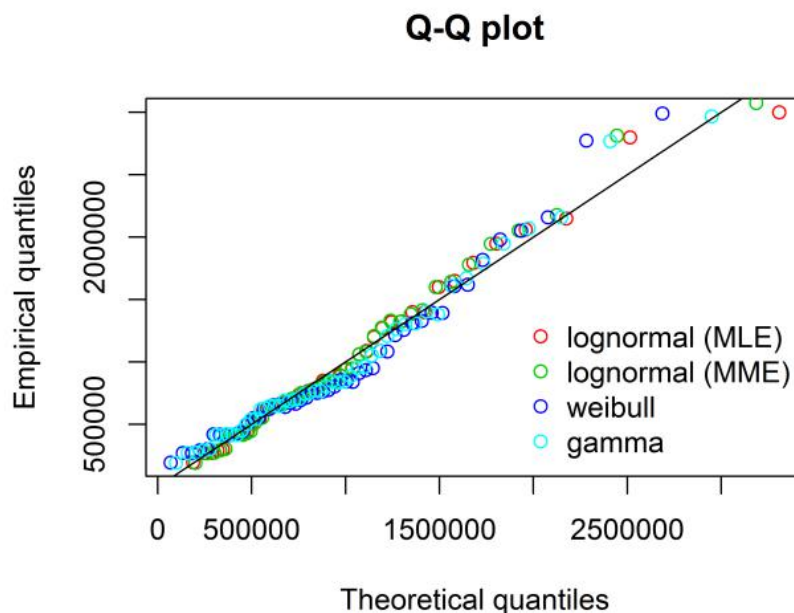


Figura 5 - Gráfico QQ para as diversas distribuições ajustadas  
Fonte: Autor

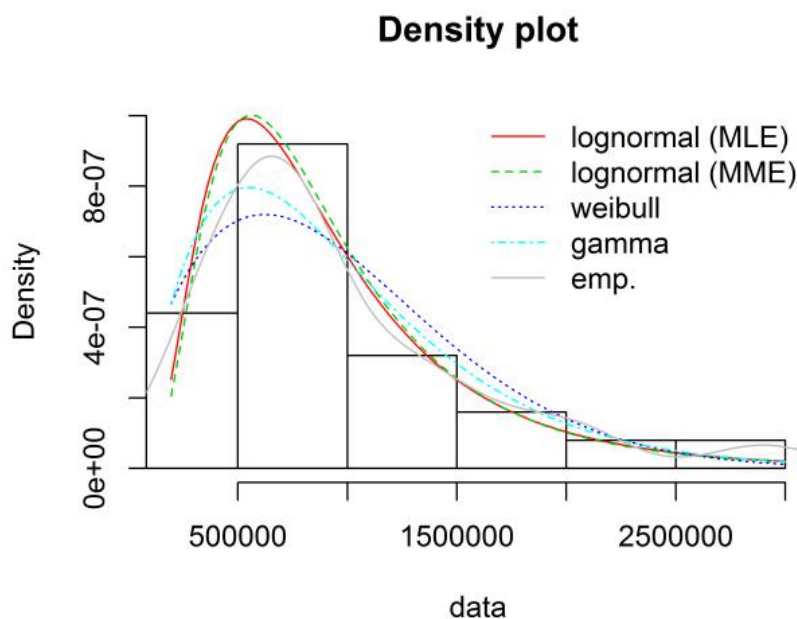


Figura 6 - Ajuste da distribuição empírica para funções de probabilidade conhecidas.  
Fonte: Autor

### 3.3 Gráficos

As figuras 7 e 8 mostram a distribuição lognormal ajustada com os parâmetros obtidos em HOCHHEIM (2015, pp. 21–22), ou seja, para o parâmetro  $\mu$  foi adotado o valor  $\mu = \overline{\ln(VT)}$  e para o parâmetro  $\sigma$  foi adotado o valor  $\sigma = \text{desvio}(\ln(VT))$ .

A figura 7 mostra o gráfico da função densidade de probabilidade (FDP) construídos com os parâmetros  $\mu$  e  $\sigma$  obtidos da variável *valor*.

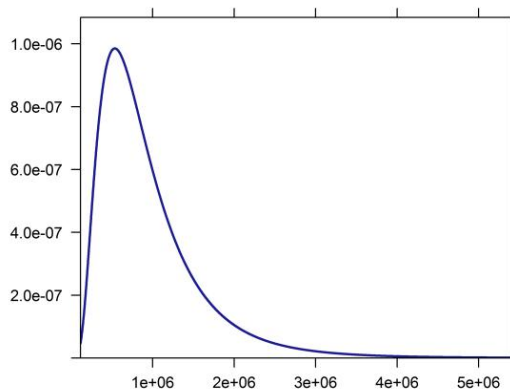


Figura 7 -FDP com parâmetros obtidos dos dados da variável *valor*

Fonte: Autor

A figura 8 mostra o histograma dos dados da variável *valor*, superposto com a curva da função densidade de probabilidade (FDP) da figura 7.

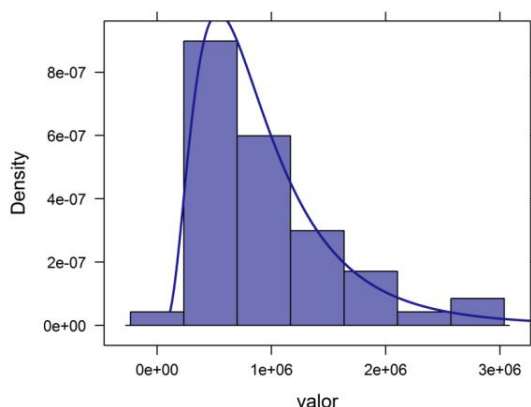


Figura 8 - Histograma das variável *valor* com FDP superposta

Fonte: Autor

A figura 9 ilustra o gráfico da função cumulativa de probabilidade teórica (FCP), em vermelho, para a variável *valor*. A curva azul representa a curva de distribuição cumulativa de probabilidades empírica para a mesma variável.

A figura 10 mostra a distribuição da variável  $\ln(VT)$ . Pode-se notar que, conforme esperado, já que a distribuição da variável *VT* é aproximadamente lognormal, seu logaritmo tem distribuição aproximadamente normal.

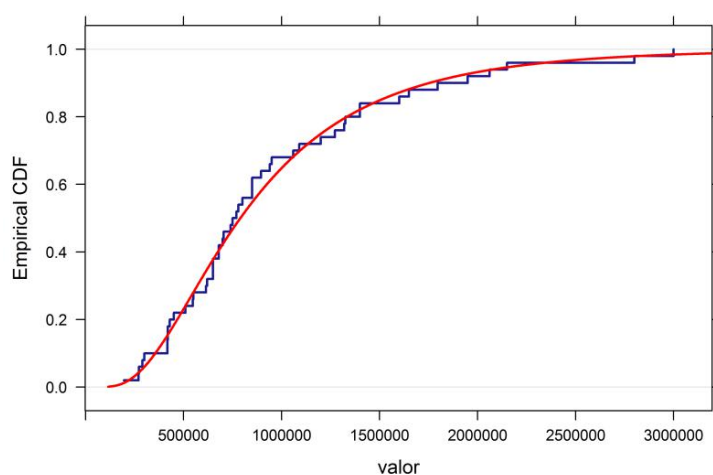
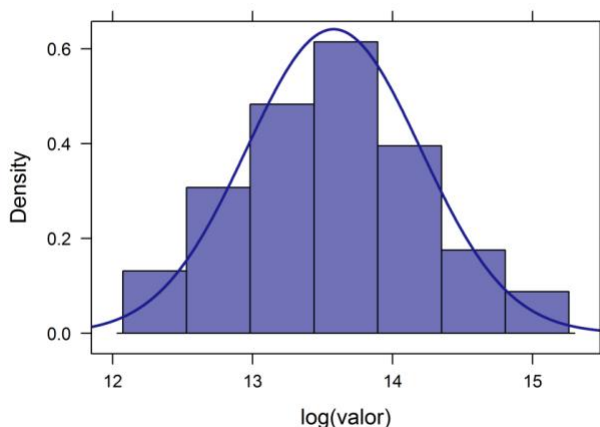


Figura 9 - FCP teórica (azul) e empírica (vermelho)

Fonte: Autor

Histograma de  $\ln(\text{valor})$



Com pdf  $N(\text{mean}(\ln(\text{valor})), \text{sd}(\ln(\text{valor})))$  superposta

Figura 10 - Histograma com função densidade de probabilidade normal superposta

Fonte: Autor

### 3.4 - Modelos

Detectando-se a presença de variável resposta com distribuição lognormal, pode-se proceder da seguinte maneira:

- Proceder com a transformação da variável resposta pela função logarítmica;
- Proceder com a variável na escala original, corrigindo posteriormente a heteroscedasticidade com o método de Eicker-White;
- Proceder com o método dos mínimos quadrados ponderados.

É fácil mostrar que o modelo linear com a variável resposta logaritmizada, ou seja, com distribuição normal, é melhor ajustado que o modelo linear de uma variável resposta lognormal. Na tabela 1, no entanto, mostra-se que, para o presente caso, esta melhora de ajuste é modesta, próxima a 4,5%.

No entanto, o modelo sem transformação, segundo o teste de Breusch-Pagan, apresenta heteroscedasticidade, enquanto no modelo transformado pela função logaritmo natural, esta heteroscedasticidade desaparece.

A função máxima verossimilhança de Box-Cox também apresenta como transformação ótima a transformação logarítmica, como demonstra a figura 11.

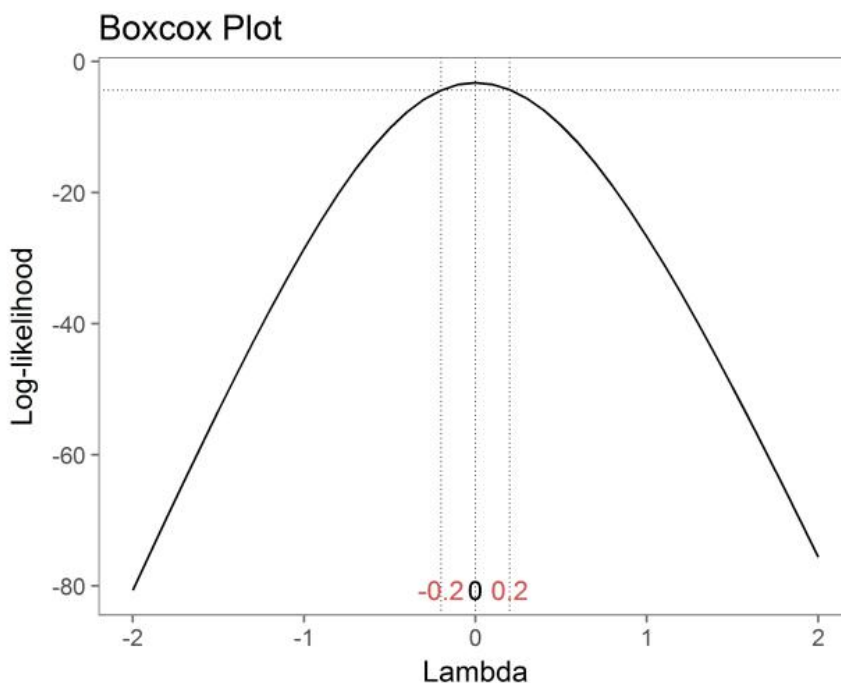


Figura 11 - Gráfico da função verossimilhança de Box-Cox  
Fonte: Autor

Tabela 1- Comparação entre modelos com e sem transformação da variável resposta.

	<i>Dependent variable:</i>	
	valor (1)	log(valor) (2)
area_total	2,893.178 (2,065.405, 3,720.951) t = 6.850 p = 0.00000***	0.002 (0.001, 0.002) t = 4.886 p = 0.00002***
quartos	73,524.375 (-34,814.143, 181,862.894) t = 1.330 p = 0.191	0.169 (0.084, 0.255) t = 3.870 p = 0.0004* * *
suítes	111,000.591 (8,045.131, 213,956.052) t = 2.113 p = 0.041**	0.088 (0.007, 0.170) t = 2.121 p = 0.040**
garagens	148,427.448 (49,657.102, 247,197.795) t = 2.945 p = 0.006***	0.175 (0.097, 0.253) t = 4.394 p = 0.0001***
dist_b_mar	-223.217 (-434.862, -11.571) t = -2.067 p = 0.045**	-0.0003 (-0.0004, -0.0001) t = -3.215 p = 0.003***
padraomedio	-146,549.393 (-354,850.457, 61,751.672) t = -1.379 p = 0.176	0.268 (0.103, 0.433) t = 3.190 p = 0.003* * *
padraoalto	-56,064.550 (-264,003.525, 151,874.426) t = -0.528 p = 0.600	0.334 (0.169, 0.498) t = 3.975 p = 0.0003* * *
Constant	33,953.788 (-267,469.800, 335,377.375) t = 0.221 p = 0.827	12.315 (12.076, 12.553) t = 101.170 p = 0.000***
Observations	50	50
R <sup>2</sup>	0.906	0.940
Adjusted R <sup>2</sup>	0.890	0.930
Akaike Inf. Crit.	1,375.659	-29.275
Residual Std. Error (df = 42)	207.903.003	0.165
F Statistic (df = 7; 42)	57.731***	94.063***

Note:

\*p<0.1; \*\* p<0.05; \*\*\* p<0.01

Fonte: Autor

### 3.4.1 Modelo linear com a variável resposta transformada

#### 3.4.1.1 Retransformação de variáveis

O problema da transformação da variável resposta no logaritmo da variável resposta original, é que devemos estudar como proceder na retransformação da variável, para efetuar a avaliação do imóvel.

#### 3.4.1.2 A desigualdade de Jensen

Segundo Matloff (2017, p. 142), a desigualdade de Jensen (aplicada à estatística) se traduz na expressão seguinte (12), válida para funções convexas:

$$E[h(V)] \geq h(E[V]) \quad (12)$$

Aplicando-se a equação (12) à função logarítmica, que é uma função côncava, obtém-se a desigualdade expressa na equação (13) (MATLOFF, 2017, p. 142):

$$E[\ln Y \mid X = t] \leq \ln(E[Y \mid X = t]) \quad (13)$$

Para Matloff, então, como a igualdade só irá acontecer em poucos casos especiais, a função de regressão de  $\ln(Y)$  será quase sempre menor do que o logaritmo natural da função de regressão de  $Y$ , de tal forma que a suposição que dado uma variável aleatória  $Y$  tal que assumimos que  $E(Y \mid X = t) = e^{\beta_0 + \beta_1 t}$ , não podemos concluir de imediato que um modelo linear razoável seria da forma  $E(\ln Y \mid X = t) = \beta_0 + \beta_1 t$ , pois, pela desigualdade de Jensen, se temos dados significativamente heteroscedásticos da variável original ( $Y$ ), a discrepância entre os dois lados da desigualdade acima poderia variar bastante com  $t$ , potencialmente produzindo uma grande distorção à forma da curva de regressão (MATLOFF, 2017, p. 143).

Segundo Becker (2012, p. 4), a desigualdade de Jensen pode ser transformada numa igualdade do tipo expressa na equação (14):

$$E[f(X)] = f(E[X]) + \Delta \quad (14)$$

E, de acordo com o mesmo (2012, p. 5), o valor de, chamado de Defeito de Holder, é proporcional à variância da variável aleatória, tal que se é duas vezes continuamente diferenciável e existem limites finitos e tais que para todo, então existe um valor para o qual a equação (15) abaixo é válida:

$$E[f(X)] - f(E[X]) = \frac{1}{2} \mu \text{Var}(X) \quad (15)$$

Em suma, o valor de  $\Delta$  é proporcional à variância de  $X$  ( $\Delta \propto \sigma^2(X)$ ).

Deste modo, existem na literatura diversos estudos sobre qual seria o “melhor” estimador – paramétrico ou não-paramétrico – para a variável resposta original, quando da ocorrência da transformação da variável pela função logaritmo natural, como pode ser visto em DUAN (1983), MEULENBERG (1965) e SHEN; ZHU (2008).

De acordo com Shen e Zhu (2008), com a simples aplicação da transformação inversa (exponencial) à aplicada na variável dependente (logarítmica), chegamos ao *Back-Transform* (BT) *Estimator*, que tem performance “muito pior do que os outros estimadores” (2008, p. 554). De fato, o estimador BT seria mais apropriado para estimar a mediana, no entanto a equação de regressão linear é uma equação para a média da variável dependente. Métodos de regressão à mediana (KOENKER; HALLOCK, 2001), então, seriam mais apropriados para este fim.

Se a distribuição dos erros do modelo for aproximadamente normal, pode-se considerar suficiente a adoção do estimador teórico da equação (16) (DUAN, p. 607-608). Porém, destaca-se que, em muitos casos, os estimadores não-paramétricos podem apresentar-se mais eficientes que o estimador teórico (ver DUAN, pp. 607-609).

$$E[Y | X] = \exp(\beta_0 + X\beta_1 + 0,5\sigma^2) \quad (16)$$

### 3.4.1 Modelo linear com erros robustos

Uma vez detectada a heteroscedasticidade, uma outra possibilidade é permanecer com os valores da variável original e proceder com o cálculo de erros robustos pelo método de Eicker-White. Computados os erros robustos, podem ser calculados os intervalos de confiança para os regressores e os novos p-valores dos testes de significância. A tabela 2 mostra o cálculo destes valores calculados segundo método de Eicker-White. Percebe-se que os p-valores são, em regra, menores para o modelo com erros robustos, com exceção para o p-valor da variável *area\_total*.

Problemas quanto à significância da variável *padrão* podem também ser notados na tabela 2. Como observamos na tabela 1, no entanto, a variável apresentou-se significativa quando da transformação da variável dependente. Possivelmente com uma melhor análise de regressão (*outliers*, pontos influenciantes, multicolinearidade e outras verificações).

### 3.4.2 Mínimos quadrados ponderados

De acordo com Matloff (2017, p.139), em princípio, o Método dos Mínimos Quadrados ponderados (MQP) fornece melhores estimativas para os coeficientes e inferência estatística correta mesmo na presença de heteroscedasticidade.



Tal método consiste, analogamente ao Método dos Mínimos Quadrados Ordinários (MQO), em uma minimização. No caso dos MQO, o processo de minimização consiste em minimizar-se a quantidade descrita na equação (17) (MATLOFF, p. 69):

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{X}_i' b)^2 \quad (17)$$

Enquanto no método MQP, minimiza-se a quantidade expressa pela equação (18) (MATLOFF, p. 133):

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{w_i} (Y_i - \tilde{X}_i' b)^2 \quad (18)$$

Onde  $w_i = \sigma^2(X_i)$ , ou seja, são atribuídos pesos a cada observação, de acordo com os erros computados no modelo original.

#### 4 CONCLUSÃO

Foi possível demonstrar de maneira gráfica que os dados da variável *valor* apresentados se ajustam bem a uma distribuição lognormal equivalente. Por definição, então, o logaritmo da variável possui distribuição normal.

O uso da variável explicada sem transformação adequada conduziu a um modelo com resíduos heterocedásticos. Neste caso, uma alternativa seria a aplicação da regressão linear ponderada (ou mínimos quadrados ponderados). No entanto, o método de Eicker-White mostrou-se adequado e sua aplicação é menos complexa do que o método de regressão ponderada.

Já a transformação da variável dependente pela função logarítmica também mostrou-se adequada. Porém, neste caso, tem-se o problema da retransformação da variável à escala original, assunto que ainda não foi pacificado na engenharia de avaliações, onde a escolha do estimador fica a critério do avaliador.



Tabela 2: Comparação entre modelos com e sem erros robustos

	<i>Dependent variable:</i>	
	default (1)	robust (2)
area_total	2.893,178 (2.065,405, 3.720,951) t = 6,850 p = 0,0000***	2.893,178 (1.712,055, 4.074,302) t = 4,801 p = 0,0000***
quartos	73.524,375 (-34.814,143, 181.862,894) t = 1,330 p = 0,191	73.524,375 (-31.897,755, 178.946,506) t = 1,367 p = 0,172
suítes	111.000,591 (8.045,131, 213.956,052) t = 2,113 p = 0,041**	111.000,591 (18.671,249, 203.329,934) t = 2,356 p = 0,019**
garagens	148.427,448 (49.657,102, 247.197,795) t = 2,945 p = 0,006***	148.427,448 (73.906,905, 222.947,992) t = 3,904 p = 0,0001***
dist_b_mar	-223,217 (-434,862, -11,571) t = -2,067 p = 0,045**	-223,217 (-406,588, -39,845) t = -2,386 p = 0,018**
padraomedio	-146.549,393 (-354.850,457, 61.751,672) t = -1,379 p = 0,176	-146.549,393 (-322.293,226, 29.194,441) t = -1,634 p = 0,103
padraoalto	-56.064,550 (-264.003,525, 151.874,426) t = -0,528 p = 0,600	-56.064,550 (-197.646,893, 85.517,794) t = -0,776 p = 0,438
Constant	33.953,788 (-267.469,800, 335.377,375) t = 0,221 p = 0,827	33.953,788 (-191.823,586, 259.731,161) t = 0,295 p = 0,769
Observations	50	50
R <sup>2</sup>	0,906	0,906
Adjusted R <sup>2</sup>	0,890	0,890
Akaike Inf. Crit.	1.375,659	1.375,659
Residual Std. Error (df = 42)	207.903,003	207.903,003
F Statistic (df = 7; 42)	57,731***	57,731***

Note:

\*p<0,1; \*\* p<0,05; \*\*\* p<0,01

Fonte: Autor

## Referências Bibliográficas

BALASUBRAMANIAN, V. Occam's Razor and Statistical Mechanics on The Space of Probability distributions. **eprint arXiv:cond-mat/9601030**, 1996.

BECKER, R. A. The variance drain and Jensen's inequality. **CAEPR Working Paper**, 2012. Disponível em: <<http://dx.doi.org/10.2139/ssrn.2027471>>..

DUAN, N. Smearing estimate: A nonparametric retransformation method. **Journal of the American Statistical Association**, v. 78, n. 383, p. 605–610, 1983. Taylor & Francis. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/01621459.1983.10478017>>..

FDA. **Guideline for the format and content of the clinical and statistical sections of new drug applications**. Food and Drug Administration, Public Health Service, US Department of Health and Human Services, 1988.

HOCHHEIM, N. **Engenharia de avaliações - módulo básico**. Florianópolis: IBAPE - SC, 2015.

KEENE, O. N. The log transformation is special. **Statistics in Medicine**, v. 14, p. 811–819, 1985.

KOENKER, R.; HALLOCK, K. F. Quantile regression. **Journal of Economic Perspectives**, v. 15, n.4, p. 143–156, 2001

LIMPERT, E.; STAHEL, W. A.; ABBT, M. Log-normal Distributions across the Sciences: Keys and Clues. **BioScience**, v. 51, n.5, p. 341–352, 2001

MATLOFF, N. S. **From algorithms to z-scores: Probabilistic and statistical modeling in computer science**. Davis, California: Orange Grove Books, 2009.

MATLOFF, N. **Statistical regression and classification: From linear models to machine learning**. Boca Raton, Florida: Chapman & Hall, 2017.

MEULENBERG, M. T. G. On the estimation of an exponential function. **Econometrica**, v. 33, n. 4, p. 863–868, 1965. [Wiley, Econometric Society]. Disponível em: <<http://www.jstor.org/stable/1910362>>..

SHEN, H.; ZHU, Z. Efficient mean estimation in log-normal linear models. **Journal of Statistical Planning and Inference**, v. 138, p. 552–567, 2008. Elsevier. Disponível em: <<https://www.unc.edu/~haipeng/publication/emplnM1.pdf>>..