BOX COX: O uso da técnica da transformação

Joyce de Souza Sforzin

Licenciatura em Matemática

Instituto de Ciências Exatas e Tecnológicas - IEF

joyce.sforzin@ufv.br

CAMPUS UFV - FLORESTAL

Introdução

A transformação de Box-Cox recebeu o nome dos estatísticos que a formularam, George E. P. Box y David Cox. George Edward Pelham Box, foi um estatístico britânico muito conhecido por ser "uma das grandes mentes estatísticas do século XX".

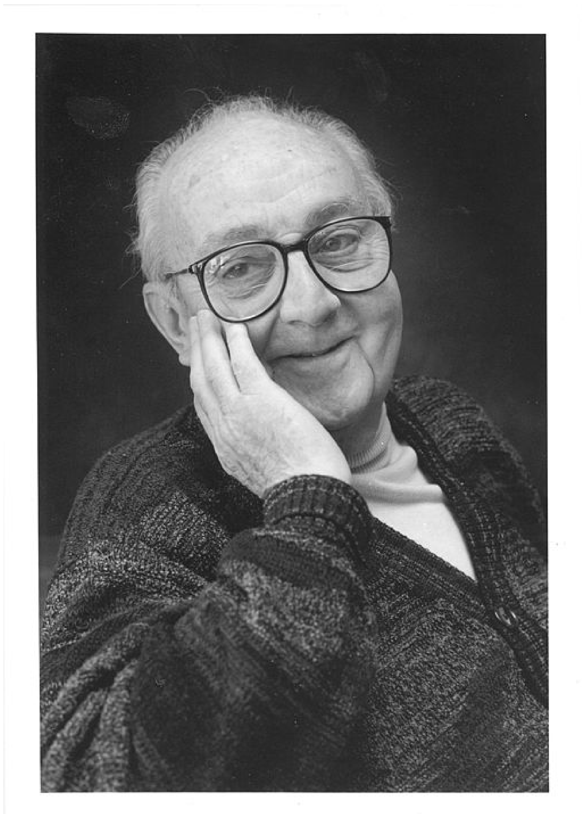


Figura 1: *George Box*

Trabalhou junto de David Roxbee Cox também um estatístico britânico de destaque. Ambos crianram o artigo chamado : "An analysis of transformations", no Journal of the Royal Statistical Society, Series B.

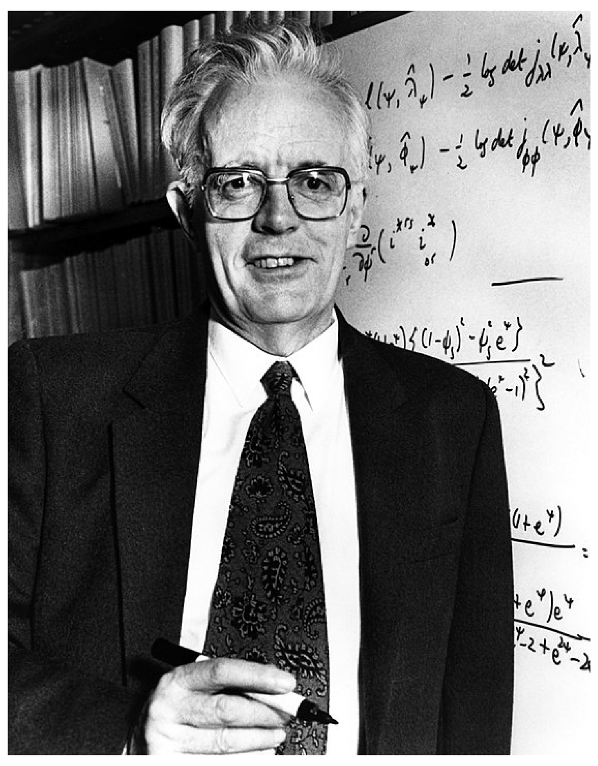


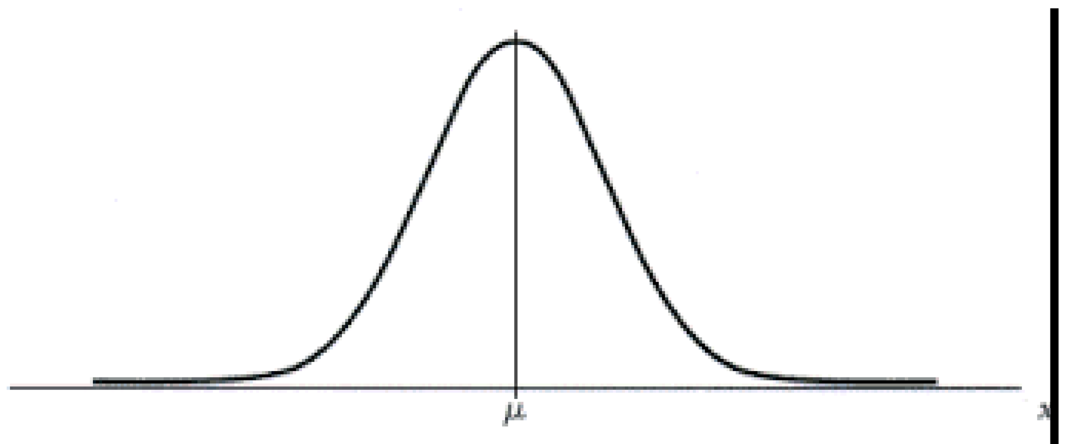
Figura 2: *David Cox*

Objetivos

A distribuição normal conhecida também como distribuição gaussiana é sem dúvida a mais importante distribuição contínua. Sua importância se deve a vários fatores. Além disso diversos estudos práticos tem como resultado uma distribuição normal. Podemos citar como exemplo a altura de uma determinada população em geral segue uma distribuição normal.Entre outras características físicas e sociais tem um comportamento gaussiano, ou seja, segue uma distribuição normal.Uma variável aleatória contínua X tem distribuição Normal se sua função densidade de probabilidade for dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], x \in (-\infty, \infty) \tag{1}$$

O grafico de distribuição é normal quando tem a forma de "sino".



Embora as distribuições de muitos processos possam assumir uma variedade de formas, muitas variáveis observadas possuem uma distribuição de frequências que é, aproximadamente, uma distribuição de probabilidade Normal. Porem quando a distribuição normal não se adéqua aos dados, muitas vezes é útil aplicar a transformação de Box-Cox para obtermos a normalidade. Considerando X1, ..., Xn os dados originais, a transformação de Box-Cox consiste em encontrar um λ tal que os dados transformados Y1, ..., Yn se aproximem de uma distribuição normal.

$$Y_i(\lambda) = \begin{cases} ln(X_i), se\lambda = 0 \\ \dfrac{X_i^\lambda - 1}{\lambda}, se\lambda \neq 0 \end{cases} \tag{2}$$

onde $i = 0, 1, 2, 3, ..., n-1.$

Como você pode ver, esta transformação tem apenas um parâmetro - lambda. Se o valor de lambda é igual a zero, a transformação logarítmica da sequência inicial é realizada, no caso em que o valor de lambda difere de zero, a transformação é por lei exponencial. Se o parâmetro lambda for igual a um, a lei de distribuição da sequência inicial permanece inalterada, embora a sequência desloque, como unidade, é subtraída de cada um dos seus valores.

Aplicação

Dependendo do valor de lambda, a transformação Box-Cox inclui os seguintes casos especiais:

$$\begin{aligned} \lambda = -1.0, &\quad x_i(\lambda) = \frac{1}{x_i} \\ \lambda = -0.5, &\quad x_i(\lambda) = \frac{1}{\sqrt{x_i}} \\ \lambda = \ \ 0.0, &\quad x_i(\lambda) = ln(x_i) \\ \lambda = \ \ 0.5, &\quad x_i(\lambda) = \sqrt{x_i} \\ \lambda = \ \ 1.0, &\quad x_i(\lambda) = x_(i)^2 \end{aligned} \tag{3}$$

A utilização da transformação Box-Cox exige que todos os valores da sequência de entrada sejam positivos e diferentes de zero. Se a sequência de entrada não atende a esses requisitos, ela pode ser movida para a área positiva, pelo volume que garante a "positividade"de todos os seus valores.O grafico a seguir mostra as curvas de transformação Box-Cox, com valores diferentes do parâmetro lambda. A grade horizontal no gráfico é dada em uma escala logarítmica.

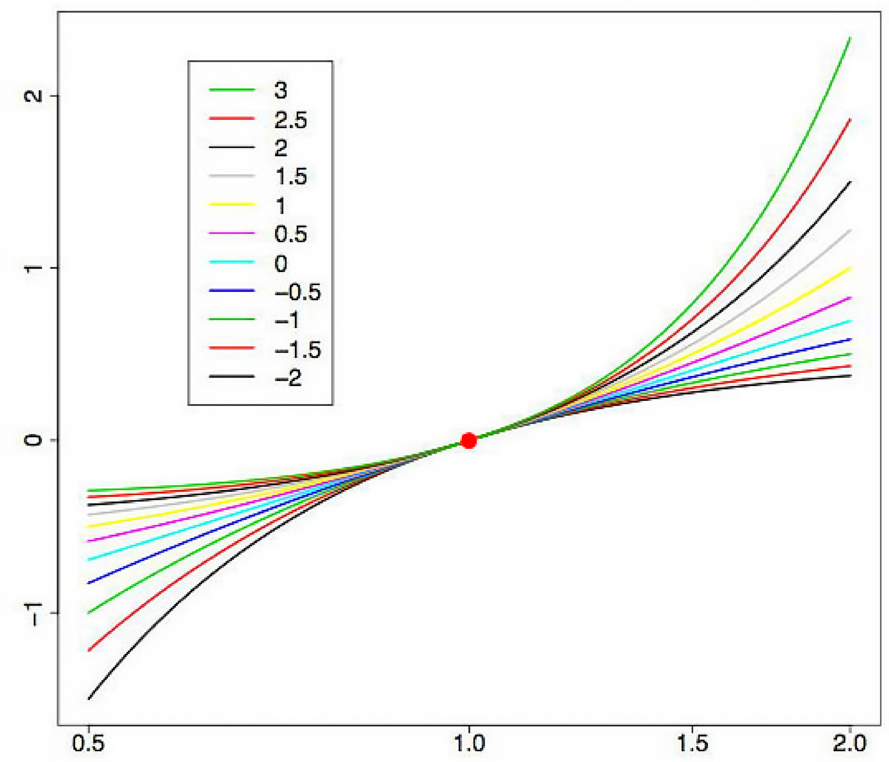


Figura 4: *Curvas de transformação Box-Cox*

Considerações finais

O parâmetro λ é estimado por uma técnica gráfica ou pelo método de máxima verossimilhança . Infelizmente, a forma fechada para o estimador $\hat{\lambda}$ pode ser raramente encontrada. Portanto, o gráfico da probabilidade máxima contra λ é útil. O valor $\hat{\lambda}$ obtido dessa maneira é tratado como se fosse um valor verdadeiro e, então, ajusta-se o modelo aos dados transformados. Essa abordagem pode ser facilmente realizada, e uma teoria assintótica associada a outros parâmetros é útil. Este tratamento tem, no entanto, algumas dificuldades, porque $\hat{\lambda}$ tem uma variabilidade e depende dos dados fornecidos. Sabe-se que a estimativa λ por testes de máxima verossimilhança e razão de verossimilhança relacionada pode ser fortemente influenciada por valores discrepantes. Além disso, em certas situações, a teoria limitativa usual baseada no conhecimento λ não se aplica no caso desconhecido. Portanto, vários procedimentos de estimativa robustos foram propostos.

Na literatura, as transformações de Box-Cox são aplicadas a distribuições básicas, por exemplo, a transformação de raiz cúbica de variáveis qui-quadrado é usada para acelerar a normalidade, e a transformação de raiz quadrada estabiliza as variações das distribuições de Poisson. As transformações de Box – Cox também são aplicadas para vincular funções em modelos lineares generalizados. As transformações visam principalmente a linearidade dos efeitos das covariáveis.

Agradecimentos

Este trabalho foi realizado com o suporte do Prof.Dr. Fernando Bastos juntamente com os alunos Thais Suelen, Michael Douglas Gomes Silva, Jose Victor, Ana Carolina Saliba, graças a eles foi possivel concluir este trabalho.

Referencias

R Nishii. "Box-Cox Transformation". Em:Encyclopedia of Mathematics,Springer, New York(2001)

TEIXEIRA, Luís, 4.1.1-transformaçao de cox-cox, http://www.portalaction.com.br/analise-de-capacidade/ 411-transformacao-de-box-cox - acesso em: 11/novembro/2019

TEIXEIRA, Luís, 6.2-distribuiçao normal, http://www.portalaction.com.br/probabilidades/ 62-distribuicao-normal -acesso em: 11/novembro/2019