Współczynnik asymetrii

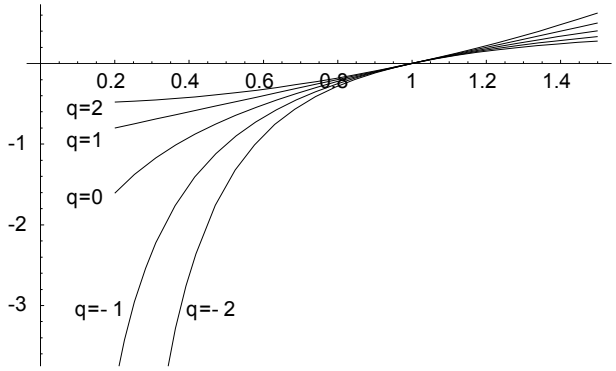
$$\gamma(\alpha)=\frac{q(1-\alpha)+q(\alpha)-2q(0{,}5)}{q(1-\alpha)-q(\alpha)},$$

gdzie $q(\alpha)$ jest kwantylem rzędu α.

Współczynnik asymetrii jest sieczną funkcji symetrii Przechodząca przez punkty o współrzędnych $\left(\frac{q(1-\alpha)-q(\alpha)}{2},\frac{q(1-\alpha)+q(\alpha)}{2}\right)$ i $(0,\tilde{x})$

Rodzina przekształceń Boxa-Coxa

$$h_q(x)=\begin{cases}\dfrac{x^q-1}{q} & q\neq 0\\ \ln(x) & q=0\end{cases}$$



Szukanie przekształcenia symetryzującego z rodziny Boxa-Coxa

Rodzina Boxa-Coxa jest używana do symetryzacji danych $x_1,x_2,\ldots,x_n$ Poszukuje się takiej potęgi p, aby wektor $h_p(x_1),h_p(x_2),\ldots,h_p(x_n)$ był symetryczny.

Przykład

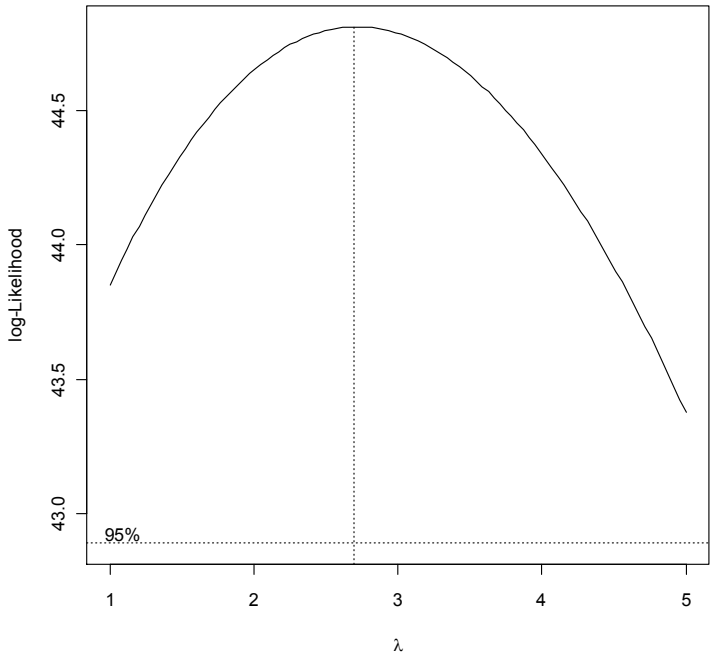
0,0682 0,0813 0,0830 0,0853 0,0982 0,1028 0,1160 0,1162 0,1208 0,1246
0,1280 0,1284 0,1294 0,1331 0,1335 0,1361 0,1402 0,1437 0,1468

Metoda prosta

Gdy asymetria ujemna –zwiększaj potęgę, gdy dodatnia - zmniejszaj potęgę

Metoda maksimum wiarygodności

Zakładając, że dane $h_p(x_1),h_p(x_2),\ldots,h_p(x_n)$ mają rozkład normalny, oblicza się logarytm funkcji wiarygodności dla tych danych. Następnie szuka się takiej wartości p, dla której logarytm funkcji wiarygodności ma największą wartość.



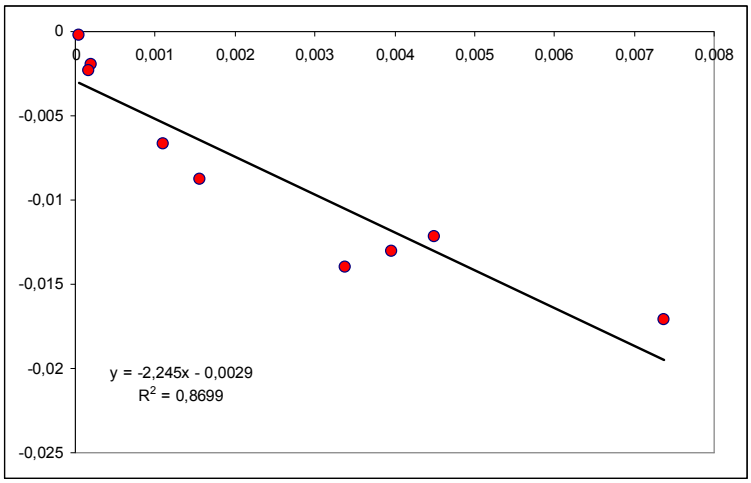
Rys. 1 Przekształcenie Boxa-Coxa. Poszukiwanie λ maksymalizującego funkcję wiarygodności (wykres z pakietu R – procedura boxcox(MASS))

Metody symetryzujące

Wykres Emmersona-Stoto:

$$\left(\frac{\left(x_{(k)}-\tilde{x}\right)^2+\left(x_{n+1-(k)}-\tilde{x}\right)^2}{4\tilde{x}}, \frac{x_{(k)}+x_{n+1-(k)}}{2}-\tilde{x} \right)$$

Gdy wykres jest liniowy to współczynnik kierunkowy =1-*p*



Rys. 2 Wykres Emmersona-Stoto. *p*≈1-(-2,245)=3,245

Metoda Hinkleya

Gdy *p* jest potęgą przekształcenia symetryzującego to zachodzi związek $\frac{x_{(k)}^p + x_{(n+1-k)}^p}{2} = \tilde{x}^p$ .Oznaczmy $u_k = \frac{x_{(k)}}{\tilde{x}}$. Wtedy zachodzi wzór Hinkleya:

$$u_k^p + u_{n+1-k}^p = 2 \text{ dla } k \le \frac{n+1}{2}$$

Wartość *p* wyznacza się metodami numerycznymi

Metoda siecznych

[plik metoda siecznych.pdf]

Outliery:

Tabela 5 liczb składa się z wartości q_0, q_1, q_2, q_3, q_4 stanowiących kwartyle[1], czyli wartości kwantyli: $\min, q(1/4), q(1/2), q(3/4), \max$.

Obliczamy: odstęp kwartylowy $IQR = q_3 - q_1$ oraz wartość kroku $h = 1{,}5\,IQR$[2].

Wartości odstające to:

- zawarte w przedziale $(q_1 - 2h, q_1 - h)$ (odstające małe)
- zawarte w przedziale $(q_3 + h, q_3 + 2h)$ (odstające duże)

Wartości ekstremalne to:

- zawarte w przedziale $(-\infty, q_1 - 2h)$ (ekstremalne małe)
- zawarte w przedziale $(q_3 + 2h, \infty)$ (ekstremalne duże)

[1] W arkuszu Excel dostępne jako funkcja statystyczna **kwartyl**

[2] Współczynnik 1,5 zapewnia dużą wartość kroku. Gdy rozkład jest normalny to 1,5 IQR odpowiada wartości 2,04σ. Wartości odległe o więcej niż krok od q_3 występują z prawdopodobieństwem 0,003 a o dwa kroki z mniejszym niż 10^(-5)