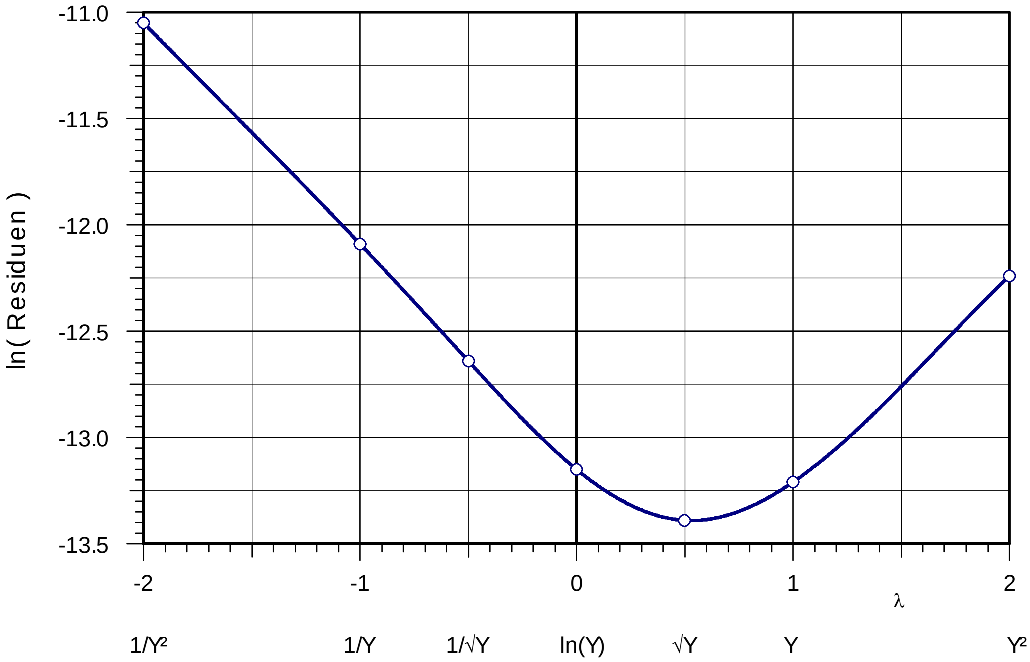
Transformation der Zielgröße (Box-Cox-Transformation)

Evtl. lässt sich das Regressionsmodell besser bestimmen, wenn die Zielgröße transformiert wird. Oft ist es z.B. notwendig, die Zielgröße y zu logarithmieren, wenn es sich bei Y um Verschleißwerte handelt.
Mit der sogenannte Box-Cox-Transformation kann bestimmt werden, welche die geeignetste Transformation ist. Nacheinander wird die Zielgröße mit durchlaufendem Exponent λ transformiert bzw. entsprechend folgender Beziehung umgeformt.

$$Y^{(\lambda)} = \begin{cases} \lambda^{-1}\,\overline{Y}^{1-\lambda}\,(Y^{\lambda}-1) & wenn\,\lambda \neq 0 \\ \overline{Y}\,\ln(Y) & wenn\,\lambda = 0 \end{cases}$$

Für den Wert λ =0 wird per Definition ln(Y) verwendet. Mit diesen neuen Y-Werten wird jeweils die Regression durchgeführt und die Residuen ermittelt. In der Regel bildet man hieraus noch den ln, um kleinere Werte zu erhalten. Je kleiner die Residuen und somit die Abweichungen vom Modell zu den Messdaten sind, desto besser ist die zu wählende Transformation (hier √Y entsprechend λ=0,5). Weitere Beschreibungen sind unter /2/ beschrieben.



Neben diesen einfachen Standardtransformationen sind beliebige komplexe Formeln möglich. Sind die Y-Werte negativ oder enthalten die 0, so ist die folgende Variante der ln-Transformation sinnvoll:

$$Y' = a' \ln\left(b'Y + c'\right)$$

Dabei wird *b'* und *c'* so angesetzt, dass im Argument keine Werte ≤ 0 auftreten können. Soll das Regressionsmodell z.B. nur Werte zwischen 0 und 1 liefern, so kann die Transformation:

$$Y' = a' Tan(b'Y + c')$$

gewählt werden. Für die Rückrechnung auf die Modellwerte ist die Umkehrfunktion anzuwenden:

$$Y = \frac{1}{b'}\left(ArcTan\left(\frac{Y'}{a'}\right) - c'\right)$$

Um grundsätzlich Werte zwischen 0 und 1 zu erhalten, ist $b'=\pi$ und c= -0,5 zu wählen. a' bestimmt, wie stark die S-Form geneigt ist. Weitere Transformation sind: