Power transform

In statistics, a **power transform** is a family of functions applied to create a monotonic transformation of data using power functions. It is a data transformation technique used to stabilize variance, make the data more normal distribution-like, improve the validity of measures of association (such as the Pearson correlation between variables), and for other data stabilization procedures.

Power transforms are used in multiple fields, including multi-resolution and wavelet analysis,[1] statistical data analysis, medical research, modeling of physical processes,[2] geochemical data analysis,[3] epidemiology[4] and many other clinical, environmental and social research areas.

Definition

The power transformation is defined as a continuous function of power parameter λ, typically given in piece-wise form that makes it continuous at the point of singularity ($\lambda = 0$). For data vectors $(y_1,..., y_n)$ in which each $y_i > 0$, the power transform is

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^\lambda - 1}{\lambda(\mathrm{GM}(y))^{\lambda-1}}, & \text{if } \lambda \neq 0 \\[2em] \mathrm{GM}(y)\ln y_i, & \text{if } \lambda = 0 \end{cases}$$

where

$$\mathrm{GM}(y) = \left(\prod_{i=1}^n y_i \right)^{\frac{1}{n}} = \sqrt[n]{y_1 y_2 \cdots y_n}$$

is the geometric mean of the observations $y_1, ..., y_n$. The case for $\lambda = 0$ is the limit as λ approaches 0. To see this, note that $y_i^\lambda = \exp(\lambda \ln(y_i)) = 1 + \lambda \ln(y_i) + O((\lambda \ln(y_i))^2)$ - using Taylor series. Then $\dfrac{y_i^\lambda - 1}{\lambda} = \ln(y_i) + O(\lambda)$, and everything but $\ln(y_i)$ becomes negligible for λ sufficiently small.

The inclusion of the ($\lambda - 1$)th power of the geometric mean in the denominator simplifies the scientific interpretation of any equation involving $y_i^{(\lambda)}$, because the units of measurement do not change as λ changes.

Box and Cox (1964) introduced the geometric mean into this transformation by first including the Jacobian of rescaled power transformation

$$\frac{y^\lambda - 1}{\lambda}.$$

with the likelihood. This Jacobian is as follows:

$$J(\lambda; y_1, \ldots, y_n) = \prod_{i=1}^{n} |dy_i^{(\lambda)}/dy| = \prod_{i=1}^{n} y_i^{\lambda-1} = GM(y)^{n(\lambda-1)}$$

This allows the normal <u>log likelihood at its maximum</u> to be written as follows:

$$\log(\mathcal{L}(\hat{\mu}, \hat{\sigma})) = (-n/2)(\log(2\pi\hat{\sigma}^2) + 1) + n(\lambda - 1)\log(GM(y))$$
$$= (-n/2)(\log(2\pi\hat{\sigma}^2 / GM(y)^{2(\lambda-1)}) + 1).$$

From here, absorbing $GM(y)^{2(\lambda-1)}$ into the expression for $\hat{\sigma}^2$ produces an expression that establishes that minimizing the sum of squares of <u>residuals</u> from $y_i^{(\lambda)}$ is equivalent to maximizing the sum of the normal <u>log likelihood</u> of deviations from $(y^\lambda - 1)/\lambda$ and the log of the Jacobian of the transformation.

The value at $Y = 1$ for any λ is 0, and the <u>derivative</u> with respect to Y there is 1 for any λ. Sometimes Y is a version of some other variable scaled to give $Y = 1$ at some sort of average value.

The transformation is a <u>power</u> transformation, but done in such a way as to make it <u>continuous</u> with the parameter λ at $\lambda = 0$. It has proved popular in <u>regression analysis</u>, including <u>econometrics</u>.

Box and Cox also proposed a more general form of the transformation that incorporates a shift parameter.

$$\tau(y_i; \lambda, \alpha) = \begin{cases} \dfrac{(y_i + \alpha)^\lambda - 1}{\lambda(GM(y+\alpha))^{\lambda-1}} & \text{if } \lambda \neq 0, \\[2em] GM(y+\alpha)\ln(y_i + \alpha) & \text{if } \lambda = 0, \end{cases}$$

which holds if $y_i + \alpha > 0$ for all i. If $\tau(Y, \lambda, \alpha)$ follows a <u>truncated normal distribution</u>, then Y is said to follow a <u>Box–Cox distribution</u>.

Bickel and Doksum eliminated the need to use a <u>truncated distribution</u> by extending the range of the transformation to all y, as follows:

$$\tau(y_i; \lambda, \alpha) = \begin{cases} \dfrac{\operatorname{sgn}(y_i + \alpha)|y_i + \alpha|^\lambda - 1}{\lambda(GM(y+\alpha))^{\lambda-1}} & \text{if } \lambda \neq 0, \\[2em] GM(y+\alpha)\operatorname{sgn}(y+\alpha)\ln(y_i + \alpha) & \text{if } \lambda = 0, \end{cases}$$

where sgn(.) is the <u>sign function</u>. This change in definition has little practical import as long as α is less than $\min(y_i)$, which it usually is.[5]

Bickel and Doksum also proved that the parameter estimates are <u>consistent</u> and <u>asymptotically normal</u> under appropriate regularity conditions, though the standard <u>Cramér–Rao lower bound</u> can substantially underestimate the variance when parameter values are small relative to the noise variance.[5] However, this problem of underestimating the variance may not be a substantive problem in many applications.[6][7]

Box–Cox transformation

The one-parameter Box–Cox transformations are defined as

$$
y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0, \\[2ex] \ln y_i & \text{if } \lambda = 0, \end{cases}
$$

and the two-parameter Box–Cox transformations as

$$
y_i^{(\lambda)} = \begin{cases} \dfrac{(y_i + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \text{if } \lambda_1 \neq 0, \\[2ex] \ln(y_i + \lambda_2) & \text{if } \lambda_1 = 0, \end{cases}
$$

as described in the original article.[8][9] Moreover, the first transformations hold for $y_i > 0$, and the second for $y_i > -\lambda_2$.[8]

The parameter λ is estimated using the profile likelihood function and using goodness-of-fit tests.[10]
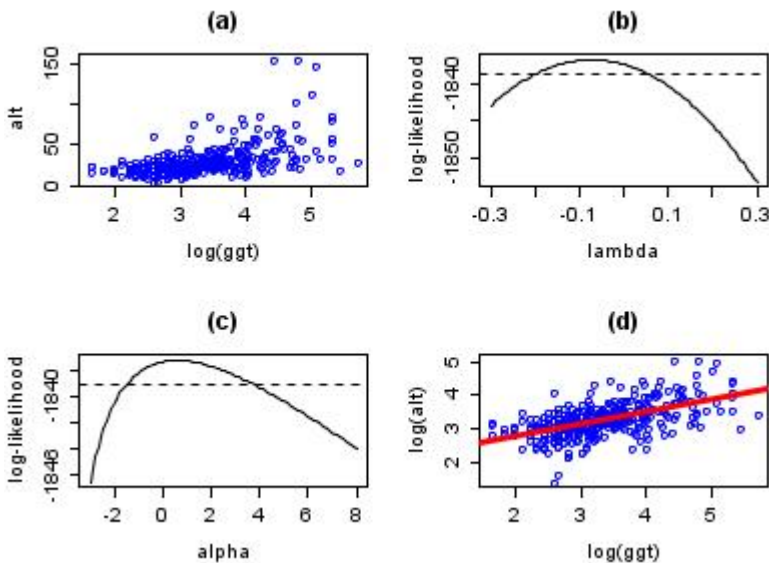
Confidence interval

Confidence interval for the Box–Cox transformation can be asymptotically constructed using Wilks's theorem on the profile likelihood function to find all the possible values of λ that fulfill the following restriction:[11]

$$
\ln\left(L(\lambda)\right) \geq \ln\left(L(\hat{\lambda})\right) - \frac{1}{2}\chi^2_{1,1-\alpha}.
$$

Example

The BUPA liver data set[12] contains data on liver enzymes ALT and γGT. Suppose we are interested in using log(γGT) to predict ALT. A plot of the data appears in panel (a) of the figure. There appears to be non-constant variance, and a Box–Cox transformation might help.

The log-likelihood of the power parameter appears in panel (b). The horizontal reference line is at a distance of $\chi_1^2/2$ from the maximum and can be used to read off an approximate 95% confidence interval for λ. It appears as though a value close to zero would be good, so we take logs.

Possibly, the transformation could be improved by adding a shift parameter to the log transformation. Panel (c) of the figure shows the log-likelihood. In this case, the maximum of the likelihood is close to zero suggesting that a shift parameter is not needed. The final panel shows the transformed data with a superimposed regression line.

Note that although Box–Cox transformations can make big improvements in model fit, there are some issues that the transformation cannot help with. In the current example, the data are rather heavy-tailed so that the assumption of normality is not realistic and a robust regression approach leads to a more precise model.

Econometric application

Economists often characterize production relationships by some variant of the Box–Cox transformation.[13]

Consider a common representation of production Q as dependent on services provided by a capital stock K and by labor hours N:

$$\tau(Q) = \alpha\tau(K) + (1 - \alpha)\tau(N).$$

Solving for Q by inverting the Box–Cox transformation we find

$$Q = \left(\alpha K^\lambda + (1 - \alpha)N^\lambda\right)^{1/\lambda},$$

which is known as the *constant elasticity of substitution (CES)* production function.

The CES production function is a homogeneous function of degree one.

When $\lambda = 1$, this produces the linear production function:

$$Q = \alpha K + (1 - \alpha)N.$$

When $\lambda \to 0$ this produces the famous Cobb–Douglas production function:

$$Q = K^\alpha N^{1-\alpha}.$$

Activities and demonstrations

The SOCR resource pages contain a number of hands-on interactive activities[14] demonstrating the Box–Cox (power) transformation using Java applets and charts. These directly illustrate the effects of this transform on Q–Q plots, X–Y scatterplots, time-series plots and histograms.

Yeo–Johnson transformation

The Yeo–Johnson transformation[15] allows also for zero and negative values of y. λ can be any real number, where $\lambda = 1$ produces the identity transformation. The transformation law reads:

$$y_i^{(\lambda)} = \begin{cases} ((y_i + 1)^\lambda - 1)/\lambda & \text{if } \lambda \neq 0, y \geq 0 \\ \ln(y_i + 1) & \text{if } \lambda = 0, y \geq 0 \\ -((-y_i + 1)^{(2-\lambda)} - 1)/(2 - \lambda) & \text{if } \lambda \neq 2, y < 0 \\ -\ln(-y_i + 1) & \text{if } \lambda = 2, y < 0 \end{cases}$$

Box-Tidwell transformation

The Box-Tidwell transformation is a statistical technique used to assess and correct non-linearity between predictor variables and the logit in a generalized linear model, particularly in **logistic regression**. This transformation is useful when the relationship between the independent variables and the outcome is non-linear and cannot be adequately captured by the standard model.

Overview

The Box-Tidwell transformation was developed by **George E. P. Box** and John W. Tidwell in 1962 as an extension of **Box-Cox transformations**, which are applied to the dependent variable. However, unlike the Box-Cox transformation, the Box-Tidwell transformation is applied to the independent variables in regression models. It is often used when the assumption of linearity between the predictors and the outcome is violated.

Method

The general idea behind the Box-Tidwell transformation is to apply a power transformation to each independent variable Xi in the regression model:

$$X_i' = X_i^\lambda$$

Where λ is the parameter estimated from the data. If Box-Tidwell Transformation is significantly different from 1, this indicates a non-linear relationship between Xi and the logit, and the transformation improves the model fit.

The Box-Tidwell test is typically performed by augmenting the regression model with terms like $X_i \log(X_i)$ and testing the significance of the coefficients. If significant, this suggests that a transformation should be applied to achieve a linear relationship between the predictor and the logit.

Applications

Stabilizing Continuous Predictors

The transformation is beneficial in logistic regression or proportional hazards models where non-linearity in continuous predictors can distort the relationship with the dependent variable. It is a flexible tool that allows the researcher to fit a more appropriate model to the data without guessing the relationship's functional form in advance.

Verifying Linearity in Logistic Regression

In logistic regression, a key assumption is that continuous independent variables exhibit a linear relationship with the logit of the dependent variable. Violations of this assumption can lead to biased estimates and reduced model performance. The Box-Tidwell transformation is a method used to assess and correct such violations by determining whether a continuous predictor requires transformation to achieve linearity with the logit.

Method for Verifying Linearity

The Box-Tidwell transformation introduces an interaction term between each continuous variable X_i and its natural logarithm $\log(X_i)$:

$$X_i \log(X_i)$$

This term is included in the logistic regression model to test whether the relationship between X_i and the logit is non-linear. A statistically significant coefficient for this interaction term indicates a violation of the linearity assumption, suggesting the need for a transformation of the predictor. the Box-Tidwell transformation provides an appropriate power transformation to linearize the relationship, thereby improving model accuracy and validity. Conversely, non-significant results support the assumption of linearity.

Limitations

One limitation of the Box-Tidwell transformation is that it only works for positive values of the independent variables. If your data contains negative values, the transformation cannot be applied directly without modifying the variables (e.g., adding a constant).

Notes

1. Gao, Peisheng; Wu, Weilin (2006). "Power Quality Disturbances Classification using Wavelet and Support Vector Machines". *Sixth International Conference on Intelligent Systems Design and Applications*. ISDA '06. Vol. 1. Washington, DC, USA: IEEE Computer Society. pp. 201–206. doi:10.1109/ISDA.2006.217 (https://doi.org/10.1109%2FISDA.2006.2 17). ISBN 9780769525280. S2CID 2444503 (https://api.semanticscholar.org/CorpusID:2444 503).
2. Gluzman, S.; Yukalov, V. I. (2006-01-01). "Self-similar power transforms in extrapolation problems". *Journal of Mathematical Chemistry*. **39** (1): 47–56. arXiv:cond-mat/0606104 (http s://arxiv.org/abs/cond-mat/0606104). Bibcode:2006cond.mat..6104G (https://ui.adsabs.harv ard.edu/abs/2006cond.mat..6104G). doi:10.1007/s10910-005-9003-7 (https://doi.org/10.100 7%2Fs10910-005-9003-7). ISSN 1572-8897 (https://search.worldcat.org/issn/1572-8897). S2CID 118965098 (https://api.semanticscholar.org/CorpusID:118965098).

3. Howarth, R. J.; Earle, S. A. M. (1979-02-01). "Application of a generalized power transformation to geochemical data". *Journal of the International Association for Mathematical Geology*. **11** (1): 45–62. doi:10.1007/BF01043245 (https://doi.org/10.1007%2FBF01043245). ISSN 1573-8868 (https://search.worldcat.org/issn/1573-8868). S2CID 121582755 (https://api.semanticscholar.org/CorpusID:121582755).

4. Peters, J. L.; Rushton, L.; Sutton, A. J.; Jones, D. R.; Abrams, K. R.; Mugglestone, M. A. (2005). "Bayesian methods for the cross-design synthesis of epidemiological and toxicological evidence". *Journal of the Royal Statistical Society, Series C*. **54**: 159–172. doi:10.1111/j.1467-9876.2005.00476.x (https://doi.org/10.1111%2Fj.1467-9876.2005.00476.x). S2CID 121909404 (https://api.semanticscholar.org/CorpusID:121909404).

5. Bickel, Peter J.; Doksum, Kjell A. (June 1981). "An analysis of transformations revisited". *Journal of the American Statistical Association*. **76** (374): 296–311. doi:10.1080/01621459.1981.10477649 (https://doi.org/10.1080%2F01621459.1981.10477649).

6. Sakia, R. M. (1992), "The Box–Cox transformation technique: a review", *The Statistician*, **41** (2): 169–178, CiteSeerX 10.1.1.469.7176 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.469.7176), doi:10.2307/2348250 (https://doi.org/10.2307%2F2348250), JSTOR 2348250 (https://www.jstor.org/stable/2348250)

7. Li, Fengfei (April 11, 2005), *Box–Cox Transformations: An Overview* (http://www.ime.usp.br/~abe/lista/pdfm9cJKUmFZp.pdf) (PDF) (slide presentation), Sao Paulo, Brazil: University of Sao Paulo, Brazil, retrieved 2014-11-02

8. Box, George E. P.; Cox, D. R. (1964). "An analysis of transformations". *Journal of the Royal Statistical Society, Series B*. **26** (2): 211–252. JSTOR 2984418 (https://www.jstor.org/stable/2984418). MR 0192611 (https://mathscinet.ams.org/mathscinet-getitem?mr=0192611).

9. Johnston, J. (1984). *Econometric Methods* (Third ed.). New York: McGraw-Hill. pp. 61–74. ISBN 978-0-07-032685-9.

10. Asar, O.; Ilk, O.; Dag, O. (2017). "Estimating Box-Cox power transformation parameter via goodness-of-fit tests". *Communications in Statistics - Simulation and Computation*. **46** (1): 91–105. arXiv:1401.3812 (https://arxiv.org/abs/1401.3812). doi:10.1080/03610918.2014.957839 (https://doi.org/10.1080%2F03610918.2014.957839). S2CID 41501327 (https://api.semanticscholar.org/CorpusID:41501327).

11. Abramovich, Felix; Ritov, Ya'acov (2013). *Statistical Theory: A Concise Introduction*. CRC Press. pp. 121–122. ISBN 978-1-4398-5184-5.

12. BUPA liver disorder dataset (https://web.archive.org/web/20171023174701/http://ftp.ics.uci.edu:80/pub/machine-learning-databases/liver-disorders/)

13. Zarembka, P. (1974). "Transformation of Variables in Econometrics". *Frontiers in Econometrics*. New York: Academic Press. pp. 81–104. ISBN 0-12-776150-0.

14. Power Transform Family Graphs (http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_PowerTransformFamily_Graphs), SOCR webpages

15. Yeo, In-Kwon; Johnson, Richard A. (2000). "A New Family of Power Transformations to Improve Normality or Symmetry". *Biometrika*. **87** (4): 954–959. doi:10.1093/biomet/87.4.954 (https://doi.org/10.1093%2Fbiomet%2F87.4.954). JSTOR 2673623 (https://www.jstor.org/stable/2673623).

References

- Box, George E. P.; Cox, D. R. (1964). "An analysis of transformations". *Journal of the Royal Statistical Society, Series B*. **26** (2): 211–252. JSTOR 2984418 (https://www.jstor.org/stable/2984418). MR 0192611 (https://mathscinet.ams.org/mathscinet-getitem?mr=0192611).
- Carroll, R. J.; Ruppert, D. (1981). "On prediction and the power transformation family" (http://wiki.stat.ucla.edu/socr/uploads/b/b8/PowerTransformFamily_Biometrica609.pdf) (PDF).

Biometrika. **68** (3): 609–615. doi:10.1093/biomet/68.3.609 (https://doi.org/10.1093%2Fbiomet%2F68.3.609).

- DeGroot, M. H. (1987). "A Conversation with George Box" (https://projecteuclid.org/journals/statistical-science/volume-2/issue-3/A-Conversation-with-George-Box/10.1214/ss/117701322 23.pdf) (PDF). *Statistical Science*. **2** (3): 239–258. doi:10.1214/ss/1177013223 (https://doi.org/10.1214%2Fss%2F1177013223).
- Handelsman, D. J. (2002). "Optimal Power Transformations for Analysis of Sperm Concentration and Other Semen Variables". *Journal of Andrology*. **23** (5).
- Gluzman, S.; Yukalov, V. I. (2006). "Self-similar power transforms in extrapolation problems". *Journal of Mathematical Chemistry*. **39** (1): 47–56. arXiv:cond-mat/0606104 (https://arxiv.org/abs/cond-mat/0606104). Bibcode:2006cond.mat..6104G (https://ui.adsabs.harvard.edu/abs/2006cond.mat..6104G). doi:10.1007/s10910-005-9003-7 (https://doi.org/10.1007%2Fs10910-005-9003-7). S2CID 118965098 (https://api.semanticscholar.org/CorpusID:118965098).
- Howarth, R. J.; Earle, S. A. M. (1979). "Application of a generalized power transformation to geochemical data". *Journal of the International Association for Mathematical Geology*. **11** (1): 45–62. doi:10.1007/BF01043245 (https://doi.org/10.1007%2FBF01043245). S2CID 121582755 (https://api.semanticscholar.org/CorpusID:121582755).
- Box, G.E.P. and Tidwell, P.W. (1962) Transformation of Independent Variables. Technometrics, 4, 531-550. https://doi.org/10.1080/00401706.1962.10490038 (a.k.a: Box-Tidwell transformation)

External links

- Nishii, R. (2001) [1994], "Box–Cox transformation" (https://www.encyclopediaofmath.org/index.php?title=Box–Cox_transformation), *Encyclopedia of Mathematics*, EMS Press (fixed link (http://www.encyclopediaofmath.org/index.php/Box%E2%80%93Cox_transformation))
- Sanford Weisberg, Yeo-Johnson Power Transformations (https://www.stat.umn.edu/arc/yjpower.pdf)