STAT 3008 Applied Regression Analysis
Tutorial 9.

XU Yongze & DONG Fangyuan

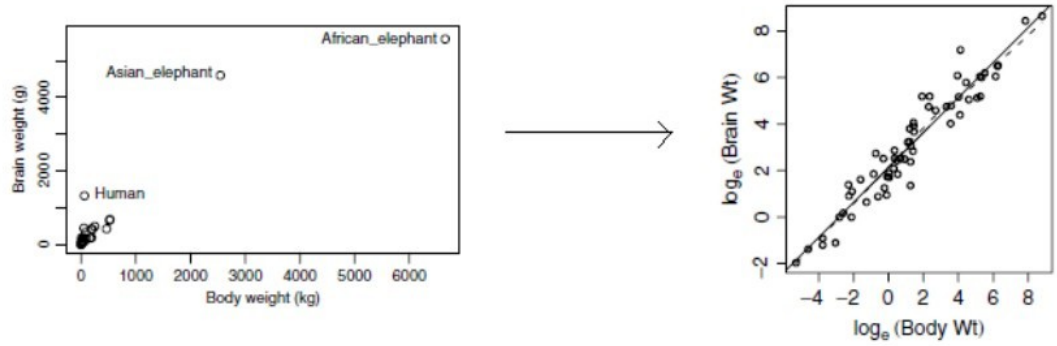
Nov 18 & 19 & 21, 2013

Contents

1 Transformation

1.1 Why transformation?

♣ The assumption of linear relation does not hold.



We can transform:

⋆ The predictor

⋆ The response

⋆ Both

to achieve the linear relationship.

1.2 Transformation methods for simple linear regression

Power Transformation

$$\psi(X,\lambda) = \begin{cases} X^{\lambda}, & if \quad \lambda \neq 0 \\ \log X, & if \quad \lambda = 0 \end{cases}$$

♣ Log transformation is useful when

1. Observations are positive
2. Range if variable is huge(i.e. $Max(X) \gg Min(X)$)

♣ Continuation of the example in part1.1
Both variables have large range. So we perform transformation on both the predictor and the response.

$$\psi(BrainWt, 0) = \alpha + \beta\psi(BodyWt, 0) + e$$

$$\Rightarrow log(BrainWt) = \alpha + \beta log(BodyWt) + e$$

$$\Rightarrow BrainWt = e^{\alpha}(BodyWt)^{\beta}e^{e} = \tilde{\alpha}(BodyWt)^{\beta}\delta$$

δ is multiplicative error.

$$\widehat{BrainWt} = e^{log(\widehat{BrainWt})} = e^{\hat{\alpha}+\hat{\beta}log(BodyWt)}$$

C.I. for $e^{\alpha+\beta log(BodyWt)} \leftarrow$ Delta method.

Scaled Power Transformation

$$\psi_s(X,\lambda) = \begin{cases} \frac{X^{\lambda}-1}{\lambda}, & if \;\; \lambda \neq 0 \\ \log X, & if \;\; \lambda = 0 \end{cases}$$

♣ Advantages:

1. Continuous function of λ
2. Preserve the direction of association.

E.g. find transformation for X for true model $E(Y|X) = \beta X^{-\frac{1}{2}}$,
Power transformation,
Let $\psi(X,\lambda) = X^{-\frac{1}{2}}$ (i.e. $\lambda = -\frac{1}{2}$), then $E(Y|X) = \beta\psi(X, -\frac{1}{2})$.
Scaled power transform,
$E(Y|X) = \tilde{\alpha} + \tilde{\beta}\psi_S(X,\lambda) = \beta X^{-\frac{1}{2}} \Rightarrow \tilde{\alpha} = \beta, \tilde{\beta} = -\frac{1}{2}\beta, \lambda = -\frac{1}{2}$.

1.3 Transformation selection for simple linear regression

Remarks:

1. It is to choose λ;
2. Suppose we only transform X now.

♣ Method 1: Draw many fitted curves

i.e. Plot $(x, \hat{y})$ for various λ.

♣ Method 2: Draw many scatter plots

i.e. Plot $(\psi(X,\lambda), y)$.

♣ Method 3: Plot λ against RSS of fitting y against $\psi(X,\lambda)$, then find the λ that minimizes RSS.

1.4 Methods for multiple regression

Method 1: Inverse fitted value plot

$\star$ Plot $\hat{Y}$ against Y.

$\star$ Find transformation for Y. $\leftarrow$ see if curve $(Y, \hat{Y_\lambda})$ matches the pattern of $(Y, \hat{Y})$

$\star$ Match $\rightarrow \hat{\beta}_0 + \hat{\beta}_1 \psi_S(Y, \lambda) = \hat{Y_\lambda} \approx \hat{Y} = X\hat{\beta}$

$\hat{Y}$ and $\psi_S(Y, \lambda)$ linearly related, while $\hat{Y}$ and $X\hat{\beta}$ linearly related, so $\psi_S(Y, \lambda)$ and $X\hat{\beta}$ linearly related.

Which is the correct order of the steps in the Inverse fitted value plot?

[a]. Plot $\hat{Y_\lambda}$ against Y.

[b]. Fit a linear regression between Y and X, obtaining $\hat{Y}$.

[c]. Specify a scaled-power transformation $\psi_S(Y; \lambda)$.

[d]. Fit $\hat{Y}$ against $\psi_S(Y; \lambda)$, obtaining $\hat{Y_\lambda}$.

[e]. Plot $\hat{Y}$ against Y.

Ans: b,e,c,d,a.

$\star$ After the above steps, we have decided an λ. Then go back to the analysis and fit $\psi_S(Y, \lambda)$ against X.

Method 2: Modified Power transformation for all predictors

Use the modified power family:

$$\psi_M(Y, \lambda) = \begin{cases} (\sqrt[n]{y_1 \cdots y_n})^{1-\lambda} \dfrac{Y^\lambda - 1}{\lambda}, & if \;\; \lambda \neq 0 \\ (\sqrt[n]{y_1 \cdots y_n})^{1-\lambda} \log Y, & if \;\; \lambda = 0 \end{cases}$$

Transform predictors so that each pair of predictors in the scatterplot matrix have a linear relationship.

$$(X_1, X_2, \cdots, X_p) \rightarrow (\psi_M(X_1, \lambda_1), \psi_M(X_2, \lambda_2), \cdots, \psi_M(X_p, \lambda_p))$$

Method 3: Box Cox Transformation

Modified power family:

$$\psi_M(Y, \lambda) = \begin{cases} (\sqrt[n]{y_1 \cdots y_n})^{1-\lambda} \dfrac{Y^\lambda - 1}{\lambda}, & if \;\; \lambda \neq 0 \\ (\sqrt[n]{y_1 \cdots y_n})^{1-\lambda} \log Y, & if \;\; \lambda = 0 \end{cases}$$

♣ Advantage:

Unit of $\psi_M(Y, \lambda)$ is the same as Y for all λ.

♣ Selection of λ.

$\star$ Fix a λ, fit model with transformed Y_λ and obtain $RSS(\lambda)$

$\star$ Try various λ and find the one which minimizes $RSS(\lambda)$

2 Example

Thomas is analyzing a data set using the Box-Cox transformation. The minimum value of RSS occurs when $\lambda = -0.979$. Therefore, he transformed the response with

$$(\sqrt[n]{y_1 \cdots y_n})^{1.979} \frac{Y^{-0.979} - 1}{-0.979}$$

Any comment?

$\lambda = -0.979$ seems to have no good interpretations or meanings. I would suggest $\lambda = -1$ as it may have a better interpretation.

3 Discussions on Exercise 8-10 solutions