应用回归分析

上海财经大学　统计与管理学院

第七章变量变换

章节概括：

- 变量变换
- 指数族变换
- 比例指数变换
- Box-Cox方法 （修改指数变换）

简单线性回归

- （近似）线性回归

$$\mathrm{E}(Y|X=x) \approx \beta_0 + \beta_1 x$$

- 二维散点图

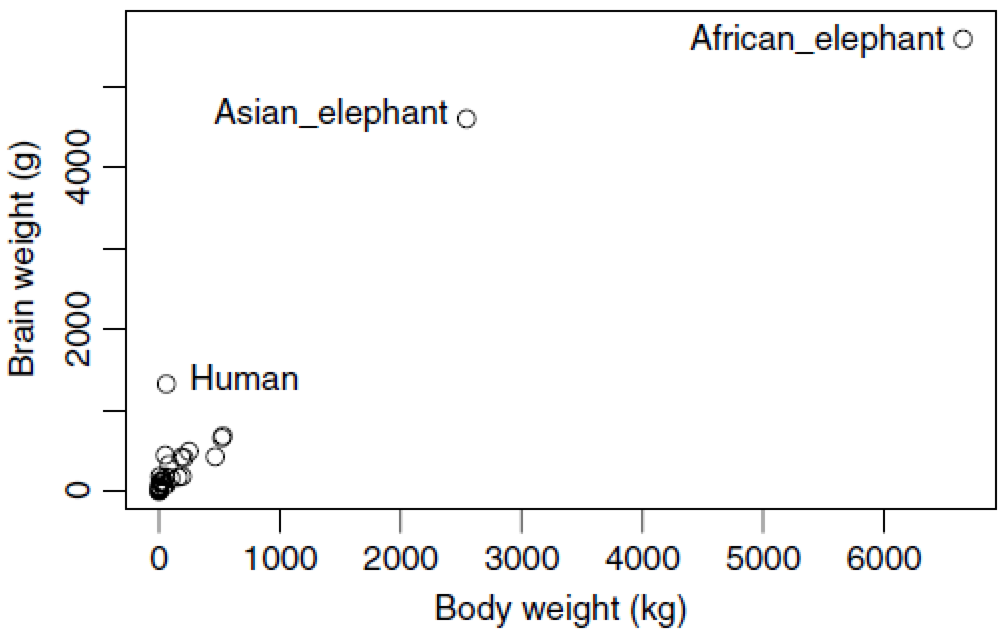

FIG. 7.1 Plot of *BrainWt* versus *BodyWt* for 62 mammal species.

指数族变换

- 指数族变换

$$\psi(U,\lambda) = U^{\lambda}$$

- 参数 λ = 0, log变换

  1, 无变换

  1/2， 1/3

  (-2,2)

- 自变量 U 非负

简单线性回归

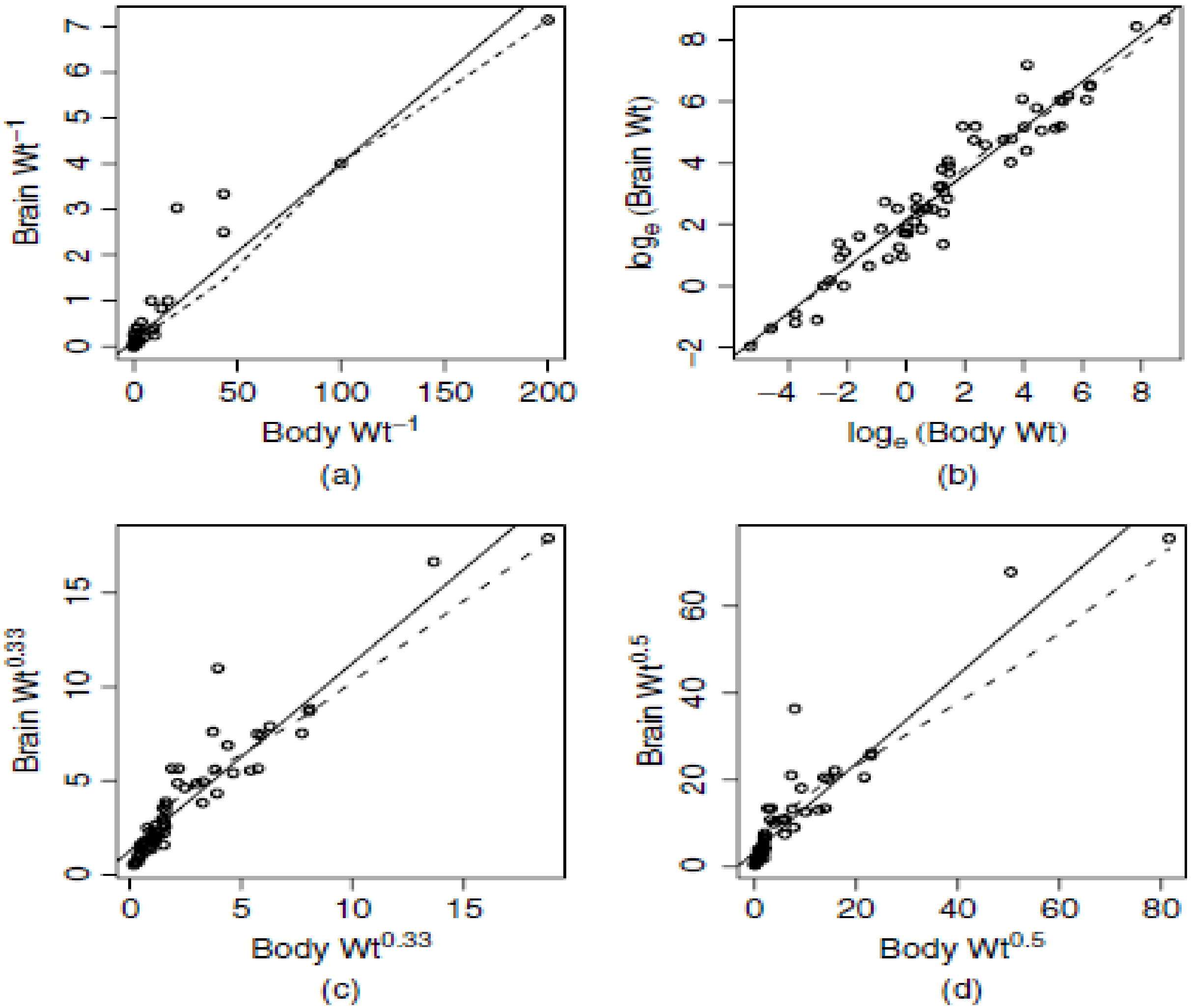

FIG. 7.2 Scatterplots for the brain weight data with four possible transformations. The solid line on each plot is the OLS line; the dashed line is a *loess* smooth.

Log规则

- Log规则，变量取值区间跨越多个的数量级且为正
- Range规则，变量取值区间在一个数量级内
- 乘法误差 $e = \log(\delta)$ 同时变换自变量和因变量

$$BrainWt = \alpha \times BodyWt^{\beta_1} \times \delta$$

$$\log(BrainWt) = \beta_0 + \beta_1 \log(BodyWt) + e$$

自变量变换

- scaled power transformation (比例指数变换）

$$\psi_S(X,\lambda) = \begin{cases} (X^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log(X) & \text{if } \lambda = 0 \end{cases}$$

- 对比指数变换

  连续性 $\lim_{\lambda \to 0} \psi_S(X,\lambda) = \log_e(X)$

  保持相关性

  比例，位置，正负
- 通常用来选择一个变换

变换参数选择

- 变换模型

$$E(Y|X) = \beta_0 + \beta_1 \psi_S(X, \lambda)$$

- 优化函数

$$RSS(\lambda) = E(Y|X) = \beta_0 + \beta_1 \psi_S(X, \lambda)$$

- 参数选择

$$\lambda \in \{-1, -1/2, 0, 1/3, 1/2, 1\}$$

变换参数选择

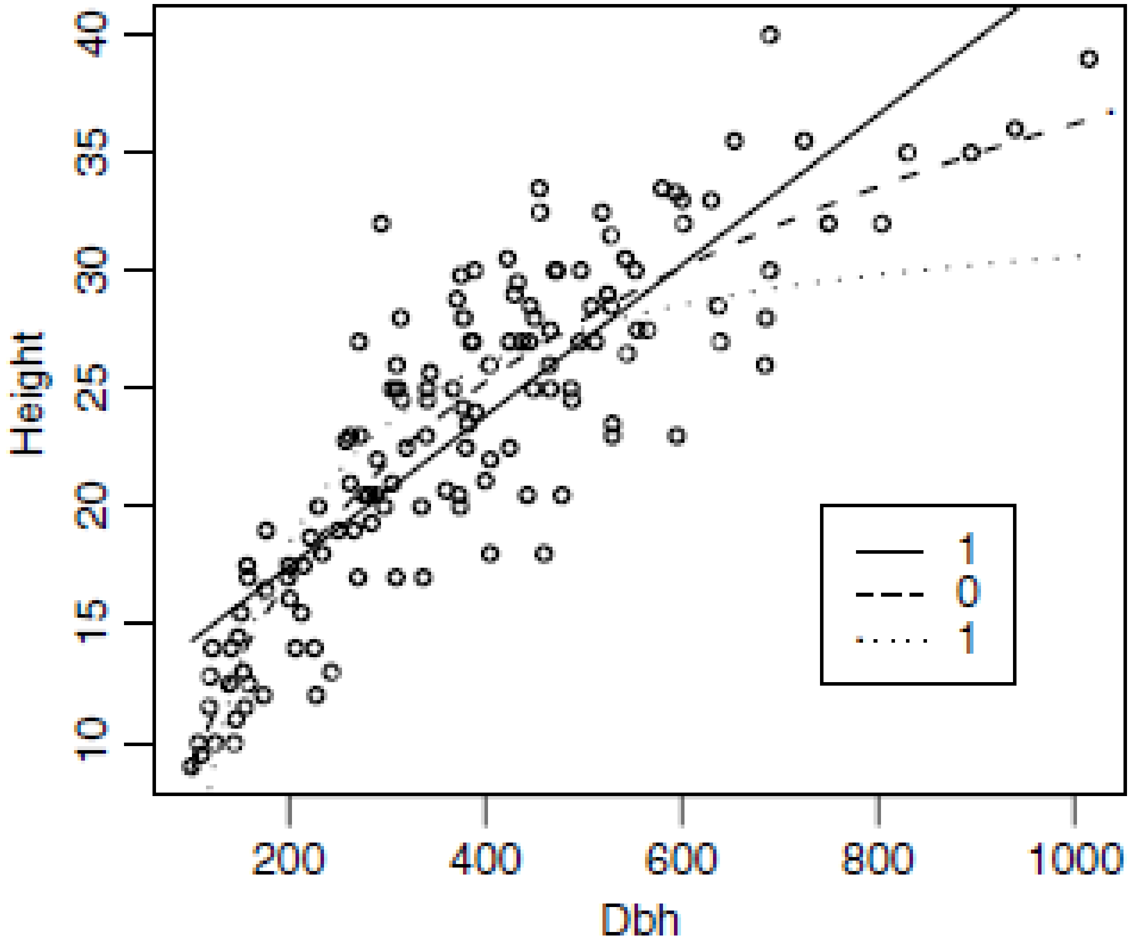

FIG. 7.3 *Height* versus *Dbh* for the red cedar data from Upper Flat Creek.

因变量变换

- 逆回归
- X和Y 转换角色
- Y 回归 X (Y自变量 / X因变量）

$$E(\hat{y}|Y) = \alpha_0 + \alpha_1 \psi_S(Y, \lambda_y)$$

- 重复比例指数变换方法，选择 λ_y

因变量变换

- Box-Cox方法 （修改指数变换）

$$\psi_M(Y,\lambda_y)=\psi_S(Y,\lambda_y)\times \mathrm{gm}(Y)^{1-\lambda_y}$$
$$=\begin{cases}\mathrm{gm}(Y)^{1-\lambda_y}\times (Y^{\lambda_y}-1)/\lambda_y & \text{if } \lambda_y\neq 0\\ \mathrm{gm}(Y)\times \log(Y) & \text{if } \lambda_y=0\end{cases}$$

$\mathrm{gm}(Y)$ 是Y的几何平均值

- 均值函数

$$\mathrm{E}(\psi_M(Y,\lambda_y)|X=\mathbf{x})=\boldsymbol{\beta}'\mathbf{x}$$

变换参数选择

- 最小函数

$$\mathrm{gm}(Y)^{1-\lambda}\, RSS(\lambda_y)$$

- 参数选择

$$\lambda \in \{-1, -1/2, 0, 1/3, 1/2, 1\}$$

正态性

- Box-Cox方法

  是为正态性而变换

  不是为线性而变换

非正变量变换

- 指数族变换

$$\psi(U,\lambda)=U^{\lambda}$$

- U非正变量
- 方法一

$$(U+\gamma)^{\lambda}$$

- 方法二（Yeo-Johnson)

$$\psi_{YJ}(U,\lambda)=\begin{cases}\psi_M(U+1,\lambda) & \text{if } U\geq 0\\ \psi_M(-U+1,2-\lambda) & \text{if } U<0\end{cases}$$

方法二

- 对比Box-Cox方法
- U > 0, (U+1), λ
- U < 0, (-U+1), $2-\lambda$
- 零点附近

Box-Cox变化大，方法二变化小

图例

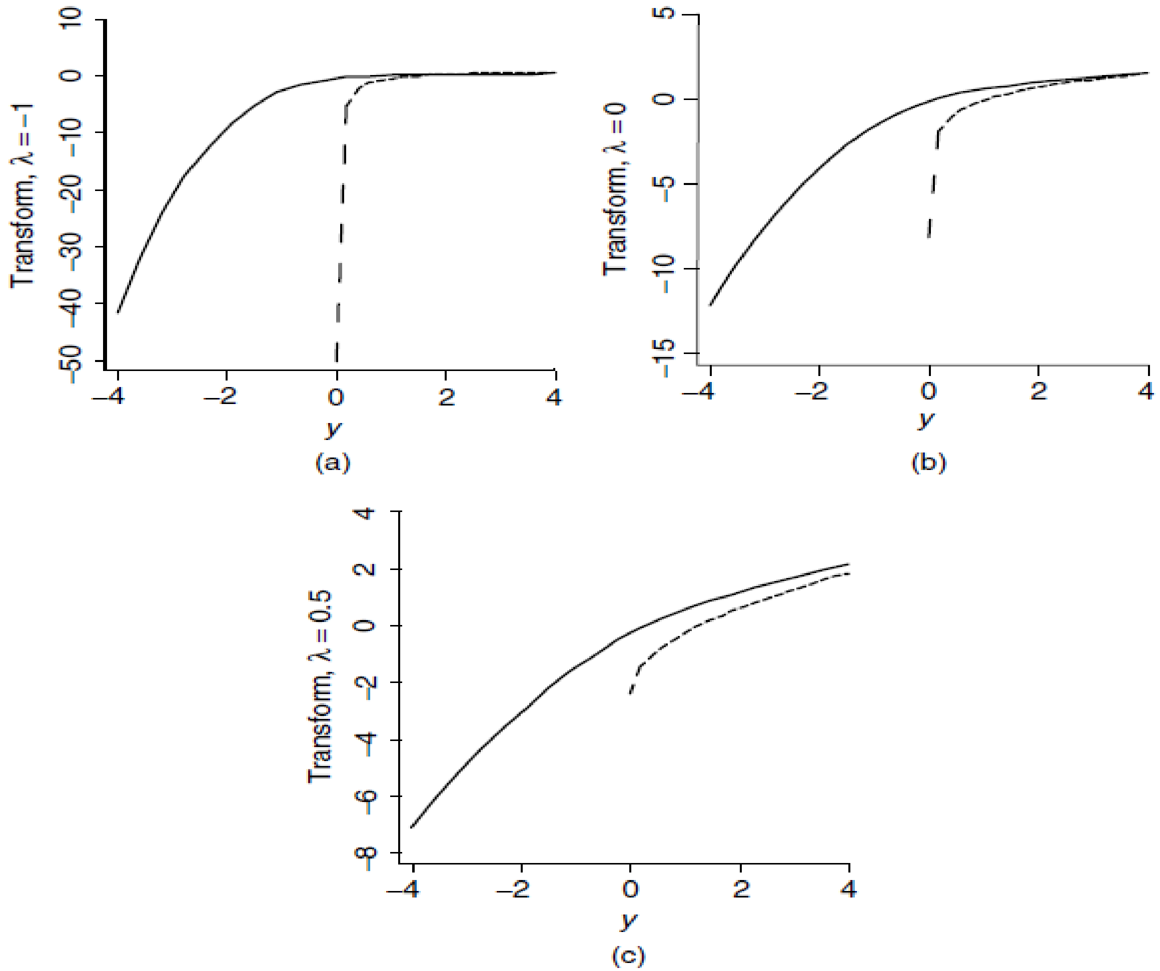

FIG. 7.9 Comparison of Box–Cox (dashed lines) and Yeo–Johnson (solid lines) power transformations for $\lambda = -1, 0, 0.5$. The Box–Cox transformations and Yeo–Johnson transformations behave differently for values of y close to zero.

Thank You !