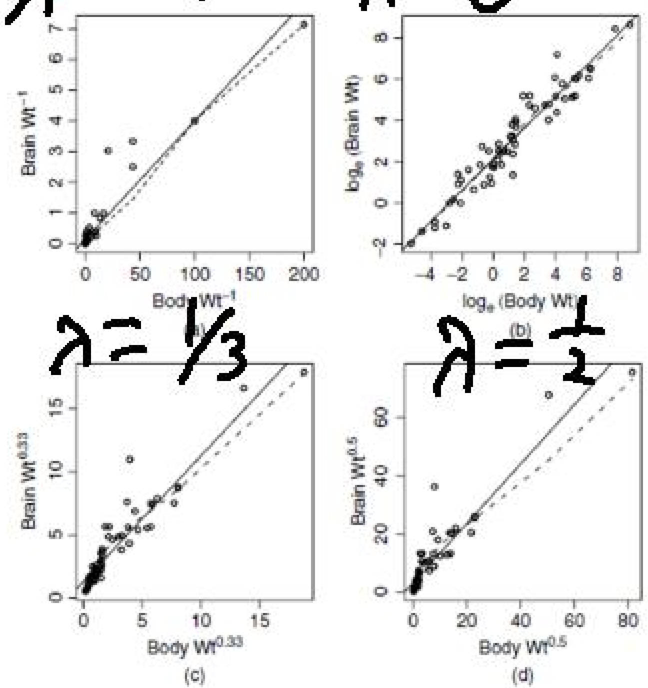
$\lambda \to 0 \quad \psi(U,\lambda) \to 1$ but not $\log U$

Power transformation

- Power transformation

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

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



- Want a linear relationship

$$\underline{\psi(BrainWt,\lambda)} = \alpha + \beta\underline{\psi(BodyWt,\lambda)} + e$$

transformed Y   transformed X

- λ=
  - a) -1
  - b) 0 (i.e. log U)
  - c) 0.33
  - d) 0.5
- Which λ will you choose?

Interpretation

$$\boxed{\psi(BrainWt,\lambda)=\alpha+\beta\psi(BodyWt,\lambda)+e}$$

- λ > 0

$$(BrainWt)^{\lambda}=\alpha+\beta(BodyWt)^{\lambda}+e$$

  - Artificial , usually has no physical meaning

- λ = 0 : log transformation
  - Corresponding to a physical model – allometric model

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

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

e^{α}

$\exp(e)$

Multiplicative error

Improving Power transformation

- Power transformation

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

$$\frac{d}{d\lambda} X^{\lambda} = (\log X)\, X^{\lambda}$$

- Scaled power transformation

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

- Advantage
  - Continuous function of λ : $\lim_{\lambda \to 0} \frac{X^{\lambda}-1}{\lambda} = \log(X)$ L'Hospital rule
  - Preserve the direction of association
    - True model : $E(Y \mid X) = \beta X^{-\frac{1}{2}}$ (negative assocation b/w Y and X)
    - Power transform: $E(Y \mid X) = \beta \psi(X, -1/2)$ (positive assocation b/w Y and ψ)
    - Scaled power transform:

$$E(Y \mid X) = \beta - \frac{1}{2} \underbrace{\beta \psi_s(X, -\frac{1}{2})}$$ (negative assocation b/w y and $\psi_s(X, -\frac{1}{2})$)

$$= \beta - \frac{\beta}{2} \frac{X^{-\frac{1}{2}}-1}{-1/2} = \beta X^{-\frac{1}{2}}$$

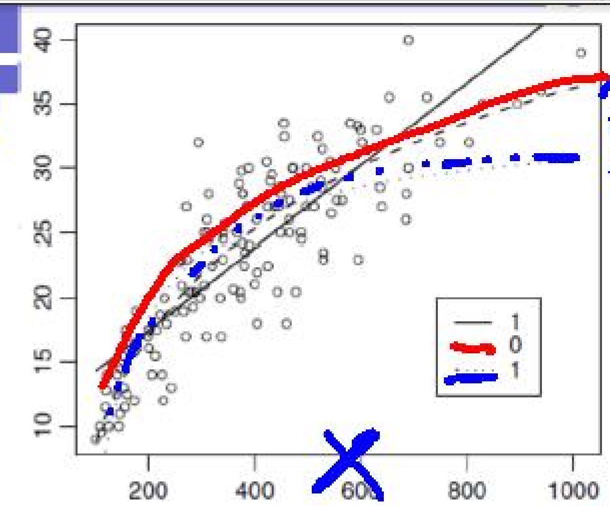
- Method 1: Draw many fitted curves

$\lambda = -1$

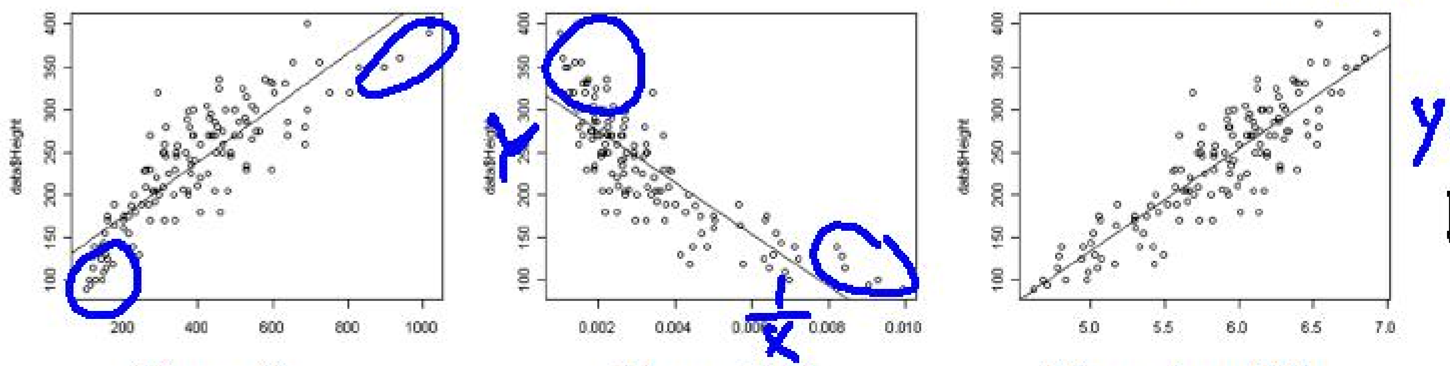
$Y = \beta_0 + \beta_1 \frac{1}{X} + e$

$i.e.$ plot $(x, \hat{y})$ for various x, where

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \psi(x, \lambda), \quad \lambda = -1, 0, 1 \ldots$$

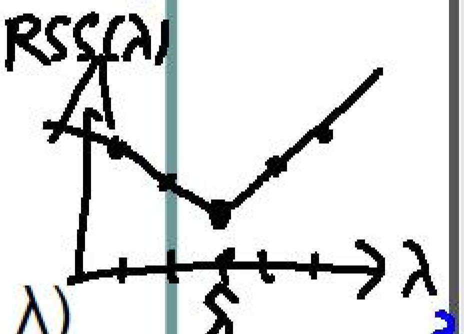


- Method 2: Draw many scatter plots



Y vs X Y vs 1/X Y vs log(X)

$y = \beta_0 + \beta_1 \log X + e$



- Method 3: plot λ against RSS of fitting y against ψ(X, λ) then find the λ that minimizes RSS. Or choose λ in the set (-1,-1/2,0,1,2)

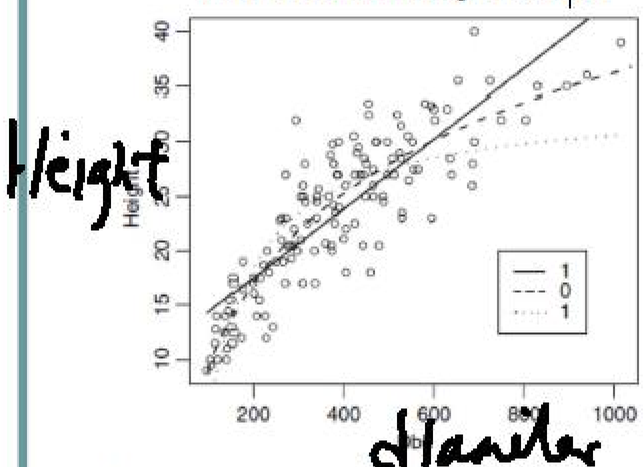
$\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 \psi(X_i, \lambda))^2$

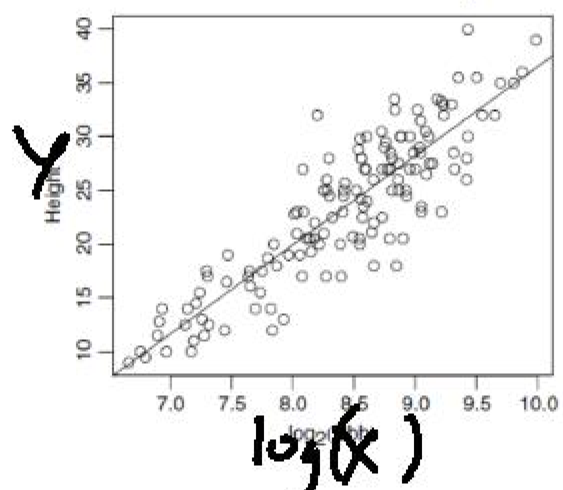
Example

- Y=Height of tree
- X=diameter of tree

M1: Draw many curves

M2: The best scatterplots

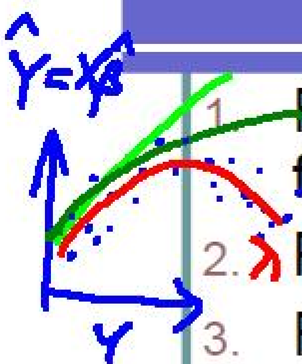




M3: Minimize RSS:

- **RSS(λ=0)=132.2, RSS(λ=1)=144.5, RSS(λ=-1)=254.8....**

Conclusion: Height = $\beta_o + \beta_1$ log(Diameter) + e



Inverse fitted value plot

1. Fit a linear regression between Y and X, get the fitted value $\hat{Y} = X\hat{\beta}$ ← Y & X may not be linearly related but $\hat{Y}$ & X are
2. Plot $\hat{Y}$ (y-axis) against Y (x-axis)
3. Fix a λ, fit $\hat{Y}$ against $\psi_s(Y,\lambda)$ and obtain
$$\boxed{\hat{Y}_\lambda = \hat{\beta}_0 + \hat{\beta}_1 \psi_s(Y,\lambda)}$$
4. Draw the fitted curve (Y, Y_λ) on the graph, see if it matches the pattern in 2).
   - Match → $\hat{\beta}_0 + \hat{\beta}_1 \psi_s(Y,\lambda) = \hat{Y}_\lambda \approx \hat{Y} = X\hat{\beta}$
5. Repeat 3)-4) to search for the best λ, say λ*

$$\boxed{\psi_s(Y,\lambda^*) \text{ and } X \text{ are linearly related} \Rightarrow \text{Regress } \psi_s(Y,\lambda^*) \text{ against } X}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad X \quad \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix}$$