

Transformation

power: $\psi(X, \lambda) = X^\lambda \leftarrow \log X \text{ when } \lambda=0$

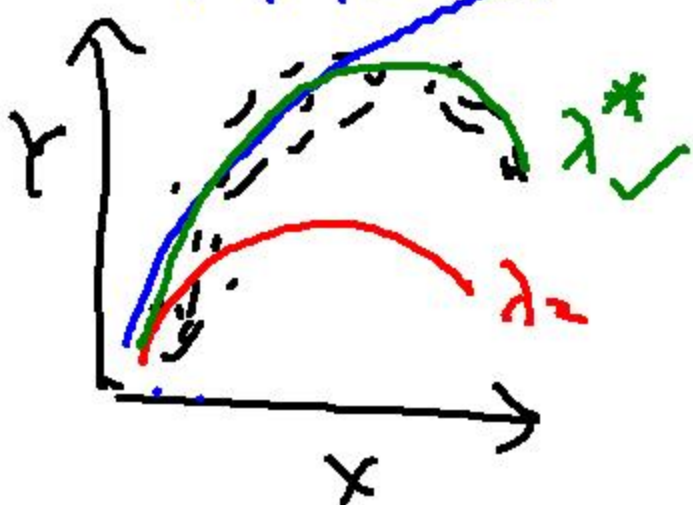
scaled power: $\psi_s(X, \lambda) = \frac{X^\lambda - 1}{\lambda}$ Adv: ① $\psi_s(X, \lambda)$ is continuous wrt λ

② $\psi_s(X, \lambda)$ preserve direction

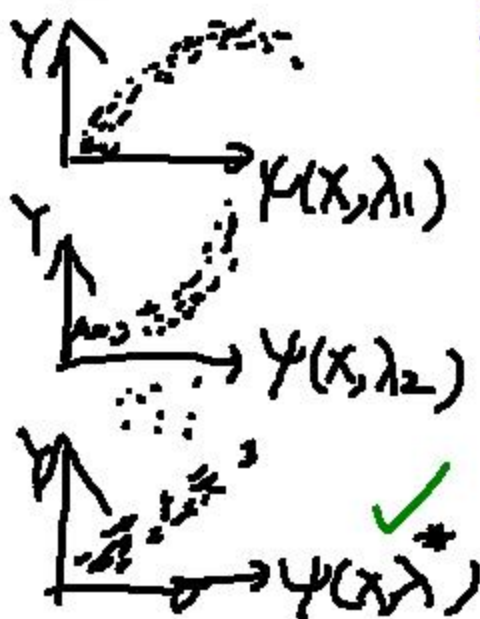
When there only 1 X variable

MI $\approx$ IFVP

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \psi(X, \lambda) \leftarrow$$

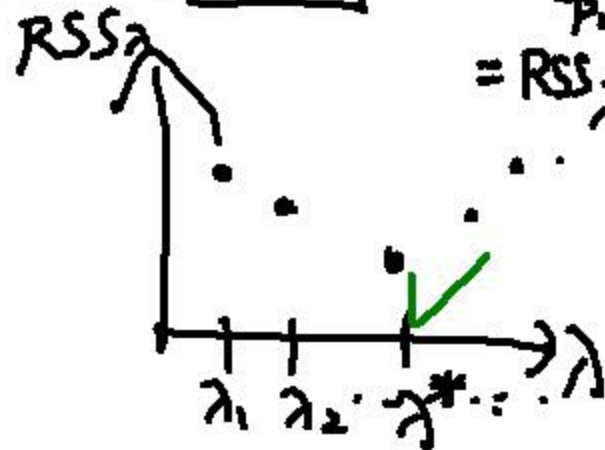


MI



$\psi_s \uparrow$ with Y
iff $X \uparrow$ with Y

Box Cox $\approx$ MI $\lambda \rightarrow \sum (Y - \beta_0 - \beta_1 \psi(X, \lambda))^2 = RSS_\lambda$



$(\hat{Y} \text{ vs } Y)$

-

$$\hat{Y}_\lambda = \hat{\beta}_0 + \hat{\beta}_1 \psi_s(Y, \lambda)$$

- ✓ match

- Match $\rightarrow \hat{\beta}_0 + \hat{\beta}_1 \psi_s(Y, \lambda) = \hat{Y}_\lambda \approx \hat{Y} = X\hat{\beta}$

5. Repeat 3)-4) to search for the best λ , say λ^*

$\psi_c(Y, \lambda^*)$ and X are linearly related $\Rightarrow$ Regress $\psi_c(Y, \lambda^*)$ against X

$$\psi_1(\gamma\lambda^*) = \frac{\hat{x}_\beta - \hat{\beta}_0}{\hat{\beta}_1} = x\hat{\alpha}$$

Chapter7.pdf - Adobe Reader

File Edit View Window Help

13 / 19

74.4%

Tools Sign Comment

power : $\psi(x, \lambda) = x^\lambda$

Scale power $\psi_s(x, \lambda) = \frac{x^\lambda - 1}{\lambda}$

Box-Cox transformation

1. Modified power family

$$\psi_M(Y, \lambda) = \psi_s(Y, \lambda) (\sqrt[n]{y_1 y_2 \dots y_n})^{1-\lambda}$$
$$= \begin{cases} (\sqrt[n]{y_1 y_2 \dots y_n})^{1-\lambda} \frac{Y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \sqrt[n]{y_1 y_2 \dots y_n} \log(Y) & \text{if } \lambda = 0 \end{cases}$$

x_{cm}
 x^λ
 cm^λ
 $cm^{1-\lambda}$
 cm

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

3. Model Assumption:

$$E(\psi_M(Y, \lambda) | X = x) = \beta' x \quad (*)$$

4. How to choose λ ?

• Fix a λ , fit model (*) for and obtain $RSS(\lambda)$

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

$$\Rightarrow \psi_M(Y, \lambda^*) = X\beta$$

$RSS(\lambda)$
 λ^*

11.69 x 8.27 in

Page 12 of 21

Recording 00:00:00

NEW PAGE

TEMPLATES

PREV

NEXT

PC MODE

POWERPOINT

BALL PEN

HIGHLIGHT

SIZE

COLOR

LINE

RECT.

CIRCLE

SELECT ERASE

ALL ERASE

REC

SAVE

OPEN

HELP

EXIT



Tools

Sign

Comment

$$E(\psi_M(Y, \lambda) | X = x) = \beta' x \quad (*)$$

Modified power family

geometric mean $\frac{\sum y_i}{n}$ - arithmetic mean

$$\psi_M(Y, \lambda) = \begin{cases} \left(\sqrt[n]{y_1 y_2 \dots y_n} \right)^{1-\lambda} \frac{Y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \sqrt[n]{y_1 y_2 \dots y_n} \log(Y) & \text{if } \lambda = 0 \end{cases}$$

```
highway.data=read.table("C:/highway.txt",header=T)
y=highway.data$Rate
n=length(y)
gm=prod(y)^{1/n}
```

#A) lambda=-1

Transform.A=-gm^2*(1/y-1)

fit.A=lm(Transform.A~log(ADT)+log(Trks)+Shld+log(Len),data=highway.data)

Rss.A=sum(fit.A\$residuals^2)

....

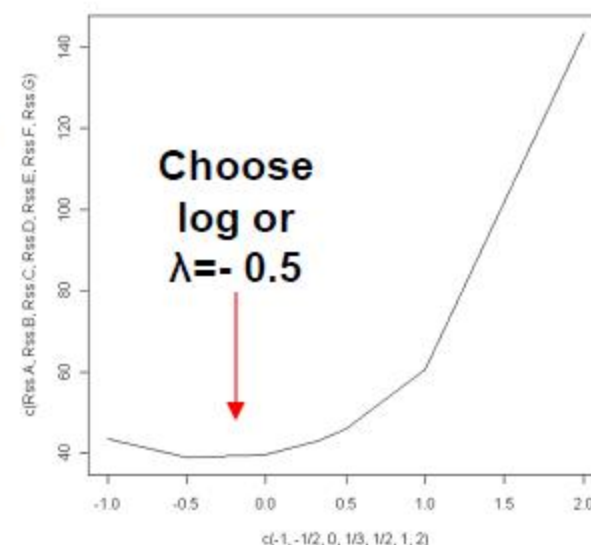
....

#G) lambda=2

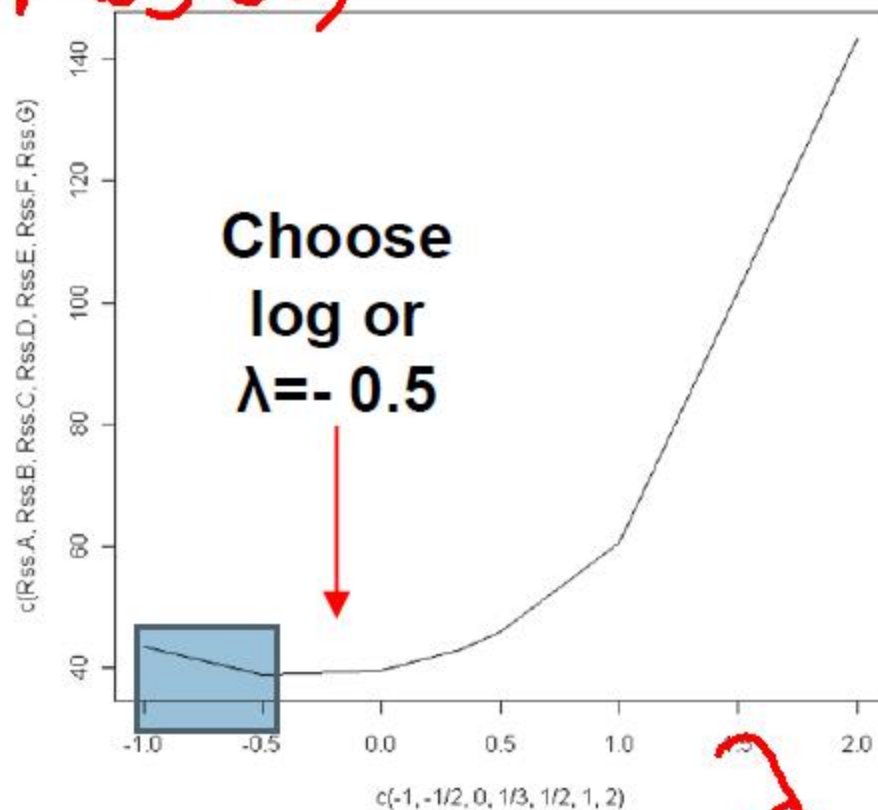
Transform.G=1/2/gm*(y^2-1)

fit.G=lm(Transform.G~log(ADT)+log(Trks)+Shld+log(Len),data=highway.data)

Rss.G=sum(fit.G\$residuals^2)



```
xt",header=T)
```

$$RSS(\lambda)$$


```
ks)+Shld+log(Len),data=highway.data)
```


$$F = \begin{cases} 1 & \text{gp 1} \\ 0 & \text{gp 2} \end{cases}$$

$$Y = \beta_0 + 3F + e \Rightarrow \begin{cases} Y = \beta_0 + 3 & \text{gp 1} \\ Y = \beta_0 & \text{gp 2} \end{cases}$$

$$\psi(F) = \begin{cases} 2 & \text{gp 1} \\ -1 & \text{gp 2} \end{cases}$$

$$Y = \beta_1 + \beta_2 \psi(F) + e \Rightarrow \begin{cases} Y = \beta_1 + 2\beta_2 & \text{gp 1} \\ Y = \beta_1 - \beta_2 & \text{gp 2} \end{cases}$$

$$\Rightarrow \begin{aligned} \beta_0 + 3 &= \beta_1 + 2\beta_2 \\ \beta_0 &= \beta_1 - \beta_2 \end{aligned} \Rightarrow \text{Solve 2 eq 2 unknown to get } \beta_1, \beta_2 \text{ in terms of } \beta_0$$

8.1. Regression Diagnostics

$$Y = X\beta + e, \quad e \sim N(0, \sigma^2)$$

- Regression diagnostic
 - Check if the assumptions (mean/var/error) are consistent with the observed data.
- Study the residuals
 - If the model works well, then the residuals matches the assumption
 - Mean zero, constant variance
 - i.e. residual plot looks like a null plot
 - what X cannot explain is random noise (no valid info)

Data

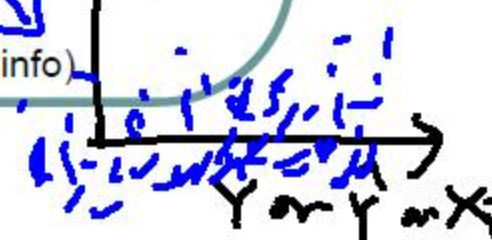
$$\hat{y} = x\hat{\beta}$$

$$\hat{e} = y - \hat{y}$$

Want

$$\hat{e} \approx e$$

$\hat{e}$



Chapter8.pdf - Adobe Reader

File Edit View Window Help

3 / 19

70.4%

Tools Sign Comment

8.1. Residuals

- Relationship between residuals and error

$$\begin{aligned}\hat{e} &= Y - \hat{Y} = \begin{pmatrix} (Y_1 - \hat{Y}_1) & (Y_1 - \hat{Y}_1) & \dots & (Y_n - \hat{Y}_n) \end{pmatrix}^T \\ &= Y - X\hat{\beta} \\ &= Y - X(X'X)^{-1}X'Y \quad (\hat{\beta} = (X'X)^{-1}X'Y) \\ &= (I - X(X'X)^{-1}X')Y \\ &= (I - H)Y, \quad \leftarrow Y = X\beta + e \quad \text{where } H = X(X'X)^{-1}X' \text{ is the Hat matrix} \\ &= (I - H)e, \quad \text{since } \underline{(I - H)X\beta} = X\beta - \underline{X(X'X)^{-1}X'X\beta} = 0\end{aligned}$$

(N(0, \sigma^2))

- To study the property of residual, we need to study the property of H.

NEW PAGE

TEMPLATES

PREV

Page 3 of 7

NEXT

PC MODE

POWERPOINT

BALL PEN

HIGH LIGHT

SIZE

COLOR

LINE

RECT.

CIRCLE

SELECT ERASE

ALL ERASE

REC

Recording 00:00:00

SAVE

OPEN

HELP

EXIT

Chapter8.pdf - Adobe Reader

File Edit View Window Help

4 / 19

70.4%

Tools Sign Comment

8.1. Residuals

- The **Hat matrix** $H = X(X'X)^{-1}X'$
 - H produces $\hat{Y}$ from Y
$$\hat{Y} = X\hat{\beta} = X \underbrace{(X'X)^{-1}X'}_{\hat{\beta}} Y = \underline{HY}$$
- Interpretation of projection
 - H projects Y onto the space of X
 - i.e. HY can be spanned (explained) by columns of X , using the weight $\hat{\beta}$:

$$\hat{Y} = HY = X\hat{\beta} = \hat{\beta}_0 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \hat{\beta}_1 \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{pmatrix} + \dots + \hat{\beta}_p \begin{pmatrix} x_{p1} \\ x_{p2} \\ \vdots \\ x_{pn} \end{pmatrix}$$

$(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$

$(1, x_1, \dots, x_p)$

$$(x_1 \dots x_n) = Y$$

$n\text{-dim space}$

11.69 x 8.27 in

NEW PAGE TEMPLATES PREV

Page 4 of 7

NEXT PC MODE POWERPOINT BALL PEN HIGH LIGHT SIZE COLOR LINE RECT. CIRCLE SELECT ERASE ALL ERASE REC

Recording 00:00:00

SAVE OPEN HELP EXIT

