Transformations and Normality

Merlise Clyde

STA721 Linear Models

Duke University

November 28, 2017

Outline

Topics

- Normality & Transformations
- Box-Cox
- Nonlinear Regression

Readings: Christensen Chapter 13 & Wakefield Chapter 6

Linear Model

Linear Model again:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Linear Model

Linear Model again:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

Linear Model

Linear Model again:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\boldsymbol{\mu} \in C(\mathbf{X}) \quad \Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

Linear Model

Linear Model again:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathrm{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

Linear Model

Linear Model again:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathsf{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

- Normal Distribution for $\boldsymbol{\epsilon}$ with constant variance
- Outlier Models
- Robustify with heavy tailed error distributions
- Computational Advantages of Normal Models

Normality

Recall

$$\mathbf{e} = (\mathbf{I} - \mathbf{P_X})\mathbf{Y}$$

[1] independent but not identically distributed

Normality

Recall

$$
\begin{aligned}
\mathbf{e} &= (\mathbf{I} - \mathbf{P_X})\mathbf{Y} \\
&= (\mathbf{I} - \mathbf{P_X})(\mathbf{X}\hat{\beta} + \epsilon)
\end{aligned}
$$

[1]independent but not identically distributed

Normality

Recall

$$\begin{aligned}\mathbf{e} &= (\mathbf{I} - \mathbf{P_X})\mathbf{Y} \\ &= (\mathbf{I} - \mathbf{P_X})(\mathbf{X}\hat{\beta} + \epsilon) \\ &= (\mathbf{I} - \mathbf{P_X})\epsilon\end{aligned}$$

[1] independent but not identically distributed

Normality

Recall

$$\begin{aligned}\mathbf{e} &= (\mathbf{I} - \mathbf{P_X})\mathbf{Y} \\ &= (\mathbf{I} - \mathbf{P_X})(\mathbf{X}\hat{\beta} + \epsilon) \\ &= (\mathbf{I} - \mathbf{P_X})\epsilon\end{aligned}$$

$$e_i = \epsilon_i - \sum_{j=1}^{n} h_{ij}\epsilon_j$$

[1] independent but not identically distributed

Normality

Recall

$$\begin{aligned} \mathbf{e} &= (\mathbf{I} - \mathbf{P_X})\mathbf{Y} \\ &= (\mathbf{I} - \mathbf{P_X})(\mathbf{X}\hat{\beta} + \epsilon) \\ &= (\mathbf{I} - \mathbf{P_X})\epsilon \end{aligned}$$

$$e_i = \epsilon_i - \sum_{j=1}^{n} h_{ij}\epsilon_j$$

Lyapunov CLT[1] implies that residuals will be approximately normal (even for modest n), if the errors are not normal

[1]independent but not identically distributed

Normality

Recall

$$\begin{aligned}\mathbf{e} &= (\mathbf{I} - \mathbf{P_X})\mathbf{Y} \\ &= (\mathbf{I} - \mathbf{P_X})(\mathbf{X}\hat{\beta} + \epsilon) \\ &= (\mathbf{I} - \mathbf{P_X})\epsilon\end{aligned}$$

$$e_i = \epsilon_i - \sum_{j=1}^{n} h_{ij}\epsilon_j$$

Lyapunov CLT[1] implies that residuals will be approximately normal (even for modest n), if the errors are not normal

"Supernormality of residuals"

[1]independent but not identically distributed

Q-Q Plots

- Order e_i: $e_{(1)} \leq e_{(2)} \ldots \leq e_{(n)}$ sample order statistics or sample quantiles

Q-Q Plots

- Order e_i: $e_{(1)} \leq e_{(2)} \ldots \leq e_{(n)}$ sample order statistics or sample quantiles
- Let $z_{(1)} \leq z_{(2)} \ldots z_{(n)}$ denote the expected order statistics of a sample of size n from a standard normal distribution "theoretical quantiles"

Q-Q Plots

- Order e_i: $e_{(1)} \leq e_{(2)} \ldots \leq e_{(n)}$ sample order statistics or sample quantiles
- Let $z_{(1)} \leq z_{(2)} \ldots z_{(n)}$ denote the expected order statistics of a sample of size n from a standard normal distribution "theoretical quantiles"
- If the e_i are normal then $\mathrm{E}[e_{(i)}] = \sigma z_{(i)}$

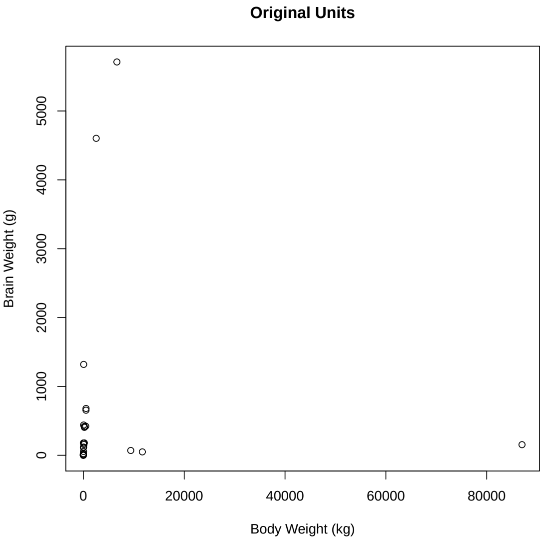
Q-Q Plots

- Order e_i: $e_{(1)} \leq e_{(2)} \ldots \leq e_{(n)}$ sample order statistics or sample quantiles
- Let $z_{(1)} \leq z_{(2)} \ldots z_{(n)}$ denote the expected order statistics of a sample of size n from a standard normal distribution "theoretical quantiles"
- If the e_i are normal then $\mathrm{E}[e_{(i)}] = \sigma z_{(i)}$
- Expect that points in a scatter plot of $e_{(i)}$ and $z_{(i)}$ should be on a straight line.

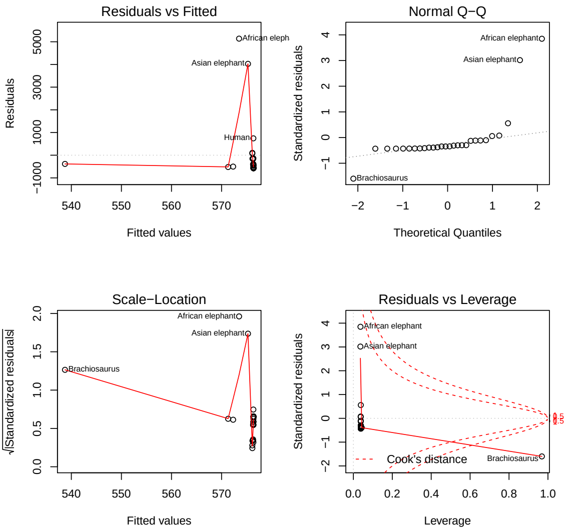
Q-Q Plots

- Order e_i: $e_{(1)} \leq e_{(2)} \ldots \leq e_{(n)}$ sample order statistics or sample quantiles
- Let $z_{(1)} \leq z_{(2)} \ldots z_{(n)}$ denote the expected order statistics of a sample of size n from a standard normal distribution "theoretical quantiles"
- If the e_i are normal then $\mathrm{E}[e_{(i)}] = \sigma z_{(i)}$
- Expect that points in a scatter plot of $e_{(i)}$ and $z_{(i)}$ should be on a straight line.
- Judgment call - use simulations to gain experience!

Animal Example



Residual Plots



Box-Cox Transformation

Box and Cox (1964) suggested a family of power transformations for $Y > 0$

Box-Cox Transformation

Box and Cox (1964) suggested a family of power transformations for $Y > 0$

$$U(\mathbf{Y}, \lambda) = Y^{(\lambda)} = \begin{cases} \frac{(Y^{\lambda} - 1)}{\lambda} & \lambda \neq 0 \\ \log(Y) & \lambda = 0 \end{cases}$$

Box-Cox Transformation

Box and Cox (1964) suggested a family of power transformations for $Y > 0$

$$U(\mathbf{Y}, \lambda) = Y^{(\lambda)} = \begin{cases} \frac{(Y^{\lambda} - 1)}{\lambda} & \lambda \neq 0 \\ \log(Y) & \lambda = 0 \end{cases}$$

- Estimate λ by maximum Likelihood

Box-Cox Transformation

Box and Cox (1964) suggested a family of power transformations for $Y > 0$

$$U(\mathbf{Y}, \lambda) = Y^{(\lambda)} = \begin{cases} \frac{(Y^{\lambda} - 1)}{\lambda} & \lambda \neq 0 \\ \log(Y) & \lambda = 0 \end{cases}$$

- Estimate λ by maximum Likelihood

$$\mathcal{L}(\lambda, \boldsymbol{\beta}, \sigma^2) \propto \prod f(y_i \mid \lambda, \boldsymbol{\beta}, \sigma^2)$$

- $U(\mathbf{Y}, \lambda) = Y^{(\lambda)} \sim \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2)$

Box-Cox Transformation

Box and Cox (1964) suggested a family of power transformations for $Y > 0$

$$U(\mathbf{Y}, \lambda) = Y^{(\lambda)} = \begin{cases} \frac{(Y^\lambda - 1)}{\lambda} & \lambda \neq 0 \\ \log(Y) & \lambda = 0 \end{cases}$$

- Estimate λ by maximum Likelihood

$$\mathcal{L}(\lambda, \boldsymbol{\beta}, \sigma^2) \propto \prod f(y_i \mid \lambda, \boldsymbol{\beta}, \sigma^2)$$

- $U(\mathbf{Y}, \lambda) = Y^{(\lambda)} \sim \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2)$
- Jacobian term is $\prod_i y_i^{\lambda - 1}$ for all λ

Box-Cox Transformation

Box and Cox (1964) suggested a family of power transformations for $Y > 0$

$$U(\mathbf{Y}, \lambda) = Y^{(\lambda)} = \begin{cases} \frac{(Y^{\lambda}-1)}{\lambda} & \lambda \neq 0 \\ \log(Y) & \lambda = 0 \end{cases}$$

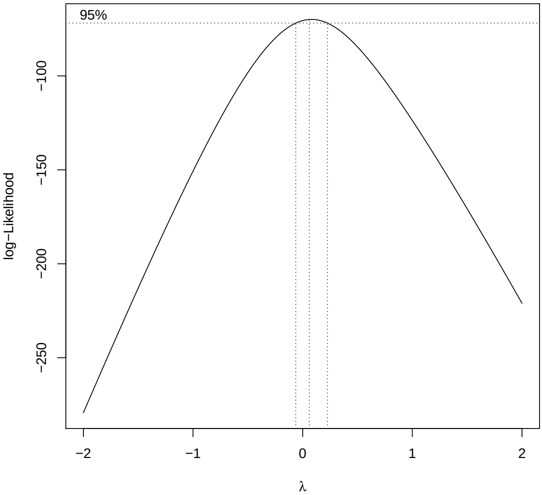
- Estimate λ by maximum Likelihood

$$\mathcal{L}(\lambda, \boldsymbol{\beta}, \sigma^2) \propto \prod f(y_i \mid \lambda, \boldsymbol{\beta}, \sigma^2)$$

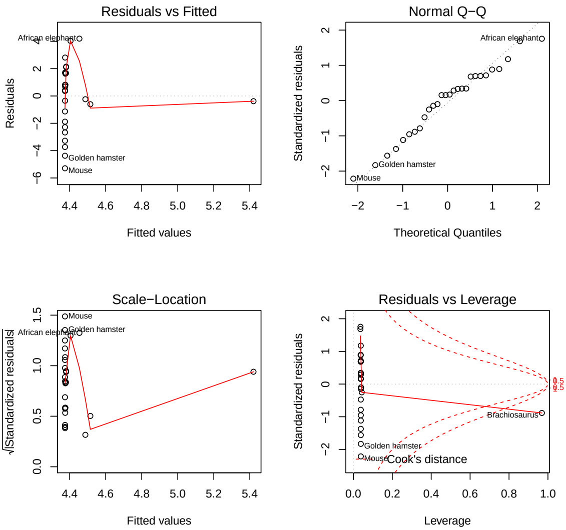
- $U(\mathbf{Y}, \lambda) = Y^{(\lambda)} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2)$
- Jacobian term is $\prod_i y_i^{\lambda-1}$ for all λ
- Profile Likelihood based on substituting MLE $\boldsymbol{\beta}$ and σ^2 for each value of λ is

$$\log(\mathcal{L}(\lambda) \propto (\lambda - 1) \sum_i \log(Y_i) - \frac{n}{2} \log(\mathrm{SSE}(\lambda))$$

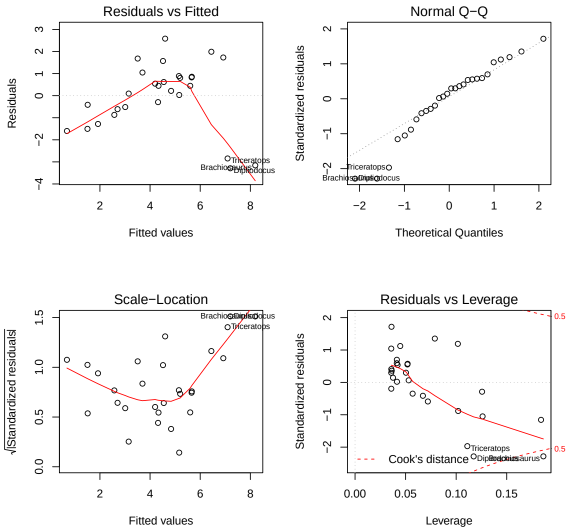
Profile Likelihood



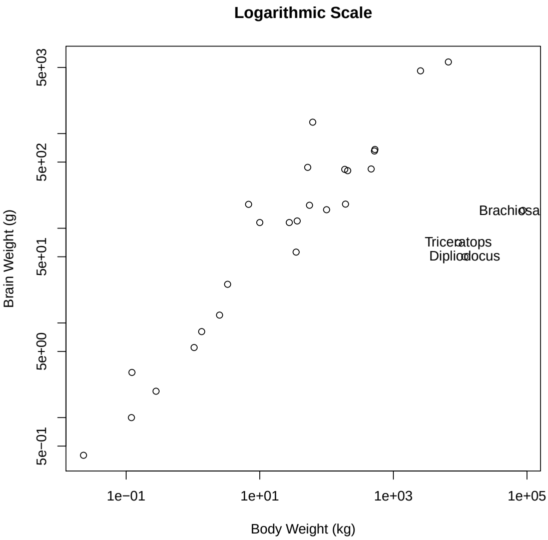
Residuals After Transformation of Response



Residuals After Transformation of Both



Transformed Data



Test that Dinos are Outliers

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	23	12.12				
2	26	60.99	-3	-48.87	30.92	0.0000

Test that Dinos are Outliers

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	23	12.12				
2	26	60.99	-3	-48.87	30.92	0.0000

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.1504	0.2006	10.72	0.0000
log(body)	0.7523	0.0457	16.45	0.0000
Triceratops	-4.7839	0.7913	-6.05	0.0000
Brachiosaurus	-5.6662	0.8328	-6.80	0.0000
Dipliodocus	-5.2851	0.7949	-6.65	0.0000

Test that Dinos are Outliers

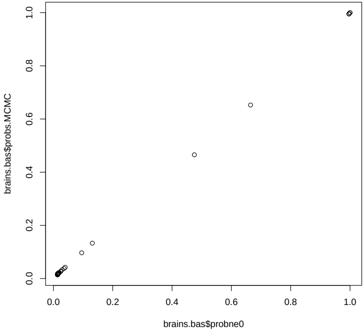
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	23	12.12				
2	26	60.99	-3	-48.87	30.92	0.0000

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.1504	0.2006	10.72	0.0000
log(body)	0.7523	0.0457	16.45	0.0000
Triceratops	-4.7839	0.7913	-6.05	0.0000
Brachiosaurus	-5.6662	0.8328	-6.80	0.0000
Dipliodocus	-5.2851	0.7949	-6.65	0.0000

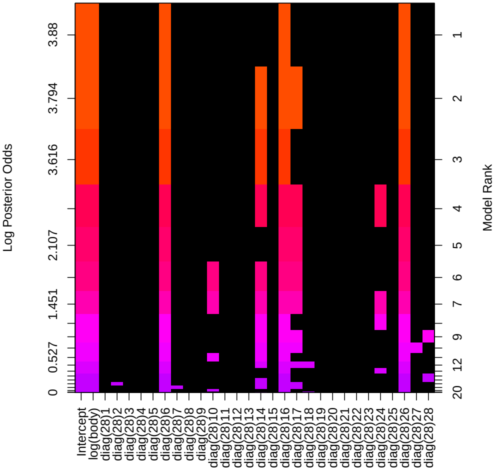
Dinosaurs come from a different population from mammals

Model Selection Priors

```
brains.bas = bas.lm(log(brain) ~ log(body) + diag(28),
   data=Animals,  prior="hyper-g-n", a=3,
   modelprior=beta.binomial(1,28),
   method="MCMC", n.models=2^17, MCMC.it=2^18)
# check for convergence
plot(brains.bas$probne0, brains.bas$probs.MCMC)
```



image(brains.bas)




```
rownames(Animals)[c(6, 14, 16, 26)]
"Dipliodocus" "Human" "Triceratops" "Brachiosaurus"
```

Variance Stabilizing Transformations

- If $Y - \mu$ (approximately) $N(0, h(\mu))$

Variance Stabilizing Transformations

- If $Y - \mu$ (approximately) $N(0, h(\mu))$
- Delta Method implies that

$$g(Y) \dot{\sim} \text{N}(g(\mu), g'(\mu)^2 h(\mu)$$

Variance Stabilizing Transformations

- If $Y - \mu$ (approximately) $N(0, h(\mu))$
- Delta Method implies that

$$g(Y) \dot{\sim} \mathrm{N}(g(\mu), g'(\mu)^2 h(\mu)$$

- Find function g such that $g'(\mu)^2 / h(\mu)$ is constant

$$g(Y) \sim N(g(\mu), c)$$

Variance Stabilizing Transformations

- If $Y - \mu$ (approximately) $N(0, h(\mu))$
- Delta Method implies that

$$g(Y) \dot{\sim} \mathrm{N}(g(\mu), g'(\mu)^2 h(\mu)$$

- Find function g such that $g'(\mu)^2/h(\mu)$ is constant

$$g(Y) \sim N(g(\mu), c)$$

- Poisson Counts ($Y > 3$): g is square root transformation

Variance Stabilizing Transformations

- If $Y - \mu$ (approximately) $N(0, h(\mu))$
- Delta Method implies that

$$g(Y) \dot{\sim} \mathrm{N}(g(\mu), g'(\mu)^2 h(\mu)$$

- Find function g such that $g'(\mu)^2/h(\mu)$ is constant

$$g(Y) \sim N(g(\mu), c)$$

- Poisson Counts ($Y > 3$): g is square root transformation
- Binomial: $\arcsin(\sqrt{Y})$

Note: transformation for normality may not be the same as the variance stabilizing transformation; boxcox assumes mean function is correct

Variance Stabilizing Transformations

- If $Y - \mu$ (approximately) $N(0, h(\mu))$
- Delta Method implies that

$$g(Y) \dot{\sim} \mathrm{N}(g(\mu), g'(\mu)^2 h(\mu)$$

- Find function g such that $g'(\mu)^2 / h(\mu)$ is constant

$$g(Y) \sim N(g(\mu), c)$$

- Poisson Counts ($Y > 3$): g is square root transformation
- Binomial: $\arcsin(\sqrt{Y})$

Note: transformation for normality may not be the same as the variance stabilizing transformation; boxcox assumes mean function is correct

Generalized Linear Models preferable

Nonlinear Models

Drug concentration of caldralazine at time X_i in a cardiac failure patient given a single 30mg dose ($D = 30$) given by

Nonlinear Models

Drug concentration of caldralazine at time X_i in a cardiac failure patient given a single 30mg dose ($D = 30$) given by

$$\mu(\boldsymbol{\beta}) = \left[\frac{D}{V}\exp(-\kappa_e x_i)\right]$$

with $\boldsymbol{\beta} = (V, \kappa_e)$ $V = \mathit{volume}$ and κ_e is the elimination rate

Nonlinear Models

Drug concentration of caldralazine at time X_i in a cardiac failure patient given a single 30mg dose ($D = 30$) given by

$$\mu(\boldsymbol{\beta}) = \left[\frac{D}{V}\exp(-\kappa_e x_i)\right]$$

with $\boldsymbol{\beta} = (V, \kappa_e)$ $V = volume$ and κ_e is the elimination rate

If $\log(Y_i) = \log(\mu(\boldsymbol{\beta})) + \epsilon_i$ with $\epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$ then the model is intrinisically linear (can transform to linear model)

$$\begin{aligned}
\log(\mu(\boldsymbol{\beta})) &= \log\left[\frac{D}{V}\exp(-\kappa_e x_i)\right] \\
&= \log[D] - \log(V) - \kappa_e x_i \\
log(Y_i) - \log[30] &= \beta_0 + \beta_1 x_i + \epsilon_i
\end{aligned}$$

Nonlinear Models

Drug concentration of caldralazine at time X_i in a cardiac failure patient given a single 30mg dose ($D = 30$) given by

$$\mu(\boldsymbol{\beta}) = \left[\frac{D}{V}\exp(-\kappa_e x_i)\right]$$

with $\boldsymbol{\beta} = (V, \kappa_e)$ $V = \mathit{volume}$ and κ_e is the elimination rate

If $\log(Y_i) = \log(\mu(\boldsymbol{\beta})) + \epsilon_i$ with $\epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$ then the model is intrinisically linear (can transform to linear model)

$$\begin{aligned} \log(\mu(\boldsymbol{\beta})) &= \log\left[\frac{D}{V}\exp(-\kappa_e x_i)\right] \\ &= \log[D] - \log(V) - \kappa_e x_i \\ log(Y_i) - \log[30] &= \beta_0 + \beta_1 x_i + \epsilon_i \end{aligned}$$

Nonlinear Least Squares

```
> conc.nlm = nls( log(y) ~ log((30/V)*exp(-k*x)),
                  data=df, start=list(V=vhat, k=khat))
> summary(conc.nlm)
Formula: log(y) ~ log((30/V) * exp(-k * x))
Parameters:
                Estimate Std. Error t value Pr(>|t|)
V.(Intercept) 16.66331    7.11923   2.341 0.057796 .
k.x            0.15211    0.02368   6.423 0.000673 ***

Residual standard error: 0.7411 on 6 degrees of freedom
Number of iterations to convergence: 0
Achieved convergence tolerance: 3.978e-09
```

Additive Errors

- under multiplicative log normal errors model is equivalent to linear model

Additive Errors

- under multiplicative log normal errors model is equivalent to linear model
- with additive Gaussian errors (or other distributions) model is intrinsically nonlinear - nonlinear least squares (or posterior sampling)

Additive Errors

- under multiplicative log normal errors model is equivalent to linear model
- with additive Gaussian errors (or other distributions) model is intrinsically nonlinear - nonlinear least squares (or posterior sampling)

$$Y_i = (30/V) * exp(-k * x_i) + \epsilon_i$$

Additive Errors

- under multiplicative log normal errors model is equivalent to linear model
- with additive Gaussian errors (or other distributions) model is intrinsically nonlinear - nonlinear least squares (or posterior sampling)

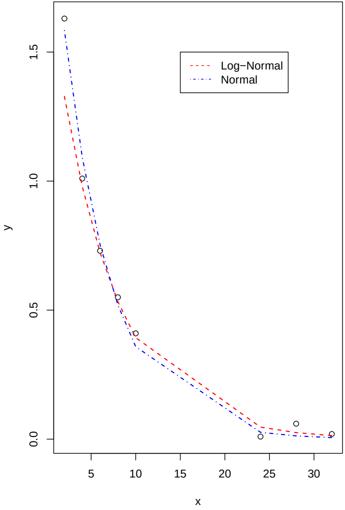
$$Y_i = (30/V) * exp(-k * x_i) + \epsilon_i$$

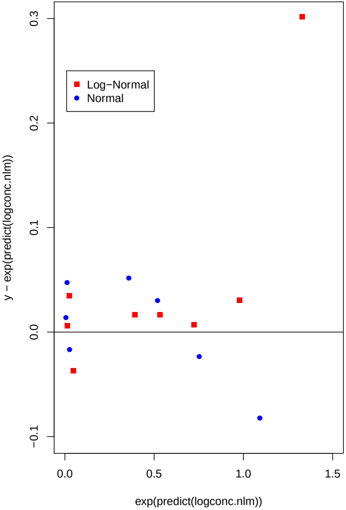
$$\epsilon_i \overset{\text{iid}}{\sim} \mathsf{N}(0, \sigma^2)$$

Intrinsically Nonlinear Model

```
> summary(conc.nlm)
Formula: y ~ (30/V) * exp(-k * x)
Parameters:
  Estimate Std. Error t value Pr(>|t|)
V 13.06506    0.60899   21.45 6.69e-07 ***
k  0.18572    0.01124   16.52 3.14e-06 ***
---
Residual standard error: 0.05126 on 6 degrees of freedom
Number of iterations to convergence: 4
Achieved convergence tolerance: 7.698e-06
```

Fitted Values & Residuals





Functions of Interest

Interest is in

- clearance: $V\kappa_e$

Functions of Interest

Interest is in

- clearance: $V\kappa_e$
- elimination half-life $x_{1/2} = \log 2/\kappa_e$

Functions of Interest

Interest is in

- clearance: $V\kappa_e$
- elimination half-life $x_{1/2} = \log 2/\kappa_e$
- Use properties of MLEs: asymptotically $\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, I(\hat{\boldsymbol{\beta}})^{-1}\right)$

Functions of Interest

Interest is in

- clearance: $V\kappa_e$
- elimination half-life $x_{1/2} = \log 2/\kappa_e$

- Use properties of MLEs: asymptotically $\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, I(\hat{\boldsymbol{\beta}})^{-1}\right)$
- (Multivariate) Delta Method for transformations

Functions of Interest

Interest is in

- clearance: $V\kappa_e$
- elimination half-life $x_{1/2} = \log 2/\kappa_e$

- Use properties of MLEs: asymptotically $\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, I(\hat{\boldsymbol{\beta}})^{-1}\right)$
- (Multivariate) Delta Method for transformations
- Asymptotic Distributions

Bayes obtain the posterior directly for parameters and functions of parameters! Priors? Constraints on Distributions?

Summary

- Optimal transformation for normality (MLE) depends on choice of mean function

Summary

- Optimal transformation for normality (MLE) depends on choice of mean function
- May not be the same as the variance stabilizing transformation

Summary

- Optimal transformation for normality (MLE) depends on choice of mean function
- May not be the same as the variance stabilizing transformation
- Nonlinear Models as suggested by Theory or Generalized Linear Models are alternatives

Summary

- Optimal transformation for normality (MLE) depends on choice of mean function
- May not be the same as the variance stabilizing transformation
- Nonlinear Models as suggested by Theory or Generalized Linear Models are alternatives
- "normal" estimates may be useful approximations for large p or for starting values for more complex models (where convergence may be sensitive to starting values)