Linear Models and Assumptions

AGRON 590 MG: Crop-Soil Modeling

Fernando E. Miguez

Iowa State University

Oct 29, 2010

Matrix Algebra

A vector

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

Matrix Algebra

A matrix

$$\mathbf{X} = \begin{bmatrix} x_{10} & x_{11} & x_{12} & x_{13} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ x_{30} & x_{31} & x_{32} & x_{33} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n0} & x_{n1} & x_{n2} & x_{n3} \end{bmatrix}$$

Linear Model

$$y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

In a linear model the second partial derivative with respect to the parameter is zero.

Linear Model in matrix notation

$$y = X\beta + \epsilon$$

$y =$ response
$X =$ inputs of explanatory variables
$\beta =$ parameters
$\epsilon =$ error random

Linear Mixed Model in matrix notation

$$y = X\beta + Zu + \epsilon$$

y = response

X = inputs of explanatory variables

β = parameters

Z = zeta

u = random component

ϵ = error random

Assumptions

1. Errors have zero mean
2. Errors are normally distributed
3. Errors have equal variance
4. Errors are independent

What does this mean in terms of the previous model?

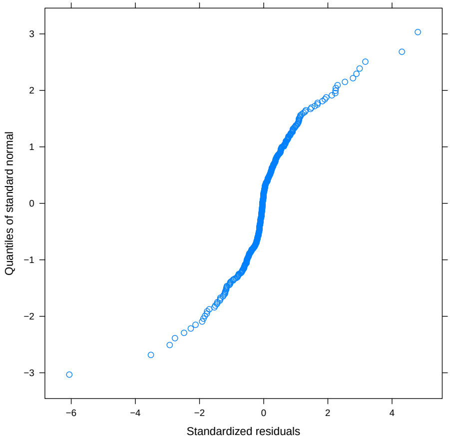
What does this mean in terms of the previous model?

1. Errors have zero mean easy to check

$$mean(\epsilon) = 0$$

What does this mean in terms of the previous model?

1. Errors have zero mean easy to check
2. Errors are normally distributed visual check



Covariance Interlude

$$Y_1 \sim \mathcal{N}(0, \sigma_1^2)$$
$$Y_2 \sim \mathcal{N}(0, \sigma_2^2)$$
$$Cov(Y_1, Y_2) = \sigma_{12}$$

Relationship between correlation and covariance

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_1^2 \sigma_2^2}}$$

Covariance Interlude
In matrix form

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathcal{N}(0, \Sigma)$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Covariance Interlude
In matrix form

Extension to three (or more) variables is straight forward

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \sim \mathcal{N}(0, \Sigma)$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

What does this mean in terms of the previous model?

1. Errors have zero mean easy to check
2. Errors are normally distributed visual check
3. Errors have equal variance

Here we are not yet assuming homogeneous variances $(n = 4)$

$$Cov(\epsilon) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix}$$

What does this mean in terms of the previous model?

1. Errors have zero mean easy to check
2. Errors are normally distributed visual check
3. Errors have equal variance

Here we assume homogeneous variances (on the diagonal)

$$Cov(\epsilon) = \begin{bmatrix} \sigma & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma \end{bmatrix}$$

What does this mean in terms of the previous model?

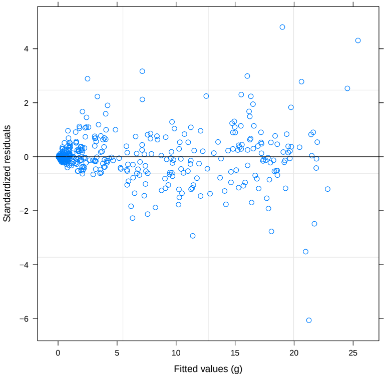
1. Errors have zero mean easy to check
2. Errors are normally distributed visual check
3. Errors have equal variance
4. Errors are independent

Here we assume homogeneous variances and independence

$$Cov(\epsilon) = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

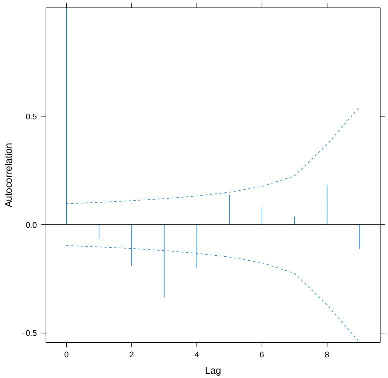
What does this mean in terms of the previous model?

1. Errors have zero mean easy to check
2. Errors are normally distributed visual check
3. Errors have equal variance visual check
4. Errors are independent



What does this mean in terms of the previous model?

1. Errors have zero mean easy to check
2. Errors are normally distributed visual check
3. Errors have equal variance visual check
4. Errors are independent visual check



Assumptions for the random effects

1. random effects have zero mean
2. random effects are normally distributed

Random effects enter linearly in the non-linear model

$$y_i = \frac{Asym + b_{1i}}{1 + \exp(((xmid + b_{2i}) - x)/(scal + b_{3i}))} + \epsilon$$

Assumptions for the random effects

1. random effects have zero mean
2. random effects are normally distributed

Random effects enter linearly in the non-linear model

$$y_i = \frac{Asym + b_{1i}}{1 + \exp(((xmid + b_{2i}) - x)/(scal + b_{3i}))} + \epsilon$$

$$\boldsymbol{b_i} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi}), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

$$\mathbf{\Psi} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

References

- Mixed Effects Models in S and S-plus. Pinheiro and Bates (2000).
- Matrix Algebra Useful for Statistics. Searle.
- Modern Applied Statistics with S. Venables and Ripley. (2002)