Mixed Models Intro

Revisiting Coefficient of Regression

$$b_{y|x} = \frac{Cov(x, y)}{V(x)}$$

- We start here because it is a simple representation of a linear predictor of y given x.
- We will expand on this to derive estimators for multiple linear regression.

Variance and Covariance

The variance of x

$$V(x) = E[(x - \mu_x)^2]$$

Using Theorem 1 you can show that:

$$V(x) = E[x^2] - \mu_x^2$$

Estimator in vector form:

$$\frac{(\mathbf{x} - \mu_x)'\,(\mathbf{x} - \mu_x)}{n - 1}$$

Covariance

If we take 2 random variables x and y, the covariance of x and y is:

$$Cov(x, y) = E[(x - \mu_x)(y - \mu_y)]$$

Again using Theorem 1 you can show that:

$$Cov(x, y) = E[xy] - E[x]E[y]$$

Estimator in vector form:

$$\frac{(\mathbf{x} - \mu_x)' (\mathbf{y} - \mu_y)}{n - 1}$$

Coefficient of Regression – Vector Notation

$$b_{y|x} = \frac{Cov(x,y)}{V(\mathrm{x})}$$

Given the vectors $\mathbf{x}$ and $\mathbf{y}$ are centered (i.e. $\mathrm{E}[\mathbf{x}] = \mathrm{E}[\mathbf{y}] = 0$):

$$b_{y|x} = \frac{\left(\frac{\mathbf{x}'\mathbf{y}}{n-1}\right)}{\left(\frac{\mathbf{x}'\mathbf{x}}{n-1}\right)} \Rightarrow \quad b_{y|x} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} \quad \Rightarrow \quad b_{y|x} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$$

Given: $\frac{\mathbf{1}}{\mathbf{x}'\mathbf{x}} = \mathbf{x}'\mathbf{x}^{-1}$

Ordinary Least Squares (OLS) and Maximum Likelihood Estimators (MLE)

Ordinary Least Squares are derived by minimizing:

$$\mathrm{e}'\mathrm{e}$$

$$(\mathrm{y} - \mathrm{X}\hat{\mathrm{b}})'(\mathrm{y} - \mathrm{X}\hat{\mathrm{b}})$$

Maximum Likelihood Estimators are derived by maximizing

$$\prod_{1}^{n} 2\sigma_e^2\pi^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_e^2}(y_i - \widehat{y_i})^2}$$

OLS

To minimize we take the derivative [1] with respect to $\hat{b}$, set it equal to zero, and solve

$$(y - X\hat{b})'(y - X\hat{b}) \qquad [1]$$

$$X'X\hat{b} = X'y$$

$$\hat{b} = (X'X)^{-1}X'y$$

Best -> min(V($\hat{\mathrm{b}}$))

Linear -> $\hat{\mathrm{b}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'y$

ex. non-linear -> $\hat{\mathrm{b}} = y^{x}$

Unbiased -> $\mathrm{E}[\hat{\mathrm{b}}] = \mathrm{E}[\mathrm{b}\,] = \mathrm{b}$

Variance of $\hat{\mathrm{b}}$

From Theorem 1

$$V(ax) = a^2 V(x)$$

In matrix form

$$V(\mathrm{A}x) = \mathrm{A}V(x)\mathrm{A}'$$

Variance of $\hat{b}$

$$\hat{b} = (X'X)^{-1}X'y$$

$$V(\hat{b}) = V((X'X)^{-1}X'y)$$

$$V(\hat{b}) = (X'X)^{-1}X'V(y)[(X'X)^{-1}X']'$$

$$V(\hat{b}) = (X'X)^{-1}X'V(y)X(X'X)^{-1}$$

The density of a normal distribution

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

In the simple fixed effect mode e_{ijk} is a random variable a $N(0,\sigma_e^2)$ distribution.

$$\frac{1}{\sigma_e\sqrt{2\pi}}e^{-\frac{(y_i-\widehat{y_i})^2}{2\sigma_e^2}}$$

$$2\sigma_e^2\pi^{-\frac{1}{2}}e^{-\frac{1}{2\sigma_e^2}(y_i-\widehat{y_i})^2}$$

The Likelihood

Each realized value from a normal distribution has an associated density from the probability density function (PDF).

The PDF reflects how likely you are to observe that value given the mean and variance of the distribution

$$2\sigma_e^2\pi^{-\frac{1}{2}}e^{-\frac{1}{2\sigma_e^2}(y_1-\widehat{y_1})^2}, 2\sigma_e^2\pi^{-\frac{1}{2}}e^{-\frac{1}{2\sigma_e^2}(y_2-\widehat{y_2})^2}, 2\sigma_e^2\pi^{-\frac{1}{2}}e^{-\frac{1}{2\sigma_e^2}(y_3-\widehat{y_3})^2}$$

$$, \ldots , 2\sigma_e^2\pi^{-\frac{1}{2}}e^{-\frac{1}{2\sigma_e^2}(y_n-\widehat{y_n})^2}$$

The Likelihood

The likelihood function is the product of the density for each e_{ijk}

$$\prod_{1}^{n} 2\sigma_e^2\pi^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_e^2}(y_i - \widehat{y_i})^2}$$

$$2\sigma_e^2\pi^{-\frac{n}{2}} e^{-\frac{1}{2\sigma_e^2}\sum(y_i - \widehat{y_i})^2} \quad \rightarrow \quad 2\sigma_e^2\pi^{-\frac{n}{2}} e^{-\frac{1}{2\sigma_e^2}(\mathrm{y} - \mathrm{X}\hat{\mathrm{b}})'(\mathrm{y} - \mathrm{X}\hat{\mathrm{b}})}$$

The likelihood function reflects how likely you are to observe a set of data points given the PDF

Maximum Likelihood

- Find most likely solutions for unknowns given the data at hand

- The likelihood function is log-linear
  - Maximizing the log of the function is equivalent to solving the actual function

- Take the first derivative of log-likelihood
  - Set equal to zero, do a little matrix algebra, and you get the equations to solve for all effects

Log-likelihood

$$2\sigma_e^2\pi^{-\frac{n}{2}}e^{-\frac{1}{2\sigma_e^2}(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})'(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})}$$

$$-\frac{n}{2}\ln(\sigma_e^2)-\frac{n}{2}\ln(2\pi)-\frac{1}{2\sigma_e^2}(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})'(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})$$

R= σ_e^2*I

$$-\frac{1}{2}(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})'\mathrm{R}^{-1}(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})$$

MLE and OLS

OLS $\hat{b} = (X'X)^{-1}X'y$

MLE $\hat{b} = (X'R^{-1}X)^{-1}X'R^{-1}y$

Mixed model

- Has both fixed and random effects

$$y = Xb + Zu + e$$

Where $u \sim N(0, G)$ and $e \sim N(0, R)$. Given u and e are independent we can use expectations to determine the mean and variance of y

What is the advantage of modeling an effect as random?

When should an effect be treated as random versus fixed?

Mixed model

$$E[y] = E[Xb + Zu + e]$$

$$E[y] = E[Xb] + E[Zu] + E[e]$$

$$E[y] = XE[b] + ZE[u] + E[e]$$

$$E[y] = Xb + Z0 + E0 \longrightarrow E[y] = Xb$$

Mixed model

$$V(y) = V(Xb + Zu + e)$$

$$V(y) = V(Xb) + V(Zu) + V(e)$$

$$V(y) = XV(b)X' + ZV(u)Z' + V(e)$$

$$V(y) = X0X' + ZGZ' + R \longrightarrow V(y) = ZGZ' + R$$

Mixed Model Equations

y ~ N(Xb, ZGZ'+R)

Using maximum likelihood the equations for mixed models are derived as:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Mixed Model Equations

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Can also be represented as:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \alpha I^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

$$\alpha = \frac{\sigma_e^2}{\sigma_a^2}$$

When:

$$R = \sigma_e^2 I$$

$$G = \sigma_a^2 I$$

Shrinkage

We want to calculate the mean of two varieties with n_r replicates each.

Fixed Effect

$$\hat{\mathrm{b}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{y} = \begin{bmatrix} \frac{\sum y_{1\_}}{n_r} \\ \frac{\sum y_{2\_}}{n_r} \end{bmatrix}$$

Random Effect:

$$\hat{\mathrm{u}} = (\mathrm{Z}'\mathrm{Z} + \alpha * \mathrm{I})^{-1}\mathrm{Z}'\mathrm{y} = \begin{bmatrix} \frac{\sum y_{1\_}}{n_r + \sigma_e^2 / \sigma_a^2} \\ \frac{\sum y_{2\_}}{n_r + \sigma_e^2 / \sigma_a^2} \end{bmatrix}$$

Mixed Models: Variance Components

To solve for b and u, you need to estimates of the variance components, to get estimates of the variance components you need to have estimate of b and u

Solving MME

- First Derivative Methods
  - Expectation Maximization Algorithm
  - Simple to implement

- Second Derivative Methods
  - Newton-Raphson
  - Faster convergence than first derivative methods

Correlated Residuals - Repeated Measures

When the same multiple measurements are taken on the same experimental unit, residuals associated with those measurement can no longer be considered independent.

In such cases it is important to account for the correlation structures between observations.

Basic Operations –Kronecker Product

Complex covariance structures are typically specified as the **Kronecker Product** of multiple simple covariance structures

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

Kronecker Product

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}B & a_{12}B & a_{13}B \\ a_{21}B & a_{21}B & a_{23}B \\ a_{31}B & a_{32}B & a_{33}B \end{bmatrix}$$

9x9 matrix

Repeated measures – Technical Replication

The use of technical replications in experiments is common when the technology being used to take measurements is error prone and the cost of including more and there are limitations including additional experimental units.

Examples:

- Conducting an experiment in a growth chamber that has limits on the number of plants.
- Conducting a field experiment but there is limited seed for planting plots

If I were collecting gene expression data (which can be noisy) it may increase the power of my experiment to take repeated measurements

Repeated measures – Time Series

Some studies may be focused on modeling an effect and how it changes through time. In these studies repeated measurements of the experiment units are taken through time.

Examples:

- Experiments looking a disease resistance/susceptibility. In such studies it is common to take measurements at multiple time points following inoculation.

- Developmental biology. Experiments may be interested in understanding how expression of certain genes changes through time. In such cases each experimental unit may be measure at different development stages.

Repeated measures

In the case of technical replication and time series experiments it is reasonable to assume that residual error from measurements taken on the same experimental unit may be correlated.

Residual Correlation Structures

Residual correlation structures represent assumptions of the underlying factors that lead to the correlations.

The full covariance structure may be the product of multiple correlation and variance structures.

Some Common Structures

Independent and Identically Distributed

$$\mathbf{I} * \sigma^2 = \begin{bmatrix} \sigma^2 & 0 & \vdots & 0 \\ 0 & \sigma^2 & \vdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \vdots & \sigma^2 \end{bmatrix}$$

Some Common Structures

Uniform Correlation/Covariance

$$\begin{bmatrix} 1 & \rho & \vdots & \rho \\ \rho & 1 & \vdots & \rho \\ \cdots & \cdots & \ddots & \cdots \\ \rho & \rho & \vdots & 1 \end{bmatrix} * \sigma^2 = \begin{bmatrix} \sigma^2 & \sigma & \vdots & \sigma \\ \sigma & \sigma^2 & \vdots & \sigma \\ \cdots & \cdots & \ddots & \cdots \\ \sigma & \sigma & \vdots & \sigma^2 \end{bmatrix}$$

$$\sigma = \rho * \sigma^2$$

Some Common Structures

Unstructured

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \vdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \vdots & \sigma_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \vdots & \sigma_n^2 \end{bmatrix}$$

Places no restrictions on the covariance structure

Not very scalable

$$\# \, parameters = \frac{n * (n-1)}{2} + n$$

Technical Replicates – Taking Means

One approach to dealing with technical replications is to take the mean of the replicated measurements

$$\mathrm{V}(\bar{y}_i) = \frac{\sigma_e^2}{n_i}$$

$$\mathrm{V}(\bar{y}) = \begin{bmatrix} \frac{\sigma_e^2}{n_1} & 0 & \vdots & 0 \\ 0 & \frac{\sigma_e^2}{n_1} & \vdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \vdots & \frac{\sigma_e^2}{n_u} \end{bmatrix}$$

$\bar{y}$ is i.i.d when $n_1 = n_2 = \ldots = n_u$

Technical Replicates – Uniform Correlation

Simple Example – 3 repeated measurements on 5 experimental units

$$\begin{bmatrix} \sigma^2 & & & & \\ & \sigma^2 & & & \\ & & \sigma^2 & & \\ & & & \sigma^2 & \\ & & & & \sigma^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

Technical Replicates – Including random observation unit effect

Simple Example – 3 repeated measurements on 5 experimental units

$$\text{ExpUnit} \sim N\left(0, \begin{bmatrix} \sigma_u^2 & & & & \\ & \sigma_u^2 & & & \\ & & \sigma_u^2 & & \\ & & & \sigma_u^2 & \\ & & & & \sigma_u^2 \end{bmatrix}\right)$$

Residuals distributed as i.i.d. normal

$$\text{V(ExpUnit+e)} = ZZ'\sigma_u^2 + \text{I}\sigma^2$$

Technical Replicates – Including random observation unit effect

Simple Example – 3 repeated measurements on 5 experimental units

$$
\begin{bmatrix}
\sigma^2+\sigma_u^2 & \sigma_u^2 & \sigma_u^2 & & & & & & & & & & \\
\sigma_u^2 & \sigma^2+\sigma_u^2 & \sigma_u^2 & & & & & & & & & & \\
\sigma_u^2 & \sigma_u^2 & \sigma^2+\sigma_u^2 & & & & & & & & & & \\
 & & & \sigma^2+\sigma_u^2 & \sigma_u^2 & \sigma_u^2 & & & & & & & \\
 & & & \sigma_u^2 & \sigma^2+\sigma_u^2 & \sigma_u^2 & & & & & & & \\
 & & & \sigma_u^2 & \sigma_u^2 & \sigma^2+\sigma_u^2 & & & & & & & \\
 & & & & & & \sigma^2+\sigma_u^2 & \sigma_u^2 & \sigma_u^2 & & & & \\
 & & & & & & \sigma_u^2 & \sigma^2+\sigma_u^2 & \sigma_u^2 & & & & \\
 & & & & & & \sigma_u^2 & \sigma_u^2 & \sigma^2+\sigma_u^2 & & & & \\
 & & & & & & & & & \ddots & & & \\
 & & & & & & & & & & \sigma^2+\sigma_u^2 & \sigma_u^2 & \sigma_u^2 \\
 & & & & & & & & & & \sigma_u^2 & \sigma^2+\sigma_u^2 & \sigma_u^2 \\
 & & & & & & & & & & \sigma_u^2 & \sigma_u^2 & \sigma^2+\sigma_u^2
\end{bmatrix}
$$

Longitudinal Data

When taking multiple measurements through time it is reasonable to assume that residuals from measurements taken closer together in time may have higher correlations than measurements taken further apart.

In this scenario a uniform correlation structure may not be appropriate.

Longitudinal Data - Toeplitz

When working with longitudinal data the term permanent environmental effect (**pe**) is sometimes used. It represent residual effects that are consistent through time, as opposed to residual effects associate with a specific point in time.

$$\begin{bmatrix} \sigma_{pe}^2 & \sigma_{pe_1} & \sigma_{pe_2} \\ \sigma_{pe_1} & \sigma_{pe}^2 & \sigma_{pe_1} \\ \sigma_{pe_2} & \sigma_{pe_1} & \sigma_{pe}^2 \end{bmatrix} \otimes \mathbf{I}_{\boldsymbol{n}}$$

Longitudinal Data - Autoregressive

Unlike the banded structure of the Toeplitz matrix, and autoregressive function has a single parameter that decays with time/distance between points:

$$\sigma^2 * \mathbf{I}_{\boldsymbol{n}} \otimes \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Longitudinal Data

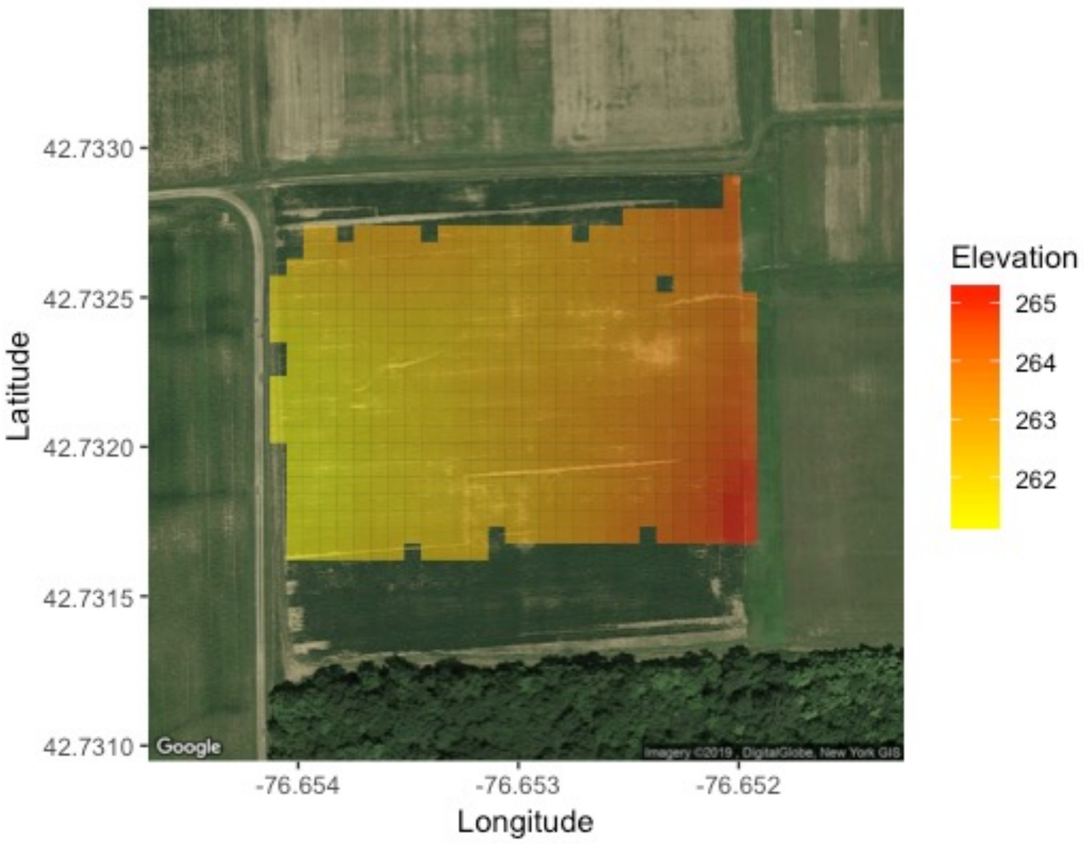
As with all covariances matrices modeling an underlying process that results in correlated effects or residuals, when in doubt there is always the unstructured covariance matrix:

$$\begin{bmatrix} \sigma_{t1}^2 & \sigma_{t12} & \sigma_{t13} \\ \sigma_{t21} & \sigma_{t2}^2 & \sigma_{t23} \\ \sigma_{31} & \sigma_{t32} & \sigma_{t3}^2 \end{bmatrix} \otimes \mathbf{I}_{\boldsymbol{n}}$$

Spatial Effects

When conducting field experiments spatial variation is common and can have a significant impact on phenotypic measures.

Factors that lead to spatial variation:
- Soil composition
- Change in elevation across the field
- Soil fertility
- Disease pressure

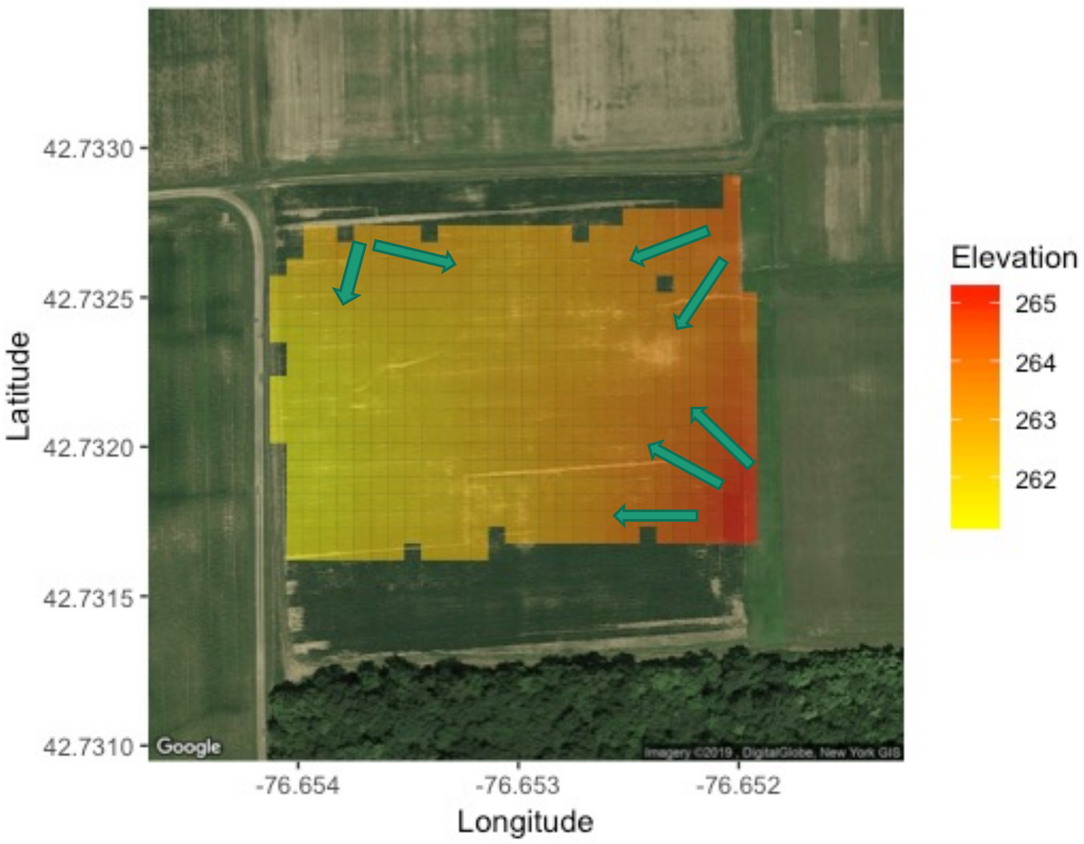


Spatial Effects

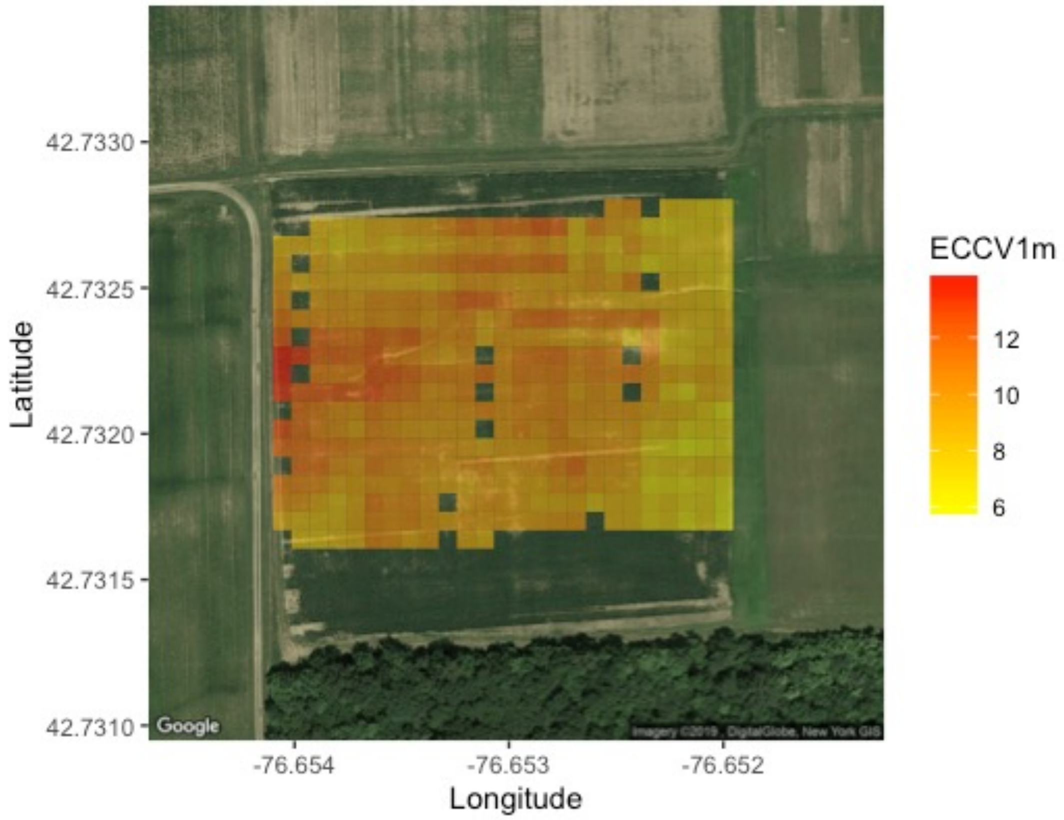
When conducting field experiments spatial variation is common and can have a significant impact on phenotypic measures.

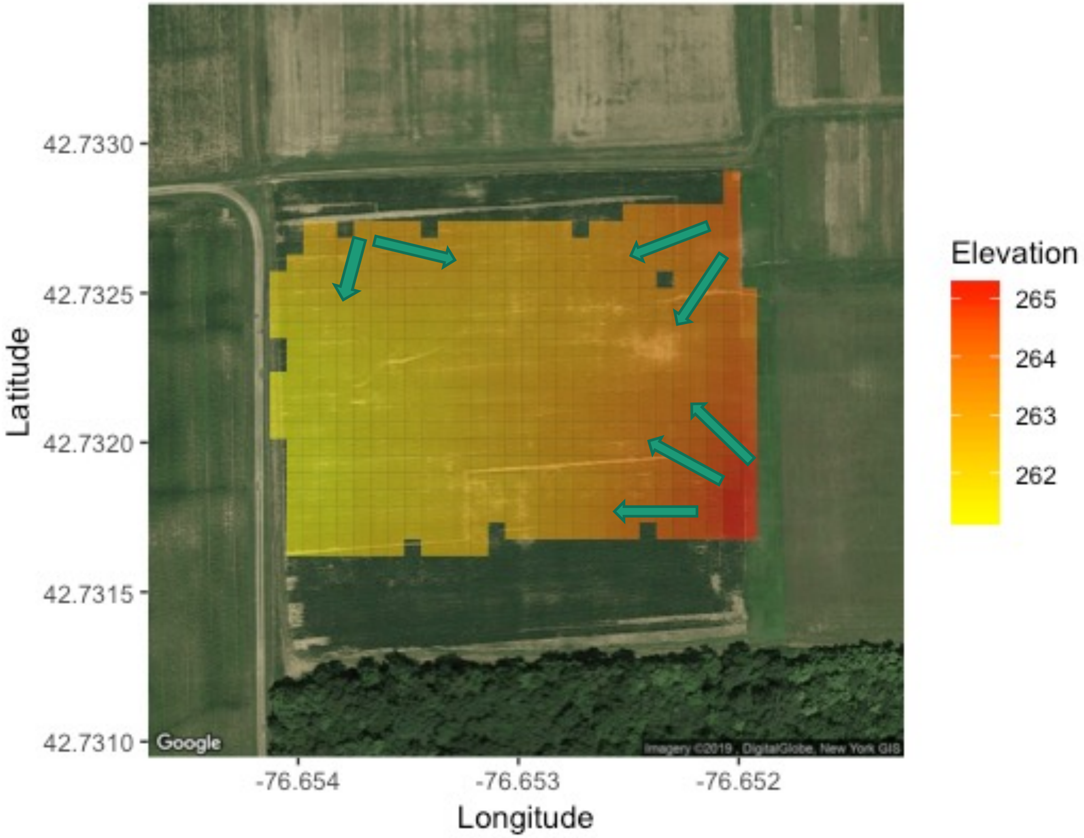
Factors that lead to spatial variation:

- Soil composition
- Change in elevation across the field
- Soil fertility
- Disease pressure



Spatial Effects





Spatial Effects – 2D Separable Autoregressive Structure

In this structure correlations are modeled in two-dimensions, both row and column

$$\sigma^2 * \begin{bmatrix} 1 & \rho_r & \rho_r^2 & \rho_r^3 & \rho_r^4 \\ \rho_r & 1 & \rho_r & \rho_r^2 & \rho_r^3 \\ \rho_r^2 & \rho_r & 1 & \rho_r & \rho_r^2 \\ \rho_r^3 & \rho_r^2 & \rho_r & 1 & \rho_r \\ \rho_r^4 & \rho_r^3 & \rho_r^2 & \rho_r & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho_c & \rho_c^2 & \rho_c^3 & \rho_c^4 \\ \rho_c & 1 & \rho_c & \rho_c^2 & \rho_c^3 \\ \rho_c^2 & \rho_c & 1 & \rho_c & \rho_c^2 \\ \rho_c^3 & \rho_c^2 & \rho_c & 1 & \rho_c \\ \rho_c^4 & \rho_c^3 & \rho_c^2 & \rho_c & 1 \end{bmatrix}$$

Spatial Effects – Moving Average

$$\sigma^2 * \begin{bmatrix} 1 & \phi_{r1} & \phi_{r2} & \phi_{r3} & 0 \\ \phi_{r1} & 1 & \phi_{r1} & \phi_{r2} & \phi_{r3} \\ \phi_{r2} & \phi_{r1} & 1 & \phi_{r1} & \phi_{r2} \\ \phi_{r3} & \phi_{r2} & \phi_{r1} & 1 & \phi_{r1} \\ 0 & \phi_{r3} & \phi_{r2} & \phi_{r1} & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \phi_{c1} & \phi_{c2} & 0 & 0 \\ \phi_{c1} & 1 & \phi_{c1} & \phi_{c2} & 0 \\ \phi_{c2} & \phi_{c1} & 1 & \phi_{c1} & \phi_{c2} \\ 0 & \phi_{c2} & \phi_{c1} & 1 & \phi_{c1} \\ 0 & 0 & \phi_{c2} & \phi_{c1} & 1 \end{bmatrix}$$

Extra Content Matrix Algebra

Matrix Notation

- Using matrix notation is less cumbersome than sums.
- Matrices provide an efficient way to organize multiple equations.
- This section will cover:
  - The basics of matrix algebra
  - Translate some of the concepts covered in the last lecture from scalar to matrix operations
  - Show you how to set up an incidence matrix for a simple linear model and solve using matrix operations in R.

Vectors and Matrices

$$b_1 + b_2 + b_3 = 7$$
$$2b_1 + b_2 + b_3 = 8$$
$$b_1 + b_2 - b_3 = 3$$

Known Coefficients	Unknown Variables	Response Variables
$1, 1, 1$	b_1	7
$2, 1, 1$	b_2	8
$1, 1, -1$	b_3	3

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 7 \\ 8 \\ 3 \end{bmatrix}$$

Basic Operations - Addition

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\mathrm{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$\mathrm{A} + \mathrm{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} \\ a_{31} + b_{31} & a_{32} + b_{32} & a_{33} + b_{33} \end{bmatrix}$$

For addition matrices must be of the same dimension

Basic Operations - Multiplication

$c = scalar\ value$

Lower case and italic is scalar

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\mathrm{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

Upper case is a matrix

$$\mathrm{a}_{1\_} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix}$$

Lower case and not italic is vector, often bold as well

$$\mathrm{b}_{\_1} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}$$

Scalar Multiplication

$$cA = \begin{bmatrix} c * a_{11} & c * a_{12} & c * a_{13} \\ c * a_{21} & c * a_{22} & c * a_{23} \\ c * a_{31} & c * a_{32} & c * a_{33} \end{bmatrix}$$

Basic Operations - Multiplication

$c = scalar\ value$

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\mathrm{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$\mathrm{a}_{1\_} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix}$$

$$\mathrm{b}_{\_1} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}$$

Vector Multiplication

Inner Product

$$\mathrm{a}_{1\_} * \mathrm{b}_{\_1} = \sum_i a_{1i} * b_{i1}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} * \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} = a_{11} * b_{11} + a_{12} * b_{21} + a_{13} * b_{31}$$

Basic Operations - Multiplication

$c = scalar\ value$

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad \mathrm{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$\mathrm{a}_{1\_} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \qquad \mathrm{b}_{\_1} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}$$

Vector Multiplication

Outer Product

$$\begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} * \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} = \begin{bmatrix} b_{11} * a_{11} & b_{11} * a_{12} & b_{11} * a_{13} \\ b_{21} * a_{11} & b_{21} * a_{12} & b_{21} * a_{13} \\ b_{31} * a_{11} & b_{31} * a_{12} & b_{31} * a_{13} \end{bmatrix}$$

Basic Operations - Multiplication

$c = scalar\ value$

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\mathrm{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$\mathrm{a}_{1\_} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix}$$

$$\mathrm{b}_{\_1} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}$$

Matrix Multiplication

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} * \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11} * b_{11} + a_{12} * b_{21} + a_{13} * b_{31} & a_{11} * b_{12} + a_{12} * b_{22} + a_{13} * b_{32} & a_{11} * b_{13} + a_{12} * b_{23} + a_{13} * b_{33} \\ a_{21} * b_{11} + a_{22} * b_{21} + a_{23} * b_{31} & a_{21} * b_{12} + a_{22} * b_{22} + a_{23} * b_{32} & a_{21} * b_{13} + a_{22} * b_{23} + a_{23} * b_{33} \\ a_{31} * b_{11} + a_{32} * b_{21} + a_{33} * b_{31} & a_{31} * b_{12} + a_{32} * b_{22} + a_{33} * b_{32} & a_{31} * b_{13} + a_{32} * b_{33} + a_{33} * b_{33} \end{bmatrix}$$

Basic Operations - Multiplication

$c = scalar\ value$

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\mathrm{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$\mathrm{a}_{1\_} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix}$$

$$\mathrm{b}_{\_1} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}$$

Kronecker Product

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}\mathrm{B} & a_{12}\mathrm{B} & a_{13}\mathrm{B} \\ a_{21}\mathrm{B} & a_{21}\mathrm{B} & a_{23}\mathrm{B} \\ a_{31}\mathrm{B} & a_{32}\mathrm{B} & a_{33}\mathrm{B} \end{bmatrix}$$

Multiplication Dimensions

$c \quad a_3 \quad b_3 \quad A_{3x3} \quad B_{3x3} \quad C_{3x4}$

$$c * A = D_{3x3}$$

Inner Product

$$a * b = d$$

Outer Product

$$a * b = D_{3x3}$$

$$A * B = D_{3x3}$$

$$A \otimes B = D_{9x9}$$

$$A * C = D_{3x4}$$

$$C * A = ??$$

$$A \otimes C = D_{9x12}$$

Transpose

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad \mathrm{A}' or \mathrm{A}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

Quadratic form

$$(\mathrm{AB})' = \mathrm{B}'\mathrm{A}' \qquad (\mathrm{ABC})' = \mathrm{C}'\mathrm{B}'\mathrm{A}' \qquad \mathrm{a}'\mathrm{Bc} = \mathrm{c}'\mathrm{B}'\mathrm{a}$$

Trace

Given A is an *n* by *n* matrix

$$Tr(A) = \sum_{i=1}^{n} a_{ii}$$

Inverse

Scalar

$$ax = y$$

$$\left(\frac{1}{a}\right) ax = \left(\frac{1}{a}\right) * y$$

$$a^{-1} = \frac{1}{a}$$

Matrix

$$\mathrm{AA^{-1}} = \mathrm{I}$$

$$\mathrm{A}\mathbf{x} = \mathbf{y}$$

$$\mathrm{A^{-1}A}\mathbf{x} = \mathrm{A^{-1}}\mathbf{y}$$

$$\mathrm{I}\mathbf{x} = \mathrm{A^{-1}}\mathbf{y}$$

$$\mathrm{I} = \left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\right]$$

$$\mathrm{A} * \mathrm{I} = \mathrm{A} \qquad \mathrm{x} * \mathrm{I} = \mathrm{x}$$

Calculating the Inverse Simple Examples

$$A = \left[\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{bmatrix}\right]$$

$$A^{-1} = \left[\begin{bmatrix} 1/a & 0 & 0 & 0 \\ 0 & 1/b & 0 & 0 \\ 0 & 0 & 1/c & 0 \\ 0 & 0 & 0 & 1/d \end{bmatrix}\right]$$

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$BB^{-1} = \begin{bmatrix} \frac{ad - bc}{ad - bc} & \frac{-ab + ab}{ad - bc} \\ \frac{cd - dc}{ad - bc} & \frac{-cb + ad}{ad - bc} \end{bmatrix}$$

$$B^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Determinants

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad |B| = ad - bc \qquad B^{-1} = \frac{1}{|B|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Laplace's Formula

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

Minor

$$|A| = (-1)^{1+1} * a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + (-1)^{1+2} * a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + (-1)^{1+3} * a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{32} & a_{32} \end{vmatrix}$$

Can be applied recursively for larger matrices

Calculating the Inverse of a Larger Matrix

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

Step 1 : replace each element with the determinant of it's minor

$$\begin{bmatrix} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{bmatrix}$$

Step 2 : Multiply each determinant of the minor by $(-1)^{i+j}$

Step 3: Transpose and multiply by $\frac{1}{|A|}$

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{bmatrix}^T$$

Properties of the Inverse and Determinant

$(\mathrm{A}^{-1})^{-1} = \mathrm{A}$

$(b\mathrm{A})^{-1} = b^{-1}\mathrm{A}^{-1}$

$(\mathrm{BA})^{-1} = \mathrm{B}^{-1}\mathrm{A}^{-1}$

$(\mathrm{A}')^{-1} = (\mathrm{A}^{-1})'$

$|\mathrm{A}^{-1}| = (|\mathrm{A}|)^{-1}$

$|\mathrm{A}'| = |\mathrm{A}|$

$|\mathrm{BA}| = |\mathrm{B}|\,|\mathrm{A}|$

For *n* by *n* matrix

$|b\mathrm{A}| = b^{n}\,|\mathrm{A}|$

Determinant of a diagonal matrix is the product of the diagonal elements

Extra Content: Solving Mixed Models

EM Algorithm

E step: solve using current estimate of R and G

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \alpha A^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \alpha A^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

M step: estimate new R and G using current solutions from MME

$$\hat{\sigma}_e^2 = {y'y - \widehat{b}'X'y - \widehat{u}'Z'Y} \Big/ {(N - rank(X))}$$

$$\hat{\sigma}_a^2 = {(\hat{u}'A^{-1}\hat{u} + trace(A^{-1}C_{22})\hat{\sigma}_e^2)} \Big/ {k}$$

EM Implementation

- Step 1: Solve for $\widehat{\mathrm{b}}$ and $\hat{\mathrm{u}}$ given the current solutions of R and G
- Step 2:
  - Solve $\hat{\sigma}_e^2 = \mathrm{y}'\mathrm{y} - \widehat{\mathrm{b}}'\mathrm{X}'\mathrm{y} - \widehat{\mathrm{u}}'\mathrm{Z}'\mathrm{Y} \big/ (N - rank(\mathrm{X}))$
  - Solve $\hat{\sigma}_a^2 = (\hat{\mathrm{u}}'\mathrm{A}^{-1}\hat{\mathrm{u}} + trace(\mathrm{A}^{-1}\mathrm{C}_{22})\hat{\sigma}_e^2) \big/ k$
- Repeat steps 1 and 2 until solutions converge.

Newton-Raphson

Based on Taylor's series expansion

$$f(\theta_t) = f(\theta_{t-1}) + \frac{f'(\theta_{t-1})}{1!}(\theta_t - \theta_{t-1}) + \frac{f''(\theta_{t-1})}{2!}(\theta_t - \theta_{t-1})^2 + \cdots + \frac{f^{n\prime}(\theta_{t-1})}{n!}(\theta_t - \theta_{t-1})^n$$

Newton-Raphson

Based on Taylor's series expansion

$$f(\theta_t) = f(\theta_{t-1}) + \frac{f'(\theta_{t-1})}{1!}(\theta_t - \theta_{t-1}) + \frac{f''(\theta_{t-1})}{2!}(\theta_t - \theta_{t-1})^2 + \cdots + \frac{f^{n\prime}(\theta_{t-1})}{n!}(\theta_t - \theta_{t-1})^n$$

This approach is useful for approximating values that are difficult or impossible to solve for directly.

A common example is approximating the value of the mathematical constant *e* using a Taylor series example.

Newton-Raphson

Based on Taylor's series expansion

$$f(\theta_t) = f(\theta_{t-1}) + \frac{f'(\theta_{t-1})}{1!}(\theta_t - \theta_{t-1}) + \frac{f''(\theta_{t-1})}{2!}(\theta_t - \theta_{t-1})^2 + \cdots + \frac{f^{n\prime}(\theta_{t-1})}{n!}(\theta_t - \theta_{t-1})^n$$

This approach is useful when $f(\theta)$ is undefined at θ_t . We can choose some value θ_{t-1} for which $f(\theta)$ is defined and use it to approximate $f(\theta_t)$.

In the case of mixed models our solutions $\hat{\mathrm{b}}$ and $\hat{\mathrm{u}}$ are undefined because we don't know the variance components

Newton-Raphson

We want solutions that maximize the likelihood function, or more specifically the log-likelihood function.

$\Delta\theta$

$$0 = f(\theta_{t-1}) + \frac{f'(\theta_{t-1})}{1!}(\theta_t - \theta_{t-1})$$

First derivative of the log-likelihood function

Second derivative of the log-likelihood function

$$\Delta\theta = -f(\theta_{t-1})f'(\theta_{t-1})^{-1}$$

$$\theta_t - \theta_{t-1} = -f(\theta_{t-1})f'(\theta_{t-1})^{-1}$$

$$\theta_t = \theta_{t-1} - f(\theta_{t-1})f'(\theta_{t-1})^{-1}$$

$$\theta_t = \theta_{t-1} + \Delta\theta$$

Newton-Raphson

In our case θ is a vector of unknowns so:

$$f(\theta) = \begin{bmatrix} \frac{\partial L}{\partial \theta_1} & \frac{\partial L}{\partial \theta_1 \partial \theta_2} & \frac{\partial L}{\partial \theta_1 \partial \theta_n} \\ & \frac{\partial L}{\partial \theta_2} & \frac{\partial L}{\partial \theta_2 \partial \theta_n} \\ & & \frac{\partial L}{\partial \theta_n} \end{bmatrix}$$

$$f'(\theta) = \begin{bmatrix} \frac{\partial^2 L}{\partial^2 \theta_1} & \frac{\partial^2 L}{\partial^2 \theta_1 \partial^2 \theta_2} & \frac{\partial^2 L}{\partial^2 \theta_1 \partial^2 \theta_n} \\ & \frac{\partial^2 L}{\partial^2 \theta_2} & \frac{\partial^2 L}{\partial^2 \theta_2 \partial^2 \theta_n} \\ & & \frac{\partial^2 L}{\partial^2 \theta_n} \end{bmatrix}$$

Hessian Matrix

Quasi-Newton Methods: Use an approximation of the Hessian

Fischer's Scoring Method: Replace the Hessian with Fischer Information Matrix/Expected Information Matrix

Newton-Raphson

Step 1:

Set initial value as θ_{t-1}

Step2:

Calculate $\Delta\theta = -f(\theta_{t-1})f'(\theta_{t-1})^{-1}$

Step 3:

Calculate $\theta_t = \theta_{t-1} + \Delta\theta$

Step 4:

Set new θ_{t-1} equal to θ_t

Repeat until convergence

Taking a Step Back – Big Picture

- The expected value, E [f(x)], of some function f of the random variable x is just the mean value of that function as n approaches infinity

$$E[x] = \mu$$

Expectations of a random variable

For a discrete random variable, the expected value of x (x being the observed value of a random variable is:

$$\sum_{1}^{n} x * p(x)$$

For all possible values of x

Expectations of a random variable

For a continuous random variable x:

$$\int_{-\infty}^{\infty} x * p(x)dx$$

Calculating the expectation of $(x-\mu)^2$ gives the variance of the distribution.

$$E[(x-\mu)^2] = \sigma^2$$

Moments of a statistical distribution

$E[(x-\mu)^r]$ is the rth moment if a distribution.

r = 3 = Skewness – Is a measure of the symmetry of a distribution around μ. It is 0 for the normal distribution.

r = 4 = Kurtosis – Is a measure of the tails of a distribution, or how likely it is to observe “extreme” values. For a normal distribution it is $3\sigma^4$

In most cases we are only interested in the mean (r=1) and variance (r=2)

Properties of Expectations

Theorem 1: Let c be a constant, and $g_1(x), g_2(x), \ldots, g_k(x)$ be functions of a continuous random variable x.

1) $E[c] = c$

2) $E[cg(x)] = cE[g(x)]$

3) $E[g_1(x) + g_2(x) + \cdots + g_k(x)] = E[g_1(x)] + E[g_2(x)] + \cdots + E[g_k(x)]$

Covariance

$Cov(x,y) > 0$ then there is a positive linear relationship between x and y

$Cov(x,y) < 0$ then there is a negative linear relationship between x and y

$Cov(x,y) = 0$ does not necessarily mean there is no relationship between x and y, but if there is no relationship between x and y $Cov(x,y)$ will always equal 0.

Variance and Covariance

Using expectations it is easy to show:

$$Cov(x, x) = V(x)$$

And using Theorem 1 we can show that:

$$Cov(ax, y) = aCov(x, y)$$

$$V(ax) = a^2V(x)$$

OLS

To minimize we take the derivative [1] with respect to $\hat{b}$, set it equal to zero, and solve

$$(y - X\hat{b})'(y - X\hat{b}) \qquad [1]$$

$$X'X\hat{b} = X'y$$

$$\hat{b} = (X'X)^{-1}X'y$$