MAXIMUM LIKELIHOOD ESTIMATION IN THE GENERAL LINEAR MODEL

Suppose $f(\underset{\sim}{w} \mid \underset{\sim}{\theta})$ is the probability density function (pdf) or probability mass function (pmf) of a random vector $\underset{\sim}{W}$, where $\underset{\sim}{\theta}$ is a $k \times 1$ parameter.

Given a value of the parameter vector $\underset{\sim}{\theta}$, $f(\underset{\sim}{w} \mid \underset{\sim}{\theta})$ is a real-valued function of $\underset{\sim}{w}$.

The likelihood function $L(\underset{\sim}{\theta} \mid \underset{\sim}{w}) = f(\underset{\sim}{w} \mid \underset{\sim}{\theta})$ is a real-valued function of $\underset{\sim}{\theta}$ for a given value of $\underset{\sim}{w}$.

For any potential observed value $\underset{\sim}{w}$, define $\hat{\underset{\sim}{\theta}}(\underset{\sim}{w})$ to be a parameter value at which $L(\underset{\sim}{\theta} \mid \underset{\sim}{w})$ attains its maximum value. A <u>maximum</u> <u>likelihood</u> <u>estimator</u> (MLE) of $\underset{\sim}{\theta}$ is $\hat{\underset{\sim}{\theta}}(\underset{\sim}{W}) = \hat{\underset{\sim}{\theta}}$.

Invariance Property of MLEs:

The MLE of a function of $\underset{\sim}{\theta}$, say $\underset{\sim}{g}(\underset{\sim}{\theta})$, is the function evaluated at the MLE of $\underset{\sim}{\theta}$:

$$\widehat{\underset{\sim}{g}(\underset{\sim}{\theta})} = \underset{\sim}{g}(\hat{\underset{\sim}{\theta}})$$

Let $\ell(\underset{\sim}{\theta} \mid \underset{\sim}{w}) = \log L(\underset{\sim}{\theta} \mid \underset{\sim}{w})$.

If $\ell(\underset{\sim}{\theta} \mid \underset{\sim}{w})$ is differentiable, candidates for the MLE of $\underset{\sim}{\theta}$ can be found by equating the <u>score function</u>

$$\frac{\partial \ell(\underset{\sim}{\theta} \mid \underset{\sim}{w})}{\partial \underset{\sim}{\theta}} \equiv \begin{bmatrix} \dfrac{\partial \ell(\underset{\sim}{\theta} \mid \underset{\sim}{w})}{\partial \theta_1} \\ \vdots \\ \dfrac{\partial \ell(\underset{\sim}{\theta} \mid \underset{\sim}{w})}{\partial \underset{\sim}{\theta}_k} \end{bmatrix}$$

to $\underset{\sim}{0}$ and solving for $\underset{\sim}{\theta}$.

The likelihood equations are

$$\frac{\partial \ell(\underset{\sim}{\theta} | \underset{\sim}{w})}{\partial \underset{\sim}{\theta}} = \underset{\sim}{0} \iff \frac{\partial \ell(\underset{\sim}{\theta} | \underset{\sim}{w})}{\partial \theta_j} = 0 \quad \forall j = 1, \ldots, K.$$

One strategy for obtaining an MLE is to find values of $\underset{\sim}{\theta}$ that satisfy the likellhood equations and verify that at least one such value maximizes $\ell(\underset{\sim}{\theta} | \underset{\sim}{w})$.

Example: Normal Theory Gauss-Markov Linear Model

$$\underset{n\times 1}{\underset{\sim}{y}} = \underset{n\times p}{X}\underset{p\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{\varepsilon}}, \quad \underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I). \quad \underset{(p+1)\times 1}{\underset{\sim}{\theta}} = \begin{bmatrix} \underset{\sim}{\beta} \\ \sigma^2 \end{bmatrix}$$

$$f(\underset{\sim}{y}|\underset{\sim}{\theta}) = \frac{\exp\left\{-\frac{1}{2}(\underset{\sim}{y}-X\underset{\sim}{\beta})'(\sigma^2 I)^{-1}(\underset{\sim}{y}-X\underset{\sim}{\beta})\right\}}{(2\pi)^{n/2}\,|\sigma^2 I|^{1/2}}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{y}-X\underset{\sim}{\beta})'(\underset{\sim}{y}-X\underset{\sim}{\beta})\right\}$$

$$\ell(\underset{\sim}{\theta}|\underset{\sim}{y}) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}(\underset{\sim}{y}-X\underset{\sim}{\beta})'(\underset{\sim}{y}-X\underset{\sim}{\beta})$$

The score function is

$$\frac{\partial \ell(\underline{\theta}|\underline{y})}{\partial \underline{\theta}} = \begin{bmatrix} \dfrac{\partial \ell(\underline{\theta}|\underline{y})}{\partial \underline{\beta}} \\[2ex] \dfrac{\partial \ell(\underline{\theta}|\underline{y})}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma^2}\left(X'\underline{y} - X'X\underline{\beta}\right) \\[2ex] \dfrac{(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})}{2\sigma^4} - \dfrac{n}{2\sigma^2} \end{bmatrix}$$

The likelihood equations are

$$\frac{\partial \ell(\underline{\theta}|\underline{y})}{\partial \underline{\theta}} = \underline{0} \iff \begin{array}{l} X'X\underline{\beta} = X'\underline{y} \\ \sigma^2 = (\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})/n \end{array}$$

Thus, a solution to the likelihood equations

is $\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ (\boldsymbol{y} - X\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})/n \end{bmatrix}$, where $\hat{\boldsymbol{\beta}}$ is any

solution to the normal equations.

We already know that $(\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta})$ is minimized over $\boldsymbol{\beta} \in \mathbb{R}^p$ by any solution to the normal equations.

Thus, $\forall\, \sigma^2 > 0$, $\ell\left(\begin{bmatrix}\hat{\boldsymbol{\beta}} \\ \sigma^2\end{bmatrix} \middle| \boldsymbol{y}\right) \geq \ell\left(\begin{bmatrix}\boldsymbol{\beta} \\ \sigma^2\end{bmatrix} \middle| \boldsymbol{y}\right)$ $\forall\, \boldsymbol{\beta} \in \mathbb{R}^p$.

Examining the 2^{ND} derivative of $\ell(\boldsymbol{\theta} | \boldsymbol{y})$ with respect to σ^2 shows that

$$\begin{bmatrix}\hat{\boldsymbol{\beta}} \\ (\boldsymbol{y} - X\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})/n\end{bmatrix} \text{ is an MLE of } \boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta} \\ \sigma^2\end{bmatrix}.$$

Thus, if $C\boldsymbol{\beta}$ is estimable, the MLE of $C\boldsymbol{\beta}$ is $C\hat{\boldsymbol{\beta}}$ (by the Invariance Property of MLEs), which is the BLUE of $C\boldsymbol{\beta}$.

Note that the MLE of σ^2 is not the unbiased estimator we have been using.

$$E\left[\frac{(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}})}{n}\right] = E\left(\frac{SSE}{n}\right)$$

$$= E\left[\frac{(n-r)MSE}{n}\right] = \frac{n-r}{n}\sigma^2 < \sigma^2.$$

Thus, the MLE underestimates σ^2 on average.

Now consider the general linear model

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}, \quad \underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \Sigma), \text{ where}$$

Σ is a positive definite covariance matrix whose entries depend on unknown parameters in some vector $\underset{\sim}{\gamma}$.

For example, $\Sigma = \begin{bmatrix} \sigma_e^2+\sigma_u^2 & \sigma_u^2 & 0 \\ \sigma_u^2 & \sigma_e^2+\sigma_u^2 & 0 \\ 0 & 0 & \sigma_e^2+\sigma_u^2 \end{bmatrix}$, $\underset{\sim}{\gamma} = \begin{bmatrix} \sigma_e^2 \\ \sigma_u^2 \end{bmatrix}$

Our linear mixed effects model

is a special case where

$\underset{\sim}{\varepsilon} = Z\underset{\sim}{u} + \underset{\sim}{e}$ and $\Sigma = ZGZ' + R$.

In general, $\underset{\sim}{\theta} = \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{bmatrix}$ and

$$f(\underset{\sim}{y} \mid \underset{\sim}{\theta}) = \frac{\exp\left\{ -\frac{1}{2} (\underset{\sim}{y} - X\underset{\sim}{\beta})' \Sigma^{-1} (\underset{\sim}{y} - X\underset{\sim}{\beta}) \right\}}{(2\pi)^{n/2} \, |\Sigma|^{1/2}}$$

$$\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}) = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (\underset{\sim}{y} - X\underset{\sim}{\beta})' \Sigma^{-1} (\underset{\sim}{y} - X\underset{\sim}{\beta}) - \frac{n}{2} \log(2\pi)$$

We know that for any positive definite covariance matrix Σ, $(\underset{\sim}{y} - X\underset{\sim}{\beta})'\Sigma^{-1}(\underset{\sim}{y} - X\underset{\sim}{\beta})$ is minimized over $\underset{\sim}{\beta} \in \mathbb{R}^p$ by

$$\hat{\underset{\sim}{\beta}}_\Sigma = (X'\Sigma^{-1}X)^- X'\Sigma^{-1}\underset{\sim}{y}.$$

Thus, for any $\underset{\sim}{\gamma}$ such that Σ is a positive definite covariance matrix,

$$\ell\left(\begin{bmatrix} \hat{\underset{\sim}{\beta}}_\Sigma \\ \underset{\sim}{\gamma} \end{bmatrix} \middle| \underset{\sim}{y} \right) \geq \ell\left(\begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{bmatrix} \middle| \underset{\sim}{y} \right) \quad \forall\, \underset{\sim}{\beta} \in \mathbb{R}^p.$$

We define the <u>profile log likelihood</u> for $\underset{\sim}{\gamma}$

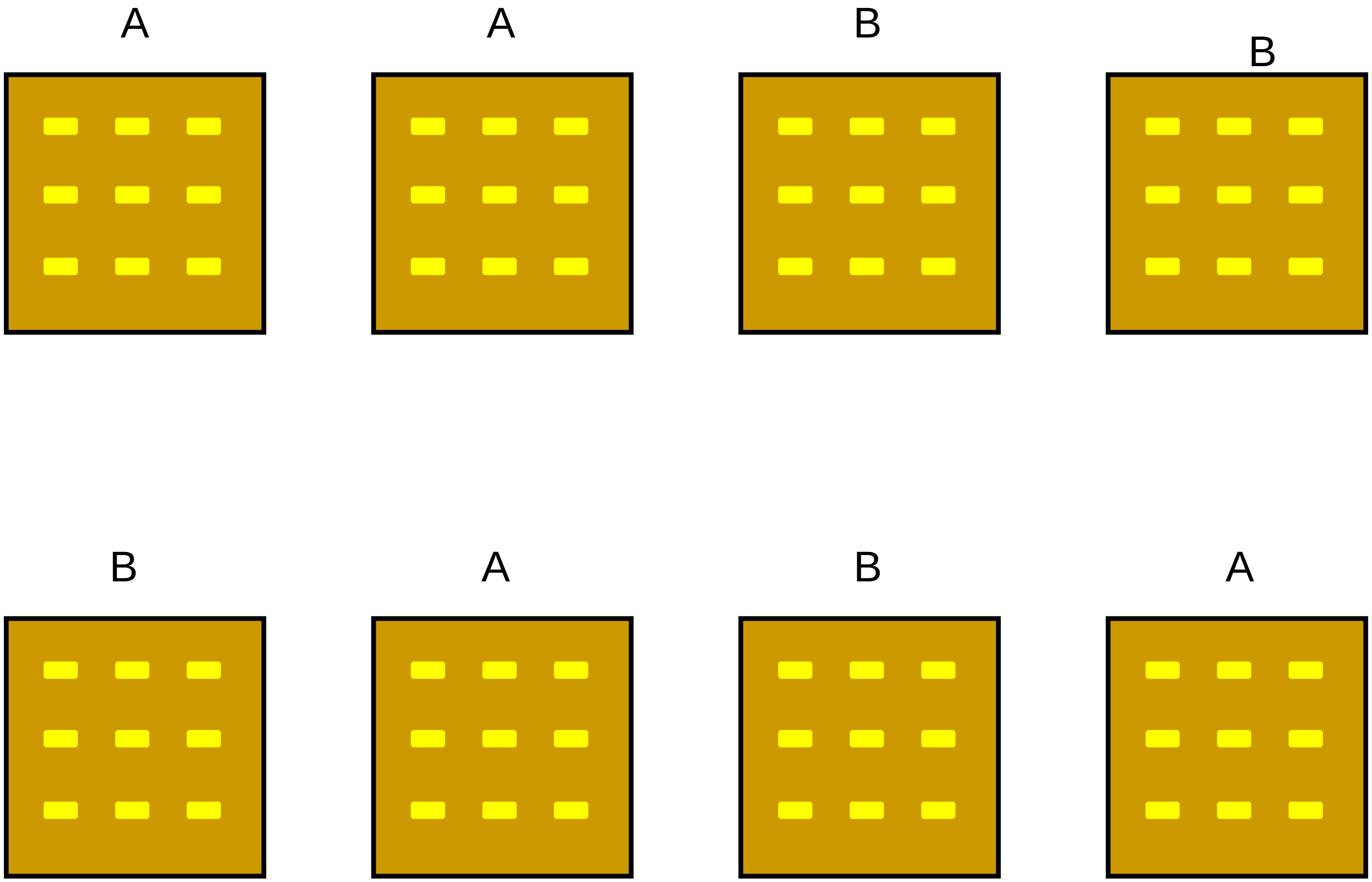
to be $\ell^*(\underset{\sim}{\gamma} \mid \underset{\sim}{y}) = \ell\left(\begin{bmatrix} \hat{\underset{\sim}{\beta}}_{\Sigma} \\ \underset{\sim}{\gamma} \end{bmatrix} \middle| \underset{\sim}{y} \right)$.

The MLE of $\underset{\sim}{\theta}$ is $\hat{\underset{\sim}{\theta}} = \begin{bmatrix} \hat{\underset{\sim}{\beta}}_{\hat{\Sigma}} \\ \hat{\underset{\sim}{\gamma}} \end{bmatrix}$,

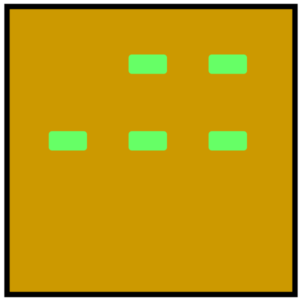
where $\hat{\underset{\sim}{\gamma}}$ is a maximizer of $\ell^*(\underset{\sim}{\gamma} \mid \underset{\sim}{y})$

and $\hat{\Sigma}$ is Σ with $\hat{\underset{\sim}{\gamma}}$ in place of $\underset{\sim}{\gamma}$.

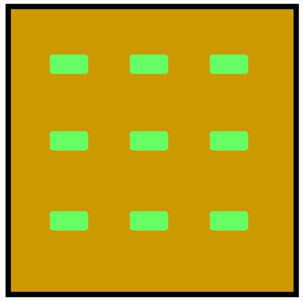
In general, numerical methods are required to find $\hat{\underline{\gamma}}$, a maximizer of $\ell^*(\underline{\gamma} \mid \underline{y})$.


A
A
B
B
B
A
B
A

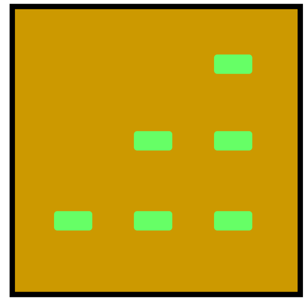
A



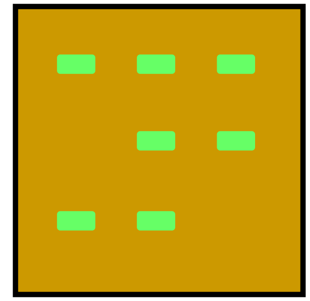
A



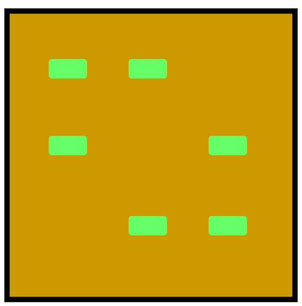
B



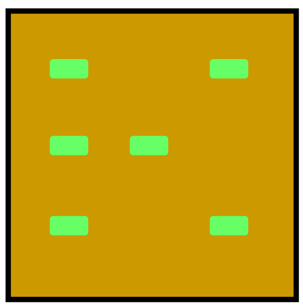
B



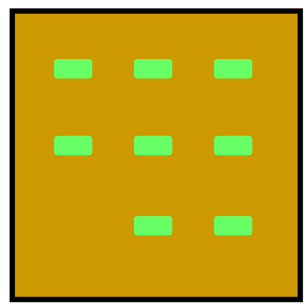
B



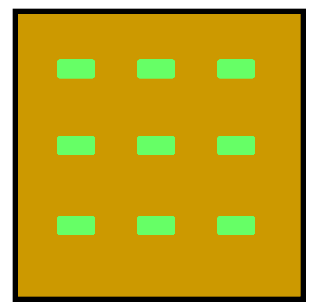
A



B



A



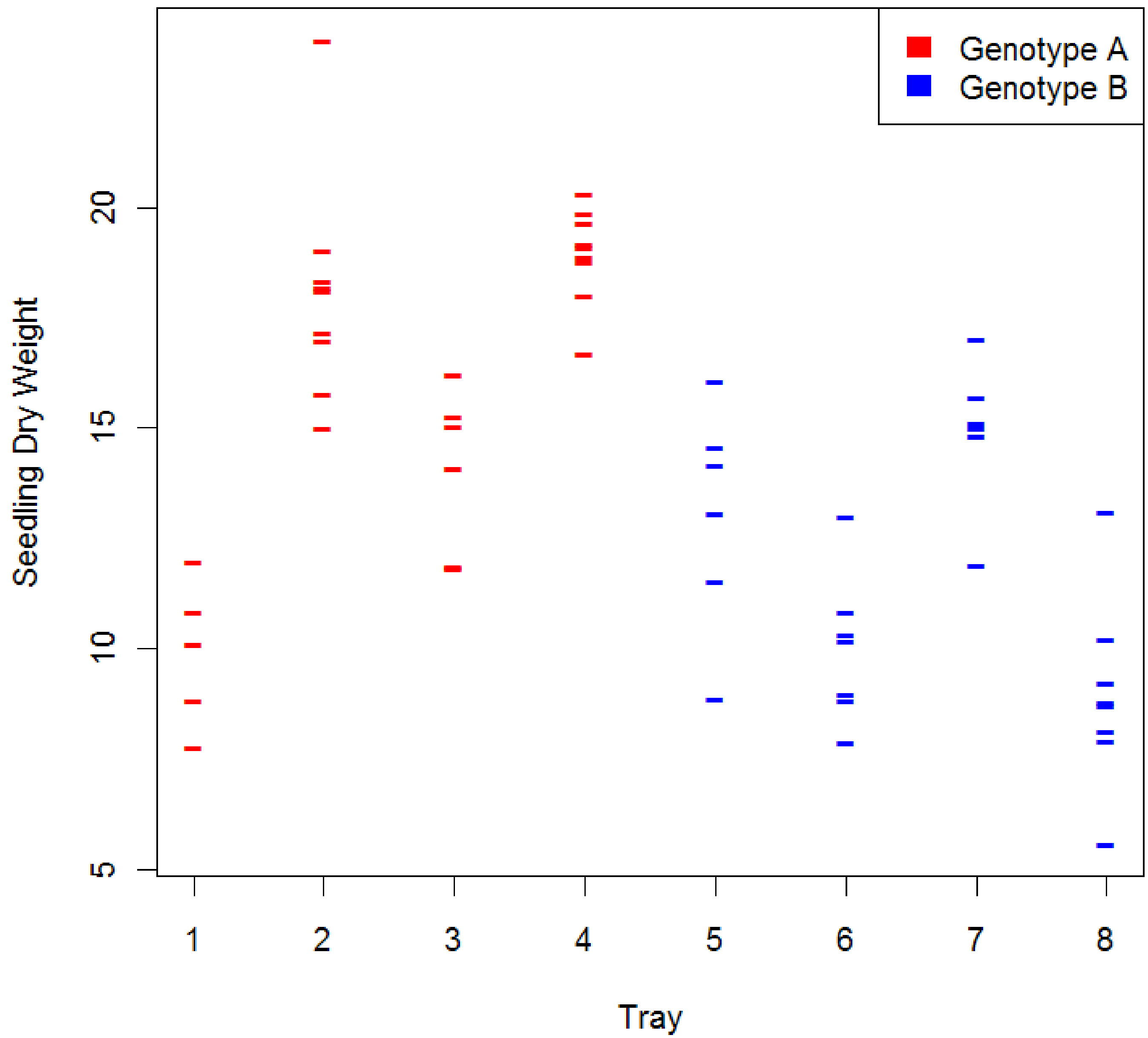
```
d=read.delim(
"http://www.public.iastate.edu/~dnett/S511/SeedlingDryWeight2.txt"
)
d
   Genotype Tray Seedling SeedlingWeight
1         A    1        1              8
2         A    1        2              9
3         A    1        3             11
4         A    1        4             12
5         A    1        5             10
6         A    2        1             17
7         A    2        2             17
8         A    2        3             16
9         A    2        4             15
10        A    2        5             19
11        A    2        6             18
12        A    2        7             18
13        A    2        8             18
14        A    2        9             24
15        A    3        1             12
16        A    3        2             12
17        A    3        3             16
```

18	A	3	4	15
19	A	3	5	15
20	A	3	6	14
21	A	4	1	17
22	A	4	2	20
23	A	4	3	20
24	A	4	4	19
25	A	4	5	19
26	A	4	6	18
27	A	4	7	20
28	A	4	8	19
29	A	4	9	19
30	B	5	1	9
31	B	5	2	12
32	B	5	3	13
33	B	5	4	16
34	B	5	5	14
35	B	5	6	14
36	B	6	1	10
37	B	6	2	10
38	B	6	3	9

39	B	6	4	8
40	B	6	5	13
41	B	6	6	9
42	B	6	7	11
43	B	7	1	12
44	B	7	2	16
45	B	7	3	17
46	B	7	4	15
47	B	7	5	15
48	B	7	6	15
49	B	8	1	9
50	B	8	2	6
51	B	8	3	8
52	B	8	4	8
53	B	8	5	13
54	B	8	6	9
55	B	8	7	9
56	B	8	8	10

```
plot(d[,2],d[,4]+rnorm(56,0,.2),
    xlab="Tray",ylab="Seedling Dry Weight",
    col=2*(1+(d[,1]=="B")),pch="-",cex=2)


legend("topright",c("Genotype A","Genotype B"),
       fill=c(2,4),border=c(2,4))
```



Genotype A
Genotype B
Seedling Dry Weight
5
10
15
20
1
2
3
4
5
6
7
8
Tray

```
library(nlme)

lme(SeedlingWeight~Genotype,random=~1|Tray,
    method="ML",data=d)

Linear mixed-effects model fit by maximum likelihood
  Data: d
  Log-likelihood: -126.3709
  Fixed: SeedlingWeight ~ Genotype
(Intercept)   GenotypeB
  15.301832   -3.567017

Random effects:
 Formula: ~1 | Tray
        (Intercept) Residual
StdDev:    2.932294 1.882470

Number of Observations: 56
Number of Groups: 8
```

```
library(lme4)

lmer(SeedlingWeight~Genotype+(1|Tray),REML=F,data=d)

Linear mixed model fit by maximum likelihood

Formula: SeedlingWeight ~ Genotype + (1 | Tray)
   Data: d
   AIC   BIC logLik deviance REMLdev
 260.7 268.8 -126.4    252.7   247.4

Random effects:
 Groups   Name        Variance Std.Dev.
 Tray     (Intercept) 8.5984   2.9323
 Residual             3.5437   1.8825

Number of obs: 56, groups: Tray, 8
```

```
Fixed effects:
            Estimate Std. Error t value
(Intercept)   15.302      1.510   10.14
GenotypeB     -3.567      2.136   -1.67

Correlation of Fixed Effects:
          (Intr)
GenotypeB -0.707
```