MA50259: Statistical Design of Investigations

Dr. Sandipan Roy

Lecture 2

Our own randomised experiment, simulating the response

In order to understand statistical analysis later, we can think of how to simulate the response in our own randomised experiment.

- Let Y_{ij} be the response for the jth experimental unit subject to the ith level of the treatment factor, $i = 1, \ldots, t$ and $j = 1, \ldots, r_i$. Y_{ij} is treated as a **random variable**
- r_i is the number of replicates in ith level of the treatment factor. Note that $r_i = r$ for all i, if the design is balanced
- Let μ_i be the **mean response** [long-run average of all possible experiments] at the ith level of the treatment factor, then

$$Y_{ij} = \mu_i + \epsilon_{ij}\,, \quad \text{where} \quad E[Y_{ij}] = \mu_i$$

- ϵ_{ij} is the [additive] experimental error and is also treated as a **random variable** such that $E[\epsilon_{ij}] = 0$

Our own randomised experiment, simulating the response

- Let μ be either the overall response average when there is no treatment or simply a reference value of interest.
- **Linear model**: Let $\mu_i = \mu + \tau_i$ so that

$$\begin{aligned} Y_{ij} &= \mu_i + \epsilon_{ij} \\ &= \mu + \tau_i + \epsilon_{ij} \end{aligned}$$

- The τ_i's are called the **treatment effects**. τ_i represents the difference between the mean response at the i-th level of the treatment factor and the reference value μ
- **Assumption:** There is an additive effect of the treatment levels on the means, e.g. $\mu_i = \mu + \tau_i$
- Note the treatment effects model has $t + 1$ unknown parameters and the means model only has t!

Our own CRD experiment, simulating the response

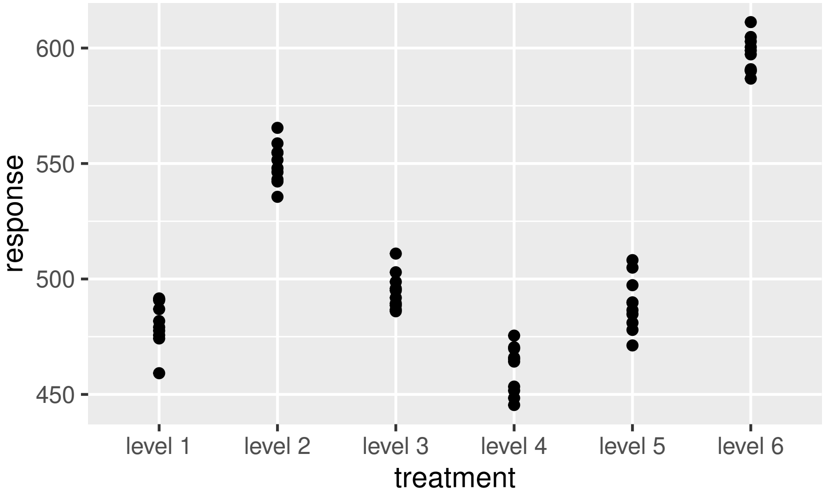
- Specifying the probability distribution of the experimental errors ϵ_{ij} determines uniquely the probability distribution of the responses Y_{ij} [and viceversa]
- Two sources of randomness: + Randomisation of units to treatement levels + Randomness in the response for different experimental units
- Due to the **randomisation** of units to treatement levels, the responses Y_{ij} (or the experimental errors ϵ_{ij}), are **mutually independent**
- **Assumption:** If the units are **homogeneous**, the probability distribution of the responses Y_{ij} (or the experimental errors ϵ_{ij}), under the same treatment level, is the **same**.
- **Assumption:** the probability distribution of the responses Y_{ij} (or the experimental errors ϵ_{ij}), is **Gaussian**.

Our own CRD experiment, simulating the response

```
mu<-500 # overall response
tau<-c(-20,50,0,-30,-10,100) # treatment effects
sd<-10 # overall standard deviation
means<-mu+tau %>% rep(each=r) # vector of means
y<-rnorm(n,mean=means,sd=sd) # units are arranged by treatment level
crd$response<-y
glimpse(crd)
Rows: 60
Columns: 3
$ units     <int> 7, 12, 14, 19, 23, 31, 33, 39, 54, 55, 18, 28, 37, 38, 41, 4~
$ treatment <fct> level 1, level 1, level 1, level 1, level 1, level 1, level ~
$ response  <dbl> 475.6323, 486.9446, 459.1966, 477.6057, 474.2420, 479.0948, ~
```

Our own CRD experiment, simulating the response

```
ggplot(crd,aes(treatment,response))+geom_point()
```



Statistical analysis, the means model

We can use maximum likelihood, which is equivalent to the method of least squares with the above assumptions, the estimates of the means are found by minimising the error sum of squares

$$ssE := \sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \mu_i)^2$$

Taking partial derivatives with respect to each μ_i and equating to zero we have

$$\frac{\partial ssE}{\partial \mu_i} = -2 \sum_{i=1}^{t} \sum_{j=1}^{r_i} (y_{ij} - \mu_i) = 0$$

which results in the estimates

$$\hat{\mu}_i = \bar{y}_{i\cdot} = \frac{1}{r_i} \sum_{j=1}^{r_i} y_{ij}$$

Must check the Hessian is indeed positive definite!

Statistical analysis, the means model

Consider a CRD with $t = 3$ levels and $r_i = 4$ replicates. We can write the means model using matrix notation as:

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\mu} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma^2\, \boldsymbol{I})$ and

$$\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \end{pmatrix}, \ \boldsymbol{Z} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \ \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \end{pmatrix}$$

Statistical analysis, the treatment effects model

Consider a CRD with $t = 3$ levels and $r_i = 4$ replicates. We can write the treatment effects model using matrix notation as:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma^2\, \boldsymbol{I})$ and

$$\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \end{pmatrix}, \; \boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \; \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}, \; \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \end{pmatrix}$$

Vector calculus recap

We will use the so-called denominator layout notation: $\boldsymbol{u}$ is a column vector and $\boldsymbol{u}^T$ is a row vector, g is a real function of $\boldsymbol{u}$

$$\frac{\partial g}{\partial \boldsymbol{u}} = \begin{pmatrix} \frac{\partial g}{\partial u_1} \\ \frac{\partial g}{\partial u_2} \\ \vdots \\ \frac{\partial g}{\partial u_p} \end{pmatrix}, \qquad \frac{\partial g}{\partial \boldsymbol{u}^T} = \left(\frac{\partial g}{\partial u_1}, \frac{\partial g}{\partial u_2}, \cdots, \frac{\partial g}{\partial u_p} \right)$$

$$\frac{\partial^2 g}{\partial \boldsymbol{u}^T \partial \boldsymbol{u}} = \begin{pmatrix} \frac{\partial^2 g}{\partial u_1 \partial u_1} & \frac{\partial^2 g}{\partial u_2 \partial u_1} & \cdots & \frac{\partial^2 g}{\partial u_p \partial u_1} \\ \frac{\partial^2 g}{\partial u_1 \partial u_2} & \frac{\partial^2 g}{\partial u_2 \partial u_2} & \cdots & \frac{\partial^2 g}{\partial u_p \partial u_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial u_1 \partial u_p} & \frac{\partial^2 g}{\partial u_2 \partial u_p} & \cdots & \frac{\partial^2 g}{\partial u_p \partial u_p} \end{pmatrix} = \frac{\partial^2 g}{\partial \boldsymbol{u} \partial \boldsymbol{u}^T}$$

Vector calculus recap

- Let $g(\boldsymbol{u}) = \boldsymbol{a}^T\boldsymbol{u}$ for some constant vector $\boldsymbol{a}$ then

$$\frac{\partial(\boldsymbol{a}^T\boldsymbol{u})}{\partial\boldsymbol{u}} = \frac{\partial(\boldsymbol{u}^T\boldsymbol{a})}{\partial\boldsymbol{u}} = \boldsymbol{a}$$

- Let $g(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u}$ for some constant matrix $\boldsymbol{A}$ then

$$\frac{\partial(\boldsymbol{A}\boldsymbol{u})}{\partial\boldsymbol{u}} = \boldsymbol{A}^T$$

- Let $g(\boldsymbol{u}) = \boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u}$ for some constant matrix $\boldsymbol{A}$ then

$$\frac{\partial(\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u})}{\partial\boldsymbol{u}} = 2\boldsymbol{A}\boldsymbol{u}$$

Statistical analysis, the means model

The Least Squares (LS) estimator for $\boldsymbol{\mu}$ is the one minimising the error sum of squares

$$ssE = \sum_{i=1}^{t}\sum_{j=1}^{r_i}(y_{ij} - \mu_i)^2 = \|\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\mu}\|^2 = (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\mu})^T(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\mu})$$

the solution is given by the so-called the *normal equations*:

$$\boldsymbol{Z^T Z}\hat{\boldsymbol{\mu}} = \boldsymbol{Z^T y}$$

since $\boldsymbol{Z^T Z}$ is non-singular then we have

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{Z^T Z})^{-1}\boldsymbol{Z^T y}$$

It is easy to verify that

$$\hat{\mu}_i = \bar{y}_{i\cdot} = \frac{1}{r_i}\sum_{j=1}^{r_i} y_{ij}$$

as verified above! We say that the model is **full rank**

Statistical analysis, the means model

```
r<-4; t<-3;levels<-c("level 1","level 2","level 3");
fact <- rep(levels,each = r) %>% factor()
Z <- model.matrix(~ fact-1); Z
   factlevel 1 factlevel 2 factlevel 3
1            1           0           0
2            1           0           0
3            1           0           0
4            1           0           0
5            0           1           0
6            0           1           0
7            0           1           0
8            0           1           0
9            0           0           1
10           0           0           1
11           0           0           1
12           0           0           1
attr(,"assign")
[1] 1 1 1
attr(,"contrasts")
attr(,"contrasts")$fact
[1] "contr.treatment"
```

Statistical analysis, the means model

```
mod.crd.means<-lm(response~treatment-1,data=crd)
summary(mod.crd.means)

Call:
lm(formula = response ~ treatment - 1, data = crd)

Residuals:
     Min       1Q   Median       3Q      Max
-21.6029  -6.9514   0.5019   6.4716  19.0177

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
treatmentlevel 1  480.800      3.015   159.5   <2e-16 ***
treatmentlevel 2  550.033      3.015   182.4   <2e-16 ***
treatmentlevel 3  494.618      3.015   164.0   <2e-16 ***
treatmentlevel 4  461.012      3.015   152.9   <2e-16 ***
treatmentlevel 5  489.149      3.015   162.2   <2e-16 ***
treatmentlevel 6  597.342      3.015   198.1   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.535 on 54 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 2.909e+04 on 6 and 54 DF,  p-value: < 2.2e-16
```

Statistical analysis, the means model

```
# compute treatment level means from original data
by_group <- group_by(crd, treatment)
means.crd<-summarize(by_group, means = mean(response))
glimpse(means.crd)
Rows: 6
Columns: 2
$ treatment <fct> level 1, level 2, level 3, level 4, level 5, level 6
$ means     <dbl> 480.7995, 550.0329, 494.6177, 461.0119, 489.1487, 597.3417
```

Statistical analysis, the treatment effects model

The LS estimator for β minimises the error sum of squares

$$ssE = \sum_{i=1}^{t}\sum_{j=1}^{r_i}(y_{ij} - \mu - \tau_i)^2 = \|\boldsymbol{y} - \boldsymbol{X}\beta\|^2 = (\boldsymbol{y} - \boldsymbol{X}\beta)^T(\boldsymbol{y} - \boldsymbol{X}\beta)$$

with corresponding normal equations:

$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\widehat{\beta} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$$

The problem now is that $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}$ is **singular** and cannot be inverted!

- Note $\boldsymbol{X}$ has $t+1$ columns!
- First column of $\boldsymbol{X}$ is the sum of the rest of the columns.
- $\boldsymbol{X'X}$ has rank$=t$
- We say that the treatment effects model is **not full rank**

Statistical analysis, the treatment effects model

By default, R makes the matrix $\boldsymbol{X}$ to have have full rank by dropping the column that corresponds to the first level of the factor

```
r<-4; t<-3;levels<-c("level 1","level 2","level 3");
fact <- rep(levels,each = r) %>% factor()
X <- model.matrix(~ fact); X
   (Intercept) factlevel 2 factlevel 3
1            1           0           0
2            1           0           0
3            1           0           0
4            1           0           0
5            1           1           0
6            1           1           0
7            1           1           0
8            1           1           0
9            1           0           1
10           1           0           1
11           1           0           1
12           1           0           1
attr(,"assign")
[1] 0 1 1
attr(,"contrasts")
attr(,"contrasts")$fact
[1] "contr.treatment"
```

Statistical analysis, the treatment effects model

```
mod.crd<-lm(response~treatment,data=crd)
summary(mod.crd)

Call:
lm(formula = response ~ treatment, data = crd)

Residuals:
     Min       1Q   Median       3Q      Max 
-21.6029  -6.9514   0.5019   6.4716  19.0177 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       480.800      3.015 159.463  < 2e-16 ***
treatmentlevel 2   69.233      4.264  16.237  < 2e-16 ***
treatmentlevel 3   13.818      4.264   3.241  0.00204 ** 
treatmentlevel 4  -19.788      4.264  -4.641 2.25e-05 ***
treatmentlevel 5    8.349      4.264   1.958  0.05540 .  
treatmentlevel 6  116.542      4.264  27.332  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.535 on 54 degrees of freedom
Multiple R-squared:  0.964, Adjusted R-squared:  0.9606 
F-statistic: 288.8 on 5 and 54 DF,  p-value: < 2.2e-16
```

Statistical analysis, the treatment effects model

```
coefs<-coef(mod.crd) # extracts coefficient values only
# reconstruct means from coefficients
taus<-c(0,coefs[2:length(coefs)])
means2<-coefs[1]+taus
means2
                 treatmentlevel 2 treatmentlevel 3 treatmentlevel 4
        480.7995         550.0329         494.6177         461.0119
treatmentlevel 5 treatmentlevel 6
        489.1487         597.3417
```

Statistical analysis of plywood experiment: linear model

```
mod.plywood<-lm(strength~glue,data=plywood)
summary(mod.plywood)

Call:
lm(formula = strength ~ glue, data = plywood)

Residuals:
   Min     1Q Median     3Q    Max 
-44.40 -17.82  -5.00  14.45  59.40 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  478.800      8.107  59.063  < 2e-16 ***
glueB         13.600     11.464   1.186   0.2407    
glueC         24.600     11.464   2.146   0.0364 *  
glueD         50.000     11.464   4.361 5.86e-05 ***
glueE         99.800     11.464   8.705 7.30e-12 ***
glueF        117.700     11.464  10.267 2.67e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25.64 on 54 degrees of freedom
Multiple R-squared:  0.7646,    Adjusted R-squared:  0.7428 
F-statistic: 35.08 on 5 and 54 DF,  p-value: 8.37e-16
```

Statistical analysis of plywood experiment: linear model

```
coefs<-coef(mod.plywood) # extracts coefficient values only
# reconstruct means from coefficients
taus<-c(0,coefs[2:length(coefs)])
means2<-coefs[1]+taus
means2
      glueB glueC glueD glueE glueF
478.8 492.4 503.4 528.8 578.6 596.5
# compute means from original data
by_group <- group_by(plywood, glue)
summaries.plywood<-summarize(by_group, means = mean(strength))

glimpse(summaries.plywood)
Rows: 6
Columns: 2
$ glue  <chr> "A", "B", "C", "D", "E", "F"
$ means <dbl> 478.8, 492.4, 503.4, 528.8, 578.6, 596.5
```

Estimability in the treatment effects model

- Theory says that normal equations:

$$\boldsymbol{X^T X}\widehat{\beta} = \boldsymbol{X^T y}$$

maybe have infinite number of solutions for $\widehat{\beta}$!!!

- Theory says any solution $\widehat{\beta}$ will be a linear function of $\boldsymbol{y}$
- We can focus on unbiased solutions!
- We will focus on finding unbiased estimators of linear combinations of the form

$$\boldsymbol{\lambda}^T \boldsymbol{\beta} = \lambda_0 \, \mu + \sum_{i=1}^{t} \lambda_i \tau_i$$

Note we write $\boldsymbol{\lambda}^T = (\lambda_0, \lambda_1, \ldots, \lambda_t)$

Estimability in the treatment effects model

- A linear combination of the parameters $\boldsymbol{\lambda}^T \beta$ is called **estimable** if

$$\boldsymbol{\lambda}^T \beta = E\left[\sum_{i=1}^{t}\sum_{j=1}^{r_i} a_{ij} Y_{ij}\right] = E[\boldsymbol{a}^T \boldsymbol{Y}], \qquad \text{for all } \beta$$

- Equivalently, a linear combination $\boldsymbol{\lambda}^T \beta$ is **estimable** if there exists an unbiased estimator of the form $\boldsymbol{a}^T \boldsymbol{Y}$
- Note that if $\boldsymbol{\lambda}^T \beta$ is estimable then

$$\boldsymbol{\lambda}^T \beta = \boldsymbol{a}^T E[\boldsymbol{Y}] = \boldsymbol{a}^T \boldsymbol{X} \beta, \qquad \text{for all } \beta$$

which implies that

$$\boldsymbol{\lambda} = \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{a}$$

which means $\boldsymbol{\lambda}$ belongs to the space generated by the rows of $\boldsymbol{X}$!

Estimability in the treatment effects model

Useful results for the CRD design

1. A linear combination $\boldsymbol{\lambda}^{\boldsymbol{T}}\beta$ is **estimable** if and only if ther exists a solution $\boldsymbol{R}$ to the system

$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\,\boldsymbol{r} = \boldsymbol{\lambda}$$

2. Both $E(\boldsymbol{Y}) = \boldsymbol{X}\beta$ and $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\beta$ are estimable
3. There are exactly t linearly independent estimable linear combinations $\boldsymbol{\lambda}_{\boldsymbol{1}}^{\boldsymbol{T}}\beta, \ldots, \boldsymbol{\lambda}_{\boldsymbol{t}}^{\boldsymbol{T}}\beta$
4. **Great Result!!** Let α be a vector of t linearly independent estimable combinations of β in the treatement effects model $\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon$. Then there exists a reparametrization to the full rank model $\boldsymbol{Y} = \boldsymbol{Z}\alpha + \epsilon$. Furthermore,

- any full rank reparametrization will give the same estimate of any estimable combination $\boldsymbol{\lambda}^{\boldsymbol{T}}\beta$
- a full rank reparametrization requires $\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{Z}$ is diagonal!

Estimability: Examples for the CRD design

- The means $\mu_i = \mu + \tau_i$ for $i = 1, .., t$ are estimable!
- Any difference $\tau_i - \tau_j$ for $i \neq j$ is estimable!
- **Definition**: Let $\boldsymbol{c}$ be known constant vector. A **contrast** is a linear combination

  $$\boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{\beta} = c_0\mu + \sum_{i=1}^{t} c_i\tau_i\,, \qquad \text{where} \quad \sum_{i=0}^{t} c_i = 0$$

  Contrasts are always estimable!
- The treatement effects themselves τ_i for $i = 1, .., t$ are not estimable!

Variances and covariances of estimable functions

Let $\boldsymbol{\lambda}_{\mathbf{1}}^{\boldsymbol{T}}\beta$ and $\boldsymbol{\lambda}_{\mathbf{2}}^{\boldsymbol{T}}\beta$ be two estimable linear combinations in the treatment effects (non full rank) model

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}\,, \qquad \epsilon \sim MVN_n(\mathbf{0}, \sigma^2\,\boldsymbol{I})$$

The corresponding variances and covariances are given by:

- $Var(\boldsymbol{a}_{\mathbf{1}}^{\boldsymbol{T}}\boldsymbol{Y}) = \sigma^2\,\boldsymbol{r}_1\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{r}_1$ where $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{r}_1 = \boldsymbol{\lambda}1$
- $Var(\boldsymbol{a}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{Y}) = \sigma^2\,\boldsymbol{r}_2\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{r}_2$ where $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{r}_2 = \boldsymbol{\lambda}2$
- $Cov(\boldsymbol{a}_{\mathbf{1}}^{\boldsymbol{T}}\boldsymbol{Y}, \boldsymbol{a}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{Y}) = \sigma^2\,\boldsymbol{r}_1\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{r}_2$

Maximum Likelihood Theory

Consider the treatment effects (non full rank) model

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}\,, \qquad \epsilon \sim MVN_n(\mathbf{0}, \sigma^2\,\boldsymbol{I})$$

Then the MLE of any estimable linear combination of β is the best possible unbiased estiamte with minimum variance where

$$\widehat{\sigma^2} = \frac{1}{n-t}(\boldsymbol{Y} - \boldsymbol{Z\alpha})^T(\boldsymbol{Y} - \boldsymbol{Z\alpha})$$

where $\boldsymbol{Y} = \boldsymbol{Z\alpha} + \boldsymbol{\epsilon}$ is a estimable reparametrization