

Unreplicated designs

Malachy Campbell

Sept. 14, 2020

Fixed vs random

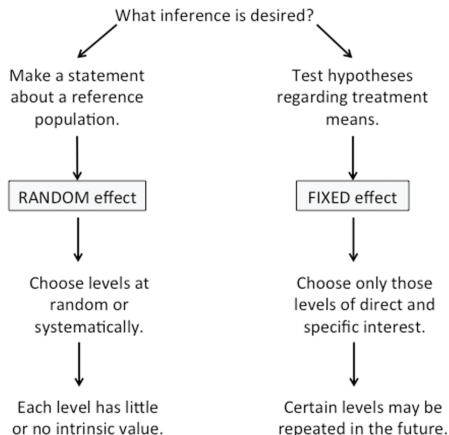


Fig. 8. Flow diagram illustrating the decision rule for fixed vs. random effects for each factor in an experiment. The first step is to ask the question about the desired inference, which leads to the decision of fixed vs. random. Following this decision, the experimental treatments can be chosen as desired to meet the chosen inference.

¹Casler, Michael D. "Fundamentals of experimental design: Guidelines for designing successful experiments." *Agronomy Journal* 107.2 (2015): 692-705.

Fixed vs random

- Five varieties (V1, V2, V3, V4, V5)
- Three Blocks (B1, B2, B3)

Equal no. observations for each variety

```
>
> exData
  Block Variety      Y
1    B1      V1 22.27660
2    B1      V2 18.84495
3    B1      V3 22.60057
4    B1      V4 14.71487
5    B1      V5 24.20826
6    B2      V1 21.49625
7    B2      V2 18.84885
8    B2      V3 20.17124
9    B2      V4 16.14260
10   B2      V5 21.11971
11   B3      V1 20.66123
12   B3      V2 14.97273
13   B3      V3 20.58808
14   B3      V4 14.81518
15   B3      V5 18.34671
>
```

Fixed vs random

Treating variety and block as fixed (BLUES)

```
>
> # Balanced #
> ## Both block and variety treated as fixed
> mod <- lm(Y ~ 0 + Variety + Block, exData)
> BLUES <- round(mod$coefficients, 1)[1:5]
> BLUES[order(BLUES)]
VarietyV4 VarietyV2 VarietyV3 VarietyV5 VarietyV1
  16.4      18.8      22.3      22.4      22.7
>
```

V4 < V2 < V3 < V5 < V1

Equal no. observations for each variety

```
>
> exData
  Block Variety      Y
1    B1      V1 22.27660
2    B1      V2 18.84495
3    B1      V3 22.60057
4    B1      V4 14.71487
5    B1      V5 24.20826
6    B2      V1 21.49625
7    B2      V2 18.84885
8    B2      V3 20.17124
9    B2      V4 16.14260
10   B2      V5 21.11971
11   B3      V1 20.66123
12   B3      V2 14.97273
13   B3      V3 20.58808
14   B3      V4 14.81518
15   B3      V5 18.34671
>
```

Fixed vs random

Treating variety and block as fixed (BLUEs)

```
>
> # Balanced #
> ## Both block and variety treated as fixed
> mod <- lm(Y ~ 0 + Variety + Block, exData)
> BLUEs <- round(mod$coefficients, 1)[1:5]
> BLUEs[order(BLUEs)]
VarietyV4 VarietyV2 VarietyV3 VarietyV5 VarietyV1
    16.4      18.8      22.3      22.4      22.7
>
```

V4 < V2 < V3 < V5 < V1

Treating variety as random (BLUPs) and block fixed

```
>
> LHS <- rbind(cbind(t(X) %>% X, t(X) %>% Z),
+              cbind(t(Z) %>% X, t(Z) %>% Z + lambda*I))
>
> RHS <- rbind(t(X) %>% exData$Y,
+              t(Z) %>% exData$Y)
>
> BLUPs <- solve(LHS) %>% RHS
> BLUPs <- round(BLUPs[4:8,], 2)
> BLUPs[order(BLUPs)]
VarietyV4 VarietyV2 VarietyV3 VarietyV5 VarietyV1
    -3.95    -1.70      1.74      1.84      2.08
>
```

V4 < V2 < V3 < V5 < V1

What if we have missing data (less information) for one variety (V1)?

Fixed vs random

- Five varieties (V1, V2, V3, V4, V5)
- Three Blocks (B1, B2, B3)

One obs. for V1, three obs. for all others

```
>
> # Unbalanced
> exData2 <- exData
> exData2 <- exData2[-c(6, 11),]
> exData2
  Block Variety      Y
1    B1      V1 22.27660
2    B1      V2 18.84495
3    B1      V3 22.60057
4    B1      V4 14.71487
5    B1      V5 24.20826
7    B2      V2 18.84885
8    B2      V3 20.17124
9    B2      V4 16.14260
10   B2      V5 21.11971
12   B3      V2 14.97273
13   B3      V3 20.58808
14   B3      V4 14.81518
15   B3      V5 18.34671
>
```

Fixed vs random

Treating variety and block as fixed (BLUES) – unbalanced

```
>
> ## Fixed block and variety
> mod <- lm(Y ~ 0 + Variety + Block, exData2)
> BLUES <- round(mod$coefficients, 1)[1:5]
> BLUES[order(BLUES)]
VarietyV4 VarietyV2 VarietyV1 VarietyV3 VarietyV5
      16.5      18.9      22.3      22.4      22.5
>
```

One obs. for V1, three obs. for all others

```
>
> # Unbalanced
> exData2 <- exData
> exData2 <- exData2[-c(6, 11), ]
> exData2
  Block Variety      Y
1    B1      V1 22.27660
2    B1      V2 18.84495
3    B1      V3 22.60057
4    B1      V4 14.71487
5    B1      V5 24.20826
7    B2      V2 18.84885
8    B2      V3 20.17124
9    B2      V4 16.14260
10   B2      V5 21.11971
12   B3      V2 14.97273
13   B3      V3 20.58808
14   B3      V4 14.81518
15   B3      V5 18.34671
>
```


Fixed vs random

Treating variety and block as fixed (BLUEs) – unbalanced

```
>
> ## Fixed block and variety
> mod <- lm(Y ~ 0 + Variety + Block, exData2)
> BLUEs <- round(mod$coefficients, 1)[1:5]
> BLUEs[order(BLUEs)]
VarietyV4 VarietyV2 VarietyV1 VarietyV3 VarietyV5
16.5      18.9      22.3      22.4      22.5
>
```

Treating variety and block as fixed (BLUEs) – balanced

```
>
> # Balanced #
> ## Both block and variety treated as fixed
> mod <- lm(Y ~ 0 + Variety + Block, exData)
> BLUEs <- round(mod$coefficients, 1)[1:5]
> BLUEs[order(BLUEs)]
VarietyV4 VarietyV2 VarietyV3 VarietyV5 VarietyV1
16.4      18.8      22.3      22.4      22.7
>
```

Ranking changes, but BLUEs are similar

Fixed vs random

Treating variety as random (BLUPs) – balanced

```
>
> LHS <- rbind(cbind(t(X) %>% X, t(X) %>% Z),
+             cbind(t(Z) %>% X, t(Z) %>% Z + lambda*I))
>
> RHS <- rbind(t(X) %>% exData$Y,
+             t(Z) %>% exData$Y)
>
> BLUPs <- solve(LHS) %>% RHS
> BLUPs <- round(BLUPs[4:8,], 2)
> BLUPs[order(BLUPs)]
VarietyV4 VarietyV2 VarietyV3 VarietyV5 VarietyV1
-3.95    -1.70     1.74     1.84     2.08
>
```

Treating variety as random (BLUPs) – unbalanced

```
>
> LHS <- rbind(cbind(t(X) %>% X, t(X) %>% Z),
+             cbind(t(Z) %>% X, t(Z) %>% Z + lambda*I))
>
> RHS <- rbind(t(X) %>% exData2$Y,
+             t(Z) %>% exData2$Y)
>
> BLUPs <- solve(LHS) %>% RHS
> BLUPs <- round(BLUPs[4:8,], 2)
> BLUPs[order(BLUPs)]
VarietyV4 VarietyV2 VarietyV1 VarietyV3 VarietyV5
-3.82    -1.58     1.57     1.86     1.96
>
```

Notice the ranking and BLUP for V1 changes

Fixed vs random

Treating variety as random (BLUPs) – balanced

```
>
> LHS <- rbind(cbind(t(X) %>% X, t(X) %>% Z),
+             cbind(t(Z) %>% X, t(Z) %>% Z + lambda*I))
>
> RHS <- rbind(t(X) %>% exData$Y,
+             t(Z) %>% exData$Y)
>
> BLUPs <- solve(LHS) %>% RHS
> BLUPs <- round(BLUPs[4:8,], 2)
> BLUPs[order(BLUPs)]
VarietyV4 VarietyV2 VarietyV3 VarietyV5 VarietyV1
-3.95    -1.70     1.74     1.84     2.08
>
```

Treating variety as random (BLUPs) – unbalanced

```
>
> LHS <- rbind(cbind(t(X) %>% X, t(X) %>% Z),
+             cbind(t(Z) %>% X, t(Z) %>% Z + lambda*I))
>
> RHS <- rbind(t(X) %>% exData2$Y,
+             t(Z) %>% exData2$Y)
>
> BLUPs <- solve(LHS) %>% RHS
> BLUPs <- round(BLUPs[4:8,], 2)
> BLUPs[order(BLUPs)]
VarietyV4 VarietyV2 VarietyV1 VarietyV3 VarietyV5
-3.82    -1.58     1.57     1.86     1.96
>
```

What if we have more data (more information) for one variety (V1)?

Fixed vs random

- Five varieties (V1, V2, V3, V4, V5)
- Three Blocks (B1, B2, B3)

Six obs. for V1, three obs. for all others

```
>
> exData3
  Block Variety      Y
1    B1      V1 22.27660
2    B1      V2 18.84495
3    B1      V3 22.60057
4    B1      V4 14.71487
5    B1      V5 24.20826
16   B1      V1 22.27660
6    B2      V1 21.49625
7    B2      V2 18.84885
8    B2      V3 20.17124
9    B2      V4 16.14260
10   B2      V5 21.11971
17   B2      V1 24.04695
11   B3      V1 20.66123
12   B3      V2 14.97273
13   B3      V3 20.58808
14   B3      V4 14.81518
15   B3      V5 18.34671
18   B3      V1 20.89857
>
```

Fixed vs random

**Treating variety as random
(BLUPs) – unbalanced data**

```
>
> LHS <- rbind(cbind(t(X) %>% X, t(X) %>% Z),
+             cbind(t(Z) %>% X, t(Z) %>% Z + lambda*I))
>
> RHS <- rbind(t(X) %>% exData3$Y,
+             t(Z) %>% exData3$Y)
>
> BLUPs <- solve(LHS) %>% RHS
> BLUPs <- round(BLUPs[4:8,], 2)
> BLUPs[order(BLUPs)]
VarietyV4 VarietyV2 VarietyV3 VarietyV5 VarietyV1
-4.05    -1.80      1.64      1.74      2.47
>
```

More observations for V1

**Six obs. for V1, three obs. for all
others**

```
>
> exData3
  Block Variety      Y
1    B1      V1 22.27660
2    B1      V2 18.84495
3    B1      V3 22.60057
4    B1      V4 14.71487
5    B1      V5 24.20826
16   B1      V1 22.27660
6    B2      V1 21.49625
7    B2      V2 18.84885
8    B2      V3 20.17124
9    B2      V4 16.14260
10   B2      V5 21.11971
17   B2      V1 24.04695
11   B3      V1 20.66123
12   B3      V2 14.97273
13   B3      V3 20.58808
14   B3      V4 14.81518
15   B3      V5 18.34671
18   B3      V1 20.89857
>
```

Fixed vs random

Treating variety as random (BLUPs) – unbalanced data

```
>
> LHS <- rbind(cbind(t(X) %>% X, t(X) %>% Z),
+             cbind(t(Z) %>% X, t(Z) %>% Z + lambda*I))
>
> RHS <- rbind(t(X) %>% exData3$Y,
+             t(Z) %>% exData3$Y)
>
> BLUPs <- solve(LHS) %>% RHS
> BLUPs <- round(BLUPs[4:8,], 2)
> BLUPs[order(BLUPs)]
VarietyV4 VarietyV2 VarietyV3 VarietyV5 VarietyV1
-4.05    -1.80      1.64      1.74      2.47
>
```

More observations for V1

Treating variety as random (BLUPs) – balanced data

```
>
> LHS <- rbind(cbind(t(X) %>% X, t(X) %>% Z),
+             cbind(t(Z) %>% X, t(Z) %>% Z + lambda*I))
>
> RHS <- rbind(t(X) %>% exData$Y,
+             t(Z) %>% exData$Y)
>
> BLUPs <- solve(LHS) %>% RHS
> BLUPs <- round(BLUPs[4:8,], 2)
> BLUPs[order(BLUPs)]
VarietyV4 VarietyV2 VarietyV3 VarietyV5 VarietyV1
-3.95    -1.70      1.74      1.84      2.08
>
```

Equal no. observations for all varieties

More observations → less shrinkage toward mean

$$y_{ijk} = \beta_i + u_j + e_{ijk}$$

- β_i : fixed effect for the i th block
- u_j : Deviation for the j th variety
 - $u_j \sim N(0, \sigma_u^2)$
- $e_{ijk} \sim N(0, \sigma_e^2)$

Linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$
$$\mathbf{u} \sim N(0, \mathbf{Z}\mathbf{G}\mathbf{Z}'\sigma_u^2); \quad \mathbf{e} \sim N(0, \mathbf{R}\sigma_e^2)$$

\mathbf{R} and \mathbf{G} are covariance matrices

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Linear mixed model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$
$$\mathbf{u} \sim N(0, \mathbf{Z}\mathbf{G}\mathbf{Z}'\sigma_u^2); \quad \mathbf{e} \sim N(0, \mathbf{R}\sigma_e^2)$$

\mathbf{R} and \mathbf{G} are covariance matrices

Solutions for fixed and random effects:

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

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})\end{aligned}$$

$$\begin{aligned}\mathbf{V} &= \mathbf{Z}\mathbf{G}\mathbf{Z}'\sigma_u^2 + \mathbf{R}\sigma_e^2 \\ &= \mathbf{Z}\mathbf{Z}'\sigma_u^2 + \mathbf{I}\sigma_e^2\end{aligned}$$

- If $\mathbf{G} = \mathbf{I}_{\text{no. var.} \times \text{no. var.}}$
- If $\mathbf{R} = \mathbf{I}_{\text{no. obs.} \times \text{no. obs.}}$

Linear mixed model

Solutions for fixed and random effects:

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

$$\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})$$

$$= \mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})$$

$$\mathbf{V} = \mathbf{ZGZ}'\sigma_u^2 + \mathbf{R}\sigma_e^2$$

$$= \mathbf{ZZ}'\sigma_u^2 + \mathbf{I}\sigma_e^2$$

```
>
> # Balanced Data
> Z <- model.matrix(~ 0 + Variety, exData)
> V = Z %>% t(Z) * gVar + diag(1, nrow(Z), nrow(Z)) * eVar
> round(t(Z) %>% solve(V), 2)
      1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
VarietyV1 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00
VarietyV2 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.32 0.00 0.00
VarietyV3 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.32 0.00
VarietyV4 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.32
VarietyV5 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.32
>
> # Unbalanced Data - one obs. V1
> Z <- model.matrix(~ 0 + Variety, exData2)
> V = Z %>% t(Z) * gVar + diag(1, nrow(Z), nrow(Z)) * eVar
> round(t(Z) %>% solve(V), 2)
      1  2  3  4  5  7  8  9 10 12 13 14 15
VarietyV1 0.9 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
VarietyV2 0.0 0.32 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.32 0.00 0.00
VarietyV3 0.0 0.00 0.32 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.32 0.00 0.00
VarietyV4 0.0 0.00 0.00 0.32 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.32 0.00
VarietyV5 0.0 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.32
>
> # Unbalanced Data - five obs. V1
> Z <- model.matrix(~ 0 + Variety, exData3)
> V = Z %>% t(Z) * gVar + diag(1, nrow(Z), nrow(Z)) * eVar
> round(t(Z) %>% solve(V), 2)
      1  2  3  4  5 16  6  7  8  9 10 17 11 12 13 14 15 18
VarietyV1 0.16 0.00 0.00 0.00 0.00 0.16 0.16 0.00 0.00 0.00 0.00 0.16 0.16 0.00 0.00 0.00 0.00 0.16
VarietyV2 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00
VarietyV3 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00
VarietyV4 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00
VarietyV5 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00 0.00 0.00 0.00 0.00 0.32 0.00
>
```

Importance of replication

T1	C	T3	T4	T2
T4	T3	C	T3	T3
T2	C	T4	T1	T2
T4	T1	T1	T2	C

Replication allows us to estimate experimental error (e)

- Error cannot be estimated ff no. treatments = no. of experimental units
 - We could use additional information to connect treatments and estimate error (e.g. genetic data, spatial data, etc.)

Importance of replication

K	S	O	E	L
G	A	Q	M	T
H	J	I	N	P
R	B	F	C	D

Sometimes we can't replicate treatments

- Treatments are too expensive
- Materials are not available
 - Treatments = experimental lines in breeding programs
 - Insufficient seed

Augmented designs

- Replicate a set of treatments (Checks)
- Get an estimate of experimental error from checks
- Use checks to “adjust” unreplicated treatments

B	M	T	J	R
K	I	CH	S	CH
P	F	G	N	H
E	CH	O	Q	L

Unreplicated (augmented block) designs

- Unreplicated designs - generally, any experiment where a few treatments are replicated and many more treatments are observed only once.
 - Augmented designs (Federer, 1956, 1975)¹
 - *p*-rep design (Cullis et al., 2006)²
- Breeders are interested in screening the progeny from many crosses (100s - 1000s of “treatments”)
- The goal is not precision – it is **scale**.
- Evaluate more lines in more environments and select the best performing lines

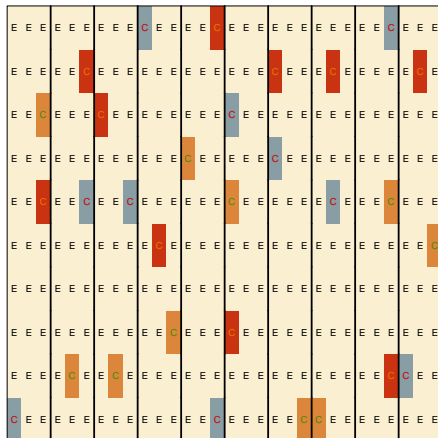
²Federer, Walter T., and D. Raghavarao. "On augmented designs." *Biometrics* (1975): 29-35.

³Cullis, Brian R., Alison B. Smith, and Neil E. Coombes. "On the design of early generation variety trials with correlated data." *Journal of agricultural, biological, and environmental statistics* 11.4 (2006): 381.

Unreplicated (augmented block) designs

Goal: Evaluate as many treatments as possible (large scale screening)

- 1 Split field into blocks
- 2 Choose a subset of treatments (checks) and allocate to each block (similar to a randomized complete block or randomized incomplete block design)
- 3 "Augment" blocks with unreplicated treatments



⁴Belamkar, Vikas, et al. "Genomic selection in preliminary yield trials in a winter wheat breeding program." G3: Genes, Genomes, Genetics 8.8 (2018): 2735-2747.

What is a check?

- A check is a set of replicated treatments
- A good check is one that allows experimental error to be adequately estimated
 - Adapted to the testing environment
 - Responsive to the environment – provides a better estimate of block-to-block variation
 - Well characterized

Analyzing augmented designs

- Should the replicated and unreplicated treatments be grouped together?
 - Are they both from the same population?
- Should checks be considered as fixed effects or random effects?
 - Do you want to estimate genetic variance for checks?

Analyzing augmented designs

- Should the replicated and unreplicated treatments be grouped together?
 - Are they both from the same population?
- Should checks be considered as fixed effects or random effects?
 - Do you want to estimate genetic variance for checks?

Both examples shown assume checks and entries are from different populations

Analyzing augmented designs – fixed checks and random entries

- Few checks (levels of fixed effect)
- Checks were deliberately selected
- Checks and entries are from different populations
 - Checks may be commercial varieties or advanced lines

Analyzing augmented designs – fixed checks and random entries

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$
$$i = (1, 2, \dots, C + 1); \quad j = (1, 2, \dots, E)$$

- For now, we are ignoring all other factors (blocks, locations, etc.)
- We have C checks and E entries
- The fixed effect (μ_i) includes a level for each check and a level that represents the overall average for all entries.
- The random effect for entry j is nested within the i th level of check ($u(\mu)_{ij}$)
 - Each entry will deviate from the fixed overall mean of the entries (μ_{C+1})

Example – fixed checks and random entries

Recoding the data

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2, \dots, C + 1); \quad j = (1, 2, \dots, E)$$

- Three check lines $C = 3$
each replicated three times
- Sixteen entries with one
observation each

Line	Y
Ck1	100.05
Ck1	101.67
Ck1	100.64
Ck2	120.25
Ck2	120.39
Ck2	121.87
Ck3	112.89
Ck3	110.28
Ck3	108.49
En1	115.12
En2	101.34
En3	109.24
En4	102.69
En5	106.97
En6	103.78
En7	104.47
En8	107.27
En9	108.10
En10	106.06
En11	104.61
En12	105.45
En13	98.30
En14	106.31
En15	97.14
En16	98.26

Example – fixed checks and random entries

Recoding the data

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2, \dots, C + 1); \quad j = (1, 2, \dots, E)$$

- 1 Create new factor for fixed effect (Chks)

Line	Y	Chks
Ck1	100.05	Ck1
Ck1	101.67	Ck1
Ck1	100.64	Ck1
Ck2	120.25	Ck2
Ck2	120.39	Ck2
Ck2	121.87	Ck2
Ck3	112.89	Ck3
Ck3	110.28	Ck3
Ck3	108.49	Ck3
En1	115.12	Ent
En2	101.34	Ent
En3	109.24	Ent
En4	102.69	Ent
En5	106.97	Ent
En6	103.78	Ent
En7	104.47	Ent
En8	107.27	Ent
En9	108.10	Ent
En10	106.06	Ent
En11	104.61	Ent
En12	105.45	Ent
En13	98.30	Ent
En14	106.31	Ent
En15	97.14	Ent
En16	98.26	Ent

Example – fixed checks and random entries

Recoding the data

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2, \dots, C + 1); \quad j = (1, 2, \dots, E)$$

- ① Create new factor for fixed effect (Chks)
- ② Create a dummy variable (New). 0 for all checks 1 for all entries
 - Allows us to estimate genetic variance for **only** entries

Line	Y	Chks	New
Ck1	100.05	Ck1	0
Ck1	101.67	Ck1	0
Ck1	100.64	Ck1	0
Ck2	120.25	Ck2	0
Ck2	120.39	Ck2	0
Ck2	121.87	Ck2	0
Ck3	112.89	Ck3	0
Ck3	110.28	Ck3	0
Ck3	108.49	Ck3	0
En1	115.12	Ent	1
En2	101.34	Ent	1
En3	109.24	Ent	1
En4	102.69	Ent	1
En5	106.97	Ent	1
En6	103.78	Ent	1
En7	104.47	Ent	1
En8	107.27	Ent	1
En9	108.10	Ent	1
En10	106.06	Ent	1
En11	104.61	Ent	1
En12	105.45	Ent	1
En13	98.30	Ent	1
En14	106.31	Ent	1
En15	97.14	Ent	1
En16	98.26	Ent	1

Example – fixed checks and random entries

Fitting the model

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2, \dots, C + 1); \quad j = (1, 2, \dots, E)$$

```
M1 <- asreml(fixed = Y ~ Chks,  
random = ~New:Line,  
data = M1data)
```

Line	Y	Chks	New
Ck1	100.05	Ck1	0
Ck1	101.67	Ck1	0
Ck1	100.64	Ck1	0
Ck2	120.25	Ck2	0
Ck2	120.39	Ck2	0
Ck2	121.87	Ck2	0
Ck3	112.89	Ck3	0
Ck3	110.28	Ck3	0
Ck3	108.49	Ck3	0
En1	115.12	Ent	1
En2	101.34	Ent	1
En3	109.24	Ent	1
En4	102.69	Ent	1
En5	106.97	Ent	1
En6	103.78	Ent	1
En7	104.47	Ent	1
En8	107.27	Ent	1
En9	108.10	Ent	1
En10	106.06	Ent	1
En11	104.61	Ent	1
En12	105.45	Ent	1
En13	98.30	Ent	1
En14	106.31	Ent	1
En15	97.14	Ent	1
En16	98.26	Ent	1

Example – fixed checks and random entries

Fitting the model

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2, \dots, C + 1); \quad j = (1, 2, \dots, E)$$

Variance component estimates:

	component	std.error	z.ratio	bound	%ch
New:Line	18.76	7.73	2.43	P	0.00
units!R	2.12	1.23	1.73	P	0.00

Line	Y	Chks	New
Ck1	100.05	Ck1	0
Ck1	101.67	Ck1	0
Ck1	100.64	Ck1	0
Ck2	120.25	Ck2	0
Ck2	120.39	Ck2	0
Ck2	121.87	Ck2	0
Ck3	112.89	Ck3	0
Ck3	110.28	Ck3	0
Ck3	108.49	Ck3	0
En1	115.12	Ent	1
En2	101.34	Ent	1
En3	109.24	Ent	1
En4	102.69	Ent	1
En5	106.97	Ent	1
En6	103.78	Ent	1
En7	104.47	Ent	1
En8	107.27	Ent	1
En9	108.10	Ent	1
En10	106.06	Ent	1
En11	104.61	Ent	1
En12	105.45	Ent	1
En13	98.30	Ent	1
En14	106.31	Ent	1
En15	97.14	Ent	1
En16	98.26	Ent	1

Example – fixed checks and random entries

Fitting the model

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2, \dots, C + 1); \quad j = (1, 2, \dots, E)$$

BLUES:

Name	BLUE
Chks_Ck1	0.00
Chks_Ck2	20.05
Chks_Ck3	9.77
Chks_Ent	3.91
(Intercept)	100.79

BLUPs:

Name	BLUP
New:Line_Ck1	0.00
New:Line_Ck2	0.00
New:Line_Ck3	0.00
New:Line_En1	9.37
New:Line_En10	1.23
New:Line_En11	-0.08
New:Line_En12	0.68
New:Line_En13	-5.74
New:Line_En14	1.45
New:Line_En15	-6.79
New:Line_En16	-5.78
New:Line_En2	-3.01
New:Line_En3	4.08
New:Line_En4	-1.80
New:Line_En5	2.04
New:Line_En6	-0.82
New:Line_En7	-0.20
New:Line_En8	2.31
New:Line_En9	3.06

Analyzing augmented designs – checks and entries are random effects

$$\begin{aligned}\hat{y}_{ij} &= \mu_i + u(\mu)_{ij} \\ i &= (1, 2) \\ j &= (1, 2, \dots, C) \text{ if } i = 1 \\ j &= (1, 2, \dots, E) \text{ if } i = 2\end{aligned}$$

- We have C checks and E entries
- The fixed effect (μ_i) has two levels: mean for all checks and mean for all entries.
- This will give us BLUPs for lines in each group
 - BLUPs are the deviation from their respective mean.
- Estimate genetic variances for each group

Example – checks and entries are random effects

Recoding the data

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2)$$

$$j = (1, 2, \dots, C) \text{ if } i = 1$$

$$j = (1, 2, \dots, E) \text{ if } i = 2$$

- 1 Create new factor for fixed effect (Type)

Line	Y	Type
Ck1	100.05	Ck
Ck1	101.67	Ck
Ck1	100.64	Ck
Ck2	120.25	Ck
Ck2	120.39	Ck
Ck2	121.87	Ck
Ck3	112.89	Ck
Ck3	110.28	Ck
Ck3	108.49	Ck
En1	115.12	Ent
En2	101.34	Ent
En3	109.24	Ent
En4	102.69	Ent
En5	106.97	Ent
En6	103.78	Ent
En7	104.47	Ent
En8	107.27	Ent
En9	108.10	Ent
En10	106.06	Ent
En11	104.61	Ent
En12	105.45	Ent
En13	98.30	Ent
En14	106.31	Ent
En15	97.14	Ent
En16	98.26	Ent

Example – checks and entries are random effects

Recoding the data

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2)$$

$$j = (1, 2, \dots, C) \text{ if } i = 1$$

$$j = (1, 2, \dots, E) \text{ if } i = 2$$

- 1 Create new factor for fixed effect (Type)
- 2 Create a dummy variables for each random effect

Line	Y	Type	ch_DV	new_DV
Ck1	100.05	Ck	1	0
Ck1	101.67	Ck	1	0
Ck1	100.64	Ck	1	0
Ck2	120.25	Ck	1	0
Ck2	120.39	Ck	1	0
Ck2	121.87	Ck	1	0
Ck3	112.89	Ck	1	0
Ck3	110.28	Ck	1	0
Ck3	108.49	Ck	1	0
En1	115.12	Ent	0	1
En2	101.34	Ent	0	1
En3	109.24	Ent	0	1
En4	102.69	Ent	0	1
En5	106.97	Ent	0	1
En6	103.78	Ent	0	1
En7	104.47	Ent	0	1
En8	107.27	Ent	0	1
En9	108.10	Ent	0	1
En10	106.06	Ent	0	1
En11	104.61	Ent	0	1
En12	105.45	Ent	0	1
En13	98.30	Ent	0	1
En14	106.31	Ent	0	1
En15	97.14	Ent	0	1
En16	98.26	Ent	0	1

Analyzing augmented designs – checks and entries are random effects

Fitting the model

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2)$$

$$j = (1, 2, \dots, C) \text{ if } i = 1$$

$$j = (1, 2, \dots, E) \text{ if } i = 2$$

```
M2 <- asreml(fixed = Y ~ Type,  
random = ~ch_DV:Line +  
new_DV:Line,  
data = M2data)
```

Line	Y	Type	ch_DV	new_DV
Ck1	100.05	Ck	1	0
Ck1	101.67	Ck	1	0
Ck1	100.64	Ck	1	0
Ck2	120.25	Ck	1	0
Ck2	120.39	Ck	1	0
Ck2	121.87	Ck	1	0
Ck3	112.89	Ck	1	0
Ck3	110.28	Ck	1	0
Ck3	108.49	Ck	1	0
En1	115.12	Ent	0	1
En2	101.34	Ent	0	1
En3	109.24	Ent	0	1
En4	102.69	Ent	0	1
En5	106.97	Ent	0	1
En6	103.78	Ent	0	1
En7	104.47	Ent	0	1
En8	107.27	Ent	0	1
En9	108.10	Ent	0	1
En10	106.06	Ent	0	1
En11	104.61	Ent	0	1
En12	105.45	Ent	0	1
En13	98.30	Ent	0	1
En14	106.31	Ent	0	1
En15	97.14	Ent	0	1
En16	98.26	Ent	0	1

Analyzing augmented designs – checks and entries are random effects

Fitting the model

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2)$$

$$j = (1, 2, \dots, C) \text{ if } i = 1$$

$$j = (1, 2, \dots, E) \text{ if } i = 2$$

Variance component estimates:

	comp.	std.error	z.ratio	bound	%ch
ch_DV:Line	99.82	100.42	0.99	P	0.10
new_DV:Line	18.77	7.73	2.43	P	0.00
units!R	2.12	1.23	1.73	P	0.00

Line	Y	Type	ch_DV	new_DV
Ck1	100.05	Ck	1	0
Ck1	101.67	Ck	1	0
Ck1	100.64	Ck	1	0
Ck2	120.25	Ck	1	0
Ck2	120.39	Ck	1	0
Ck2	121.87	Ck	1	0
Ck3	112.89	Ck	1	0
Ck3	110.28	Ck	1	0
Ck3	108.49	Ck	1	0
En1	115.12	Ent	0	1
En2	101.34	Ent	0	1
En3	109.24	Ent	0	1
En4	102.69	Ent	0	1
En5	106.97	Ent	0	1
En6	103.78	Ent	0	1
En7	104.47	Ent	0	1
En8	107.27	Ent	0	1
En9	108.10	Ent	0	1
En10	106.06	Ent	0	1
En11	104.61	Ent	0	1
En12	105.45	Ent	0	1
En13	98.30	Ent	0	1
En14	106.31	Ent	0	1
En15	97.14	Ent	0	1
En16	98.26	Ent	0	1

Analyzing augmented designs – checks and entries are random effects

Fitting the model

$$\hat{y}_{ij} = \mu_i + u(\mu)_{ij}$$

$$i = (1, 2)$$

$$j = (1, 2, \dots, C) \text{ if } i = 1$$

$$j = (1, 2, \dots, E) \text{ if } i = 2$$

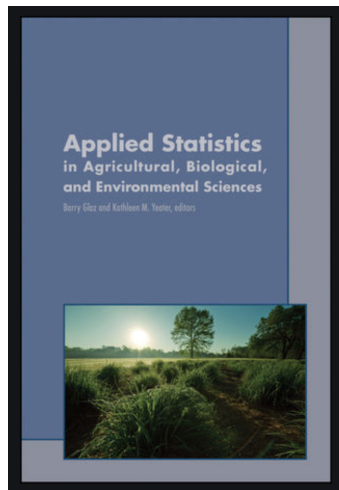
BLUES:

	BLUE
Type_Ck	0.00
Type_Ent	-6.03
(Intercept)	110.73

BLUPs:

	BLUP
ch_DV:Line_Ck1	-9.87
ch_DV:Line_Ck2	10.04
ch_DV:Line_Ck3	-0.17
ch_DV:Line_En1	0.00
	:
	:
ch_DV:Line_En9	0.00
new_DV:Line_Ck1	0.00
new_DV:Line_Ck2	0.00
new_DV:Line_Ck3	0.00
new_DV:Line_En1	9.37
new_DV:Line_En10	1.23
	:
	:
new_DV:Line_En12	0.68

Further reading



Can be accessed for free via Cornell libraries:

<https://access.onlinelibrary.wiley.com/doi/book/10.2134/appliedstatistics>