A pragmatic approach to formulating linear mixed models for randomized experiments

Prof. Dr. Hans-Peter Piepho

FG Biostatistics

Universität Hohenheim

Content

1. Introduction

- Analysis of designed experiments by software for linear models
- Often need a mixed linear model
- Main task for user: Come up with the right model
- Many practically relevant cases not found in text books

Objective:

- Develop simple rules for setting up a model
- Rules to be flexible
- Fitted models used to compute „the phenotype“

2. Randomized complete block design

Linear model for complete blocks

$y_{ij} = \mu + \alpha_i + b_j + e_{ij}$

where

y_{ij} = observed value for *i*-th treatment in *j*-th block

μ = general effect (intercept)

α_i = main effect of *i*-th treatment

b_j = main effect of *j*-th block

e_{ij} = residual error of y_{ij} (random)

$e_{ij} \sim N(0, \sigma^2)$

Null model or **block model**

$\Rightarrow$ no treatment effects

$y_{ij} = \mu + b_j + e_{ij}$

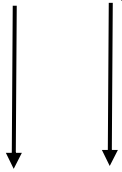
- Valid model in case no treatments or the same treatment on all plots/experimental units

Coding for statistical package (e.g. SAS)

```
GEN   BLOCK    YIELD

 2     1       34
 1     1       36
 3     1       41
 4     1       46
 4     2       35
 3     2       33
 1     2       38
 2     2       51
```

$$y_{ij} = \mu + \alpha_i + b_j + e_{ij}$$

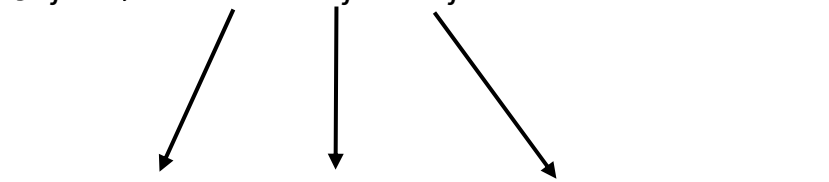


```
MODEL YIELD =  GEN BLOCK;
```

A slightly more elaborate syntax

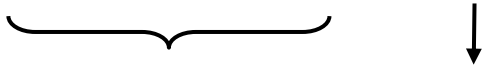
(Patterson, 1997)

$y_{ij} = \mu + \alpha_i + b_j + e_{ij}$



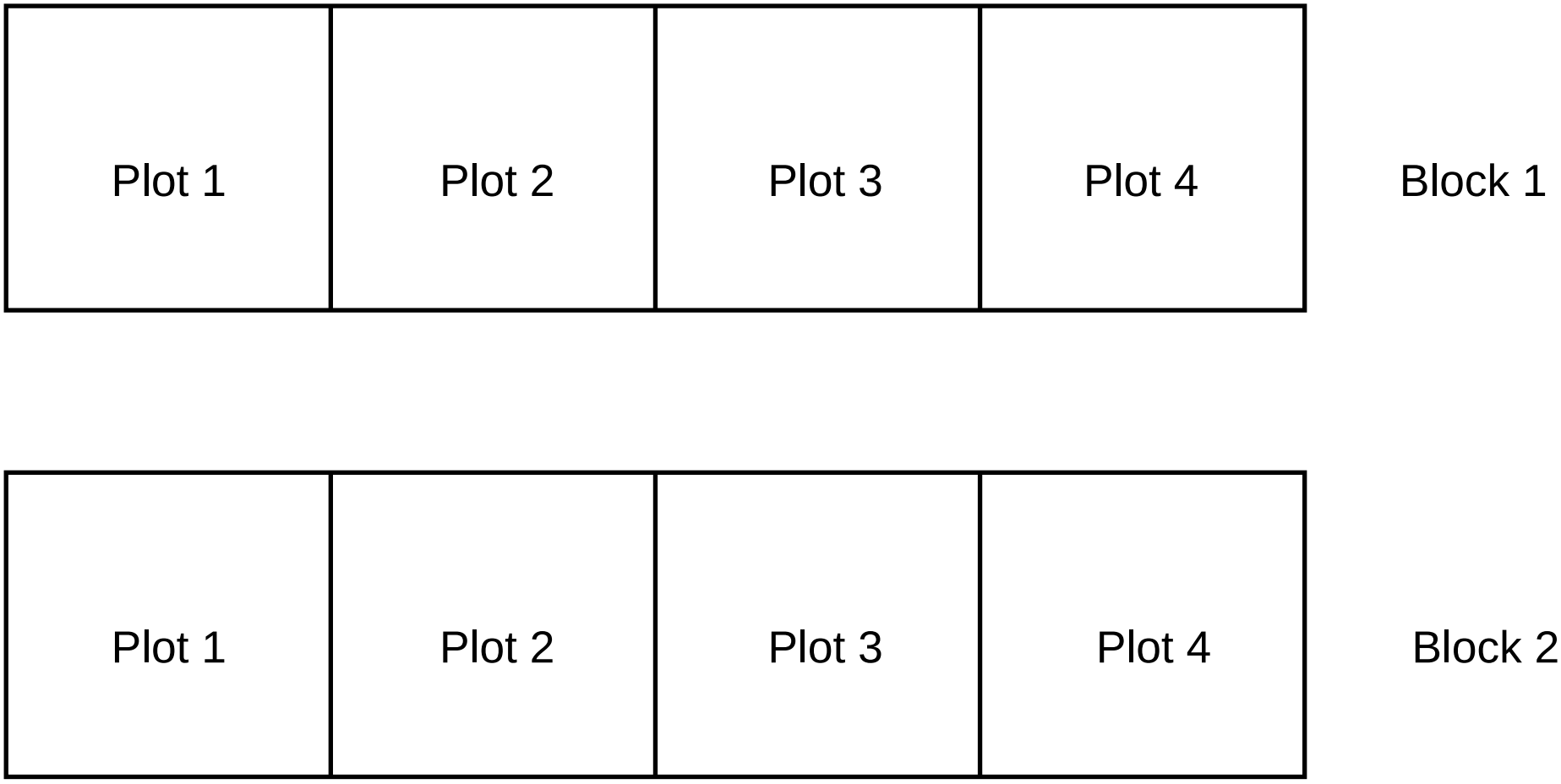
Y = GEN + BLOCK : GEN•BLOCK

"i" "j" : "i•j"



fixed : random

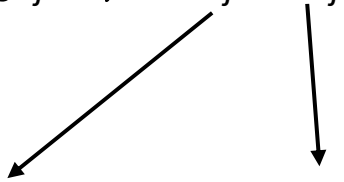
Block design without treatments (⇔ null model)



Block model (null model, design model)

GEN	BLOCK	PLOT	YIELD
2	1	1	34
1	1	2	36
3	1	3	41
4	1	4	46
4	2	1	35
3	2	2	33
1	2	3	38
2	2	4	51

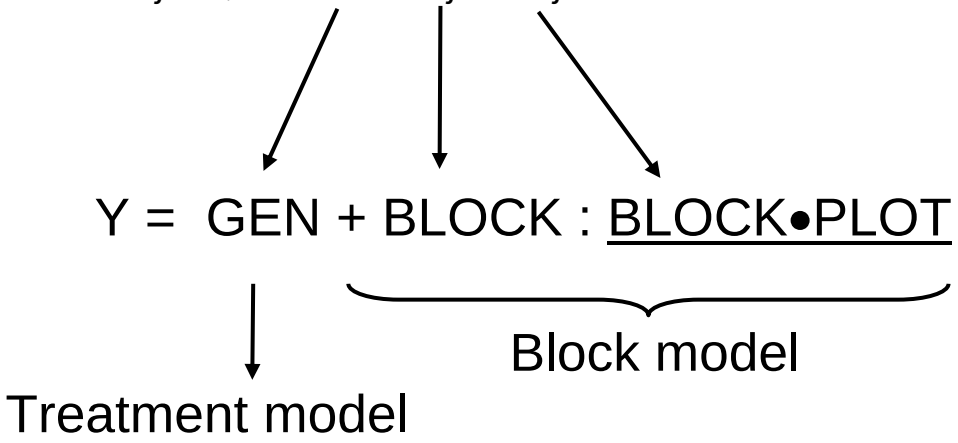
$$y_{ij} = \mu + b_j + e_{ij}$$



BLOCK : BLOCK•PLOT

Full model

$$y_{ij} = \mu + \alpha_i + b_j + e_{ij}$$



- BLOCK•PLOT is random because PLOT is randomization unit

- BLOCK•PLOT is underlined because it identifies observational unit (remainder) $\Rightarrow$ Need not be explicitly specified in statisical package

Summary

Treatment model:

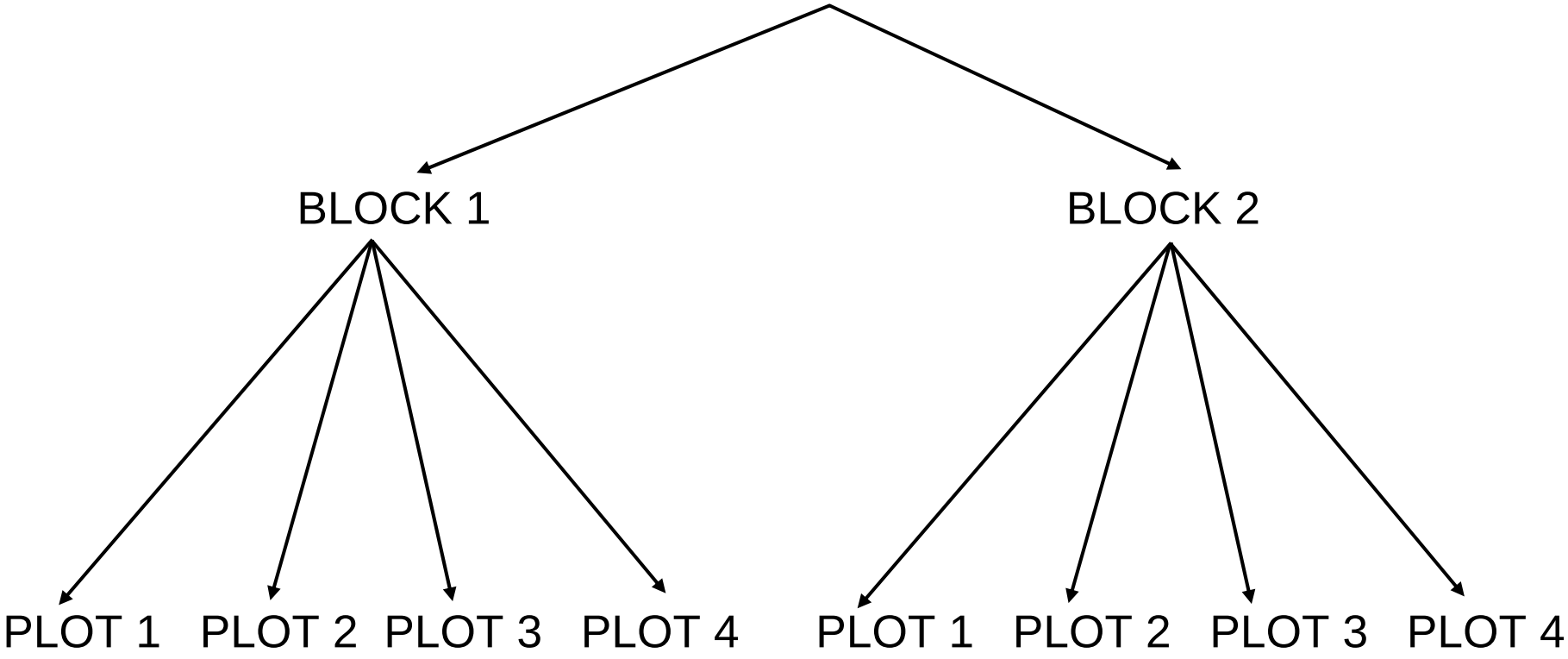
- Levels of treatment factor randomly allocated to experimental units
- Set up treatment model for treatment factors

Block model (design model):

- Block model is for observational units, irrespective of treatments applied
- Effects in block model are innate to observational units
- Can set up block model independently from treatment model

Nesting

Plots (PLOT) are nested within blocks (BLOCK):



Nesting operator

Block model:

BLOCK/PLOT = BLOCK : BLOCK•PLOT

(BLOCK•PLOT is random because PLOT is a randomization unit)

Full model:

GEN + BLOCK/PLOT = GEN + BLOCK : BLOCK•PLOT

3. Four operators

Dot operator (•):

(i) A•B = B•A

(ii) (A•B)•C = A•(B•C)

(iii) (A•C)•(B•C) = A•B•C

Product term operator [pt(.)]:

(i) pt(M) = A•B•C•…

= product term of all factors appearing in model M (A, B, C, …)

Nesting operator (/):

(i) A/B = A + A•B (A and B are factors)

(ii) A/B = A + pt(A)•B (A and B may be models)

(iii) A/(B/C) = (A/B)/C

(iv) A/(B+C) = A/B + A/C

Crossing operator (×):

(i) A×B = A + B + A•B

(ii) A×(B+C) = A + B + C + A•B + A•C

(iii) (A×B)/C = A×B + A•B•C

4. Six rules

Rule 1 (when is a factor random?)

- A factor is random

  - when the observed levels are a random sample from a defined population or

  - when it represents a randomization unit (error stratum)

- Otherwise a factor is usually considered as fixed
- When levels of a factor are to be compared (e.g. treatments), the factor is fixed, independently of whether or not it is random by design
- When a factor is random, all effects containing it are random

Example

- Series of experiments with selected set of plant varieties
- Experiments conducted at a sample of sites selected at random from target region
- Want to estimate variety means in target region

$\Rightarrow$ factors: variety and site

$\Rightarrow$ varieties fixed

$\Rightarrow$ sites random

Rule 2 (two types of factor)

Block factors:

- Randomly selected sampling units (plants, soil specimens, etc.)
- Randomisation units (rows, columns, incomplete blocks, main plots, sub plots, etc.)
- Block units, which are not themselves involved in randomization process (complete blocks, environments, etc.).

Block factors are *innate* to observational units

Treatment factors:

- Selected by experimenter to answer research question
- Levels of treatment factor are randomly allocated to observational units by a defined randomization procedure

Treatments are not *innate* to observational units

Rule 3 (Keep treatment model and block model separate)

- When setting up a full model, it is useful to (at least initially) keep treatment model and block model separete
- The treatment model can be formulated with treatment factors alone
- The block model can be formulated with block factors alone (i.e. without treatment factors!)

Rule 4 (effects of block model)

Random effects for randomization units:

- Each randomization unit (every error stratum) has its own effect
- An experimental unit or block unit becomes a randomization unit when levels of a treatment factor are randomly allocated to it
- Crossing of randomization units (error strata) produces further error effects

Fixed effects for block factors, which are not themselves part of the randomization or sampling process

Rule 5 (coding of block factors)

- Every block factor (design factor) is represented by a separate variable
- Sometimes can replace block factor by treatment factor in an effect, but this is not a necessity!
- For clarity it is better to avoid coding of a block factor by a treatment factor

5. α-design (special case: lattice design)

Replicate 1						Replicate 2						Replicate 3					
1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
11	21	23	13	17	6	8	24	12	5	2	19	11	2	17	12	21	3
4	10	14	3	15	12	20	15	11	9	18	7	1	15	18	13	22	5
5	20	16	19	7	24	14	3	21	10	13	6	14	9	4	10	16	20
22	2	18	8	1	9	4	23	17	1	22	16	19	8	6	23	24	7

plot

incomplete block

Block model:

REP/BLOCK/PLOT = REP : REP•BLOCK + REP•BLOCK•PLOT

Treatment model:

GEN

Full model

GEN + REP : REP•BLOCK + REP•BLOCK•PLOT

6. Multiple lattices

Replicate 1

Block	Lattice 1					Lattice 2					Lattice 3					Lattice 4					Lattice 5				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	14	6	C1	C3	11	23	33	39	35	29	C1	C5	42	57	50	C2	71	80	66	68	C1	85	91	C3	82
	1	C4	12	10	C2	C4	34	C5	27	37	58	60	55	C2	43	76	74	67	63	C4	93	C2	86	90	C5
	9	18	2	8	4	24	22	30	C2	C1	53	56	59	41	45	64	C1	C5	77	61	98	84	100	97	87
	13	5	19	15	3	26	21	38	32	25	49	51	46	44	47	70	78	69	C3	75	83	88	C4	94	89
	C5	17	7	16	20	40	C3	28	31	36	52	58	C3	54	C4	62	73	65	79	72	92	99	95	81	96

Replicate 2

Block	Lattice 1					Lattice 2					Lattice 3					Lattice 4					Lattice 5				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	C3	15	17	13	2	22	C3	27	C4	C2	C2	53	44	C4	58	C1	68	61	73	64	100	87	92	91	C5
	12	C4	7	C1	1	24	37	21	29	34	50	60	42	52	C5	72	69	66	67	75	90	94	96	98	95
	3	4	C5	11	C2	36	35	38	32	23	55	41	45	56	59	76	74	C5	C4	71	89	86	88	82	83
	9	19	16	18	5	C5	26	C1	33	30	51	C3	48	54	57	65	C2	62	70	C3	85	C2	C4	C3	99
	6	14	20	8	10	31	28	40	39	25	49	43	C1	46	47	79	63	78	67	80	C1	93	81	84	97

Figure: 5 simple lattices (5 x 5) with block size *k* = 5 for *v* = 100 genotypes and **5 checks (C1-C5)**

Factors

Design factors:

LAT = lattice
REP = replicate
BLK = incomplete block
PLT = plot

Treatment factors:

GEN = genotypes

Block model:

LAT/REP/BLK/PLT

Treatment model:

GEN

7. Series of experiments

Rule 6 (Interaction between block and treatment factors)

- Usually assume block-treatment additivity (no interaction)
- Check if interaction is to be expected
- Interaction likely when levels of a block factor are very diverse

Series of RCBD experiments

- RCBD for varieties (GENO)
- Experiments ar several sites (SITE)

Block model: SITE/BLOCK/PLOT = SITE + SITE•BLOCK + SITE•BLOCK•PLOT

Treatment model: GENO

Full model: GENO : SITE/BLOCK/PLOT

= GENO : SITE + SITE•BLOCK + SITE•BLOCK•PLOT

Expect GENO•SITE interaction ⇒

GENO : SITE + GENO•SITE + SITE•BLOCK + SITE•BLOCK•PLOT

Multiple lattices in several environments

Replicate 1

Block	Lattice 1					Lattice 2					Lattice 3					Lattice 4					Lattice 5				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	14	6	C1	C3	11	23	33	39	35	29	C1	C5	42	57	50	C2	71	80	66	68	C1	85	91	C3	82
	1	C4	12	10	C2	C4	34	C5	27	37	58	60	55	C2	43	76	74	67	63	C4	93	C2	86	90	C5
	9	18	2	8	4	24	22	30	C2	C1	53	56	59	41	45	64	C1	C5	77	61	98	84	100	97	87
	13	5	19	15	3	26	21	38	32	25	49	51	46	44	47	70	78	69	C3	75	83	88	C4	94	89
	C5	17	7	16	20	40	C3	28	31	36	52	58	C3	54	C4	62	73	65	79	72	92	99	95	81	96

Replicate 2

Block	Lattice 1					Lattice 2					Lattice 3					Lattice 4					Lattice 5				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	C3	15	17	13	2	22	C3	27	C4	C2	C2	53	44	C4	58	C1	68	61	73	64	100	87	92	91	C5
	12	C4	7	C1	1	24	37	21	29	34	50	60	42	52	C5	72	69	66	67	75	90	94	96	98	95
	3	4	C5	11	C2	36	35	38	32	23	55	41	45	56	59	76	74	C5	C4	71	89	86	88	82	83
	9	19	16	18	5	C5	26	C1	33	30	51	C3	48	54	57	65	C2	62	70	C3	85	C2	C4	C3	99
	6	14	20	8	10	31	28	40	39	25	49	43	C1	46	47	79	63	78	67	80	C1	93	81	84	97

Figure: 5 simple lattices (5 x 5) with block size *k* = 5 for *v* = 100 genotypes and **5 checks (C1-C5)**

Design factors:

ENV = environment
LAT = lattice
REP = replicate
BLK = incomplete block
PLT = plot

Treatment factors:

GEN = genotypes

Block model:

ENV/LAT/REP/BLK/PLT

Treatment model:

GEN

Interaction:

GEN•ENV

Block model resolved:

ENV/LAT/REP/BLK/PLT =

ENV + ENV•LAT + ENV•LAT•REP + ENV•LAT•REP•BLK +
ENV•LAT•REP•BLK•PLT

Full model:

GEN : GEN•ENV + ENV + ENV•LAT + ENV•LAT•REP + ENV•LAT•REP•BLK +
ENV•LAT•REP•BLK•PLT

8. Blocking out checks

- 5 checks (fixed)
- 1000 DHs test-crossed to tester of opposite pool (random)
- 2 testers, 500 test crosses each

	Factor/covariate	
GRP	GEN	SWITCH
Check1	1001	0
Check2	1002	0
Check3	1003	0
Check4	1004	0
Check5	1005	0
Tester1	1-500	1
Tester2	501-1000	1

Model:

GRP**/**GEN = GRP **:** GRP•GEN

Want no random effects for checks:

GRP **:** SWITCH•GRP•GEN

(Fits unchanged, but variance-structures and hence standard errors differ!!)

Blocking out checks – ridge regression:

- Want to do ridge regression (GBLUP) using marker matrix Z
- Provide known variance-covariance matrix $\Gamma = ZZ^T$ (realized relationship!)
- Do not want to / cannot include checks in Γ

$\Rightarrow$ arbitrarily set GEN equal to level of a non-check for all checks (GEN2)
$\Rightarrow$ can do this because SWITCH=0 for checks!!!

	Factor/covariate		
GRP	GEN	GEN2	SWITCH
Check1	1001	1	0
Check2	1002	1	0
Check3	1003	1	0
Check4	1004	1	0
Check5	1005	1	0
Tester1	1-500	1-500	1
Tester2	501-1000	501-1000	1

Model across environments:

- Multiple lattice experiments
- Replace GEN with GRP **:** SWITCH•GRP•GEN2

Block model:

ENV/LAT/REP/BLK/PLT

Treatment model:

GRP **:** SWITCH•GRP•GEN2

Interaction:

(GRP **:** SWITCH•GRP•GEN2)•ENV = GRP•ENV + SWITCH•GRP•GEN2•ENV

Full model:

GRP : SWITCH•GRP•GEN2 + GRP•ENV + SWITCH•GRP•GEN2•ENV +

ENV + ENV•LAT + ENV•LAT•REP + ENV•LAT•REP•BLK +
ENV•LAT•REP•BLK•PLT

Literature

Brien, C. J., 1983: Analysis of variance tables based on experimental structure. *Biometrics* **39**, 53-59.

Brien, C. J., Demetrio, C.G.B., 2010: Formulating mixed models for experiments, including longitudinal experiments. *Journal of Agricultural, Biological and Environmental Statistics* **14**, 253-180.

Nelder, J. A., 1965: The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. II. Treatment structure and the general analysis of variance. *Proceedings of the Royal Society of London A* **283**, 147-178.

Patterson, H. D., 1997: Analysis of series of variety trials. In: R. A. Kempton, and P. N. Fox (eds.), Statistical methods for plant variety evaluation, pp. 139-161. Chapman and Hall, London.

Payne, R. W., and G. N. Wilkinson, 1977: A general algorithm for analysis of variance. *Applied Statistics* **26**, 251-260.

Piepho, H.P., Büchse, A., Emrich, K., 2003: A hitchhiker's guide to the mixed model analysis of randomized experiments. *Journal of Agronomy and Crop Science* **189**, 310-322.

Piepho, H.P., Büchse, A., Richter, C., 2004: A mixed modelling approach to randomized experiments with repeated measures. *Journal of Agronomy and Crop Science* **190**, 230-247.