1. What is the **experimental unit**? WHAT IS RANDOMIZED

2. Are we taking **multiple measures** of the SAME e.u.?

    a. Subsamples AVERAGE FIRST!

    b. Repeated measures in time

3. Are there **blocking variables?**

    a. No -> CRD

    b. One->RCBD

        i.   One observation per T*Block -> TUKEY ADDITIVITY TEST

        ii. More than one per Trt*Block -> EXPLORATORY MODEL

    c. More than one blocking variable

        i.   Only one Trt per Blocks combination -> LATIN SQUARE

        ii. All Trt in each blocks combination -> TWO BLOCKING VARIABLES

4. **More than one Treatment**?

    a. Not all combinations

    b. All levels of one Factor at all levels of the other Factor->FACTORIAL

        i.   Missing data LSMEANS & TYPE IIISS

        ii. Random factors MIXED or RANDOM MODELS Random /test

        iii. Splits SPLIT PLOT, STRIP PLOT, SPLIT-SPLIT PLOT

    c. Significant Interaction -> SIMPLE EFFECTS

    d. Not significant interactions -> MAIN EFFECTS

5. **Covariable** LSMEANS & TYPE IIISS

    If too complex transform y into Z and then do normal ANOVA

ANCOVA for a 3x2 Factorial arranged as an RCBD

Data ABFact;
    Input Block A B X Y @@;
    Z = Y – Beta * (X - Xmean);   ← Beta from "solution" Xmean from "X ANOVA"
Cards;
...
;
Proc GLM;   ← Individual **ANOVAs**
    Class Block A B;
    Model X Y = Block A B A*B;
Proc GLM;   ← **ANCOVA** (Type III SS & LS means) If split?
    Class Block A B;
    Model Y = Block A B A*B X / solution;
    LSMeans A;
    LSMeans B;
    Contrast   'A linear'   A   -1 0 1;
Run; **Quit**;

If A*B interaction NS → main effects, if significant→ simple effects

1. To test normality of residuals

Proc GLM;   ← ANOVA on Z
    Class Block A B;
    Model Z = Block A B A*B;
    Output out = ABFactPRz p = Pred r = Res; ← Output can be in ANCOVA
Proc Univariate Data = ABFactPRz normal;
    Var Res;

2. To test homogeneity of variances

Create a Treatment ID (TRT) with 6 levels, one for each A-B combination (if interaction *)
or separate GLM and Levene tests for A and B (if interaction NS)

Proc GLM;
    Class TRT;
    Model Z = TRT;
    Means TRT / hovtest = Levene;   (or separate Levene test for A and B if there is no interaction)

3. To test homogeneity of slopes

Proc GLM;
    Class Block TRT;
    Model Y = Block TRT X TRT*X;

or in no interaction different **Proc GLM**
    Model Y = Block A X A*X; and Model Y = Block B X B*X;

4. To partition the A*B interaction

Proc GLM <mark>order=data</mark>; BE CAREFUL WITH DARTA ORDER!
    Class Block TRT;
    Model Y = Block TRT X;
    LSMeans TRT;
    Contrast 'A Linear'   TRT   -1 0 1 -1 0 1;
    Contrast 'A Quadratic' TRT   1 -2 1 1 -2 1;
    Contrast 'B'   TRT   1 1 1 -1 -1 -1;
    Contrast 'A Lin x B'   TRT   -1 0 1 1 0 -1;
    Contrast 'A Quad x B' TRT   1 -2 1 -1 2 -1;

---

5. To analyze a **mixed model or split plot** with covariable
Example: Factor A random, Factor B fixed
    Model Y = Block A B A*B X;
    Random Block A A*B / test;

ANOVA without X in the model

Source	Type III Expected Mean Square
Block	Var(Error) + 6 Var(Block)
A	Var(Error) + 4 Var(A*B) + 8 Var(A)
B	Var(Error) + 4 Var(A*B) + Q(B)
A*B	Var(Error) + 4 Var(A*B)

Test   h = A   e = A*B;
Contrast 'A Linear' A **-1 0 1** /e = A*B;

With X in the model the coefficients are adjusted by the covariable and the <mark>test statement is required</mark>. Contrasts and mean comparisons require Proc Mixed

Alternative: Transform Y into Z using the covariable and then used the simple model and error terms in Z (from the mixed model or from the split plot).

Model Z = Block A B A*B; with the EMS as above.
Test   h = A   e = A*B;
Contrast 'A Linear' A **-1 0 1** /e = A*B;

Mixed Model: Blocks nested within Random Locations

Five varieties of barley are tested in three locations selected at random across the Sacramento Valley. In each location, the experiment is organized as an RCBD with four blocks. The intention is to:

- **a.** Characterize the variability across the valley.
- **b.** Characterize the stability of variety performance across the valley.
- **c.** Recommend a variety(ies) for the valley.

Design:	
Response Variable:	
Experimental Unit:	

Class Variable	Block or Treatment	Number of Levels	Fixed or Random	Description
1 2 ↓ n				

Proc GLM;
    Class Location Block Variety;
    Model Yield = Location Block(Location) Variety Location*Variety;
    Random Location Block(Location) Location*Variety / test;

Source	Type III Expected Mean Square	F
Location	$\sigma^2_{Error} + b\sigma^2_{LOC*VAR} + v\sigma^2_{BLOCK(LOC)} + bv\sigma^2_{LOC}$	$(MS_L + MS_E)/(\mathbf{MS_{BL(L)} + MS_{L*V}})$
Block(Location)	$\sigma^2_{Error} + v\sigma^2_{BLOCK(LOC)}$	$\mathbf{MS_{BL(L)}} / MS_E$
Variety	$\sigma^2_{Error} + b\sigma^2_{LOC*VAR} + Q(var)$	$MS_V / \mathbf{MS_{L*V}}$
Location*Variety	$\sigma^2_{Error} + b\sigma^2_{LOC*VAR}$	MS_{L*V} / MS_E

$b=4$ $v=5$

Proc VarComp Method = Type1;
    Class Location Block Variety;
    Model Y = Variety Location Block(Location) Location*Variety / Fixed = 1;
FIXED=n: specifies that the first *n* effects in the MODEL statement are fixed effects. The remaining effects are assumed to be random. By default, PROC VARCOMP assumes that all effects are random in the model.

```
Variance Component            Estimate

σ² (Location)                  60.45326
σ² (Location*Variety)          12.52485
σ² (Block(Location))           15.16845
σ² (Error)                    150.87298
```

Three-way factorials with one split (NOT a split-split-plot)

Example 1: RCBD with…   Main plot = Factorial treatment structure (A*B)
Subplot = Factor C

	a2-b1	a1-b2	a1-b1	a2-b2
BLOCK 1	c3	c1	c2	c2
	c2	c2	c1	c3
	c1	c3	c3	c1

	a1-b1	a2-b2	a2-b1	a1-b2
BLOCK 2	c1	c2	c1	c3
	c3	c1	c2	c2
	c2	c3	c3	c1

```
Proc GLM;
      Class Block A B C;
      Model Y = Block A B A*B Block*A*B
              C A*C B*C A*B*C;
      Test h = A        e = Block*A*B;
      Test h = B        e = Block*A*B;
      Test h = A*B      e = Block*A*B;
```

If A and B would have had three levels, for example:
```
      Means A / Tukey  e = Block*A*B;
      or
      Contrast 'B Linear' B 1 0 -1 / e = Block*A*B;
```

Alternative: open up the A*B factorial, expressing the combinations as levels of some single factor TRT:

```
Proc GLM;
      Class Block TRT C;
      Model Y = Block TRT Block*TRT
              C C*TRT;
      Test h = TRT  e = Block*TRT;
      Means TRT / Tukey  e = Block*TRT;
#                        A1B1 A1B2 A2B1 A2B2:
          Contrast 'B '    TRT  1 -1  1 -1 / e = Block*TRT;
          Contrast 'A '    TRT  1  1 -1 -1 / e = Block*TRT;
          Contrast 'A*B '  TRT  1 -1 -1  1 / e = Block*TRT;
```

If you do not have the A and B factors, then you need contrasts to separate the effects of A, B and A*B

Example 2: RCBD with…   Main plot = Factor A (2 levels)
Subplot = 2x3 factorial treatment structure (B*C)

	a2		a1	
	b1-c3	**b2-c1**	b2-c2	b1-c2
BLOCK 1	**b2-c2**	**b1-c2**	b2-c1	b1-c3
	b1-c1	**b2-c3**	b2-c3	b1-c1

	a1		a2	
	b1-c1	b1-c2	**b2-c1**	**b1-c3**
BLOCK 2	b2-c3	b2-c1	**b1-c2**	**b2-c2**
	b2-c2	b1-c3	**b2-c3**	**b1-c1**

Proc GLM;
   Class Block A B C;
   Model Y =Block A **Block*A**
        B C B*C A*B A*C A*B*C;
    Test h = A   e = **Block*A**;
   Means A / Tukey   e = **Block*A**;
   Contrast 'A Linear' A 1 0 -1 / e = **Block*A**; ← If A had three levels, for example

Alternatively, open up the B*C factorial, expressing the combinations as levels of some single factor TRT:

Proc GLM;
   Class Block A TRT;
   Model Y =Block A **Block*A**
        TRT A*TRT;   TRT= B C B*C and A*TRT= A*B A*C A*B*C
    Test h = A e = **Block*A**;

The model is simpler but then you need contrasts to solve:
B
C
B*C
A*B
A*C
A*B*C

Split-split-plot with last split in time

Example: In each of 3 randomly picked clinics used as block, select 8 diabetic patients from each sex and then randomly assign within each sex the 8 patients to treatments A, B, C and D (two patients per treatment). Glucose level in the blood of the patients is monitored at 12, 24 and 36 hours.

Level one
Clinic
SEX
Clinic*SEX= **Error A**

Level two
Trt
SEX*Trt
Clinic*Trt + Clinic*SEX*Trt= **Error B**

Level three: use conservative degrees of freedom: different F threshold (df/(3-1)
time
SEX*time
Trt*time
SEX*Trt*time
Clinic*time + Clinic*SEX*time + Clinic*Trt*time + Clinic*SEX*Trt*time= **Error C**

When you cannot randomly assign one treatment but instead you need to select first one category (sex, breed, etc.) and then within that class randomize the other treatment, this needs to be treated as a split in which the first category is the main plot.

Model Glucose=   Clinic Sex Clinic*Sex
                 Trt Sex*Trt Clinic*Sex*Trt
                 Time Sex*time Trt*Time Sex*Trt*Time;
        Test h=Sex e= Clinic*Sex;
        Test h=Trt e= Clinic*Sex*Trt;
        Test h=Sex*Trt e= Clinic*Sex*Trt;
    Means Trt / tukey e= Clinic*Sex*Trt;

SPLIT PLOT

CRD		RCBD		Latin Square	
		Blocks	r-1	Rows	a-1
				Columns	a-1
A	a-1	A	a-1	A	a-1
Error A	a(r-1)	Error A	(r-1)(a-1)	Error A	(a-1)(a-2)
Total	ra-1	Total	ra-1	Total	ra-1
Factor B	b-1	**Factor B**	b-1	**Factor B**	b-1
A x B	(a-1)(b-1)	**A x B**	(a-1)(b-1)	**A x B**	(a-1)(b-1)
Error B	a(r-1)(b-1)	**Error B**	a(r-1)(b-1)	**Error B**	a(r-1)(b-1)
Total	rab-1	**Total**	rab-1	**Total**	rab-1

CRD
```
proc glm;
class rep A B;
model y= A rep*A
       B A*B;
test h=A e=rep*A;
```

RCBD
```
proc glm;
class bl A B;
    model y=Bl A bl*A
          B A*B;
 test h=A e=bl*A;
```

LS
```
proc glm;
class row col A B;
 model y=col row A col*row*A
          B A*B;
  test h=A e= col*row*A;
```

Replicated LS (= rows & columns)
```
proc glm;
class sqr row col A B;
model y=sqr col row A sqr*col*row*A
       B A*B;
test h=A e= sqr*col*row*A;
```

STRIP PLOT

```
proc GLM;
  class block A B;
  model yield = block
                A A*block
                B B*block
                B*A;
  test h=A e=A*block;
  test h=B e=B*block;
```

SPLIT SPLIT PLOT

```
proc glm;
    class Block a b c;
    model response= Block a Block*a
                    b a*b Block*b*a
                    c a*c b*c a*b*c;
    test h=a    e= Block*a;
    test h=b    e= Block*b*a;
    test h=a*b e= Block*b*a;
```

```
proc glm;
   class block A B;
   model yield=block A block*A
             B block*B A*B;
   random block block*A block*B /Test;
```

Source	Type III Expected Mean Square
block	Var(Error) + 4 Var(block*B) + 4 Var(block*A)+ 8 Var(block)
A	**Var(Error) + 4 Var(block*A)** + Q(A,A*B)
block*A	**Var(Error) + 4 Var(block*A)**
B	**Var(Error) + 4 Var(block*B)** + Q(B,A*B)
block*B	**Var(Error) + 4 Var(block*B)**
A*B	Var(Error) + Q(A*B)

```
   test h=A e=block*A;
   test h=B e=block*B;
```

Split Plot EMS with Random statement

```
proc glm;
   class block A B;
   model yield=block A block*A
             B A*B;
   random block block*A /Test;
```

Source	Type III Expected Mean Square
block	Var(Error) + 4 Var(block*A) + 8 Var(block)
A	**Var(Error) + 4 Var(block*A)** + Q(A,A*B)
block*A	**Var(Error) + 4 Var(block*A)**
B	Var(Error) + Q(B,A*B)
A*B	Var(Error) + Q(A*B)

```
   test h=A e=block*A;
```

A 4x3x3 factorial experiment, organized as an RCBD (fixed model)

Proc GLM;
    Class Block A B C;
    Model Y = Block A|B|C;

Two 2-way interactions significant: If A*B and A*C are significant, but B*C is NS:

Proc Sort;
  By A;
Proc GLM;
  Class Block B C;
  Model Y = Block B|C;
    Means B / Tukey;
    Means C / Tukey;
    Contrast 'Linear B' B 1 0 -1;
  By A;
	Proc Sort;
	  By B C;
	Proc GLM;
	  Class Block A;
	  Model Y = Block A;
	    Means A / Tukey;
	    Contrast 'Linear A' A -3 1 1 1;
	  By B C;

One 2-way interaction significant: If A*B is significant, but A*C and B*C are NS:

It is valid to analyze the main effect of C:

  e.g. Contrast 'Linear C' C 1 0 -1;
        Means C / Tukey;

But for A and B:

  1. Describe the effect of A within each level of B (across all C levels)
  2. Describe the effect of B within each level of A (across all C levels)

Proc Sort;
  By B;
Proc GLM;
  Class Block A;
  Model Y = Block A;
  By B;
	Proc Sort;
	  By A;
	Proc GLM;
	  Class Block B;
	  Model Y = Block B;
	  By A;

If C is random include C and A*C in the model and use A*C as error term

---

If all three 2-way interactions are significant:

  1. Describe the effect of A within each combination of B and C
  2. Describe the effect of B within each combination of A and C
  3. Describe the effect of C within each combination of A and B

Two nested and 1 crossed factors

Objective: extend conclusions about fire vs. not fire in Central Valley Grasslands.

Variable: diversity of nematodes in the soil

2 types of Environments
A: Fires between 1950-1990
B: No Fires

4 random locations with previous Fire and 4 within no Fire.
1 2 3 4

2 blocks
1 2

2 treatments T: Grazed and Not Grazed
1 2

```
data ex99_3;
input Fire $ L block N Diversity@@;
cards;
YES 1 1 1 3.7   YES 1 1 2 3.9   YES 1 2 1 3.9   YES 1 2 2 3.7
YES 2 1 1 3.3   YES 2 1 2 3.1   YES 2 2 1 3.5   YES 2 2 2 3.7
YES 3 1 1 3.4   YES 3 1 2 3.2   YES 3 2 1 4.0   YES 3 2 2 4.3
YES 4 1 1 3.6   YES 4 1 2 3.5   YES 4 2 1 3.8   YES 4 2 2 3.5

NO 1 1 1 4.1   NO 1 1 2 4.4   NO 1 2 1 4.0   NO 1 2 2 4.2
NO 2 1 1 3.6   NO 2 1 2 3.7   NO 2 2 1 3.9   NO 2 2 2 3.4
NO 3 1 1 3.8   NO 3 1 2 3.9   NO 3 2 1 3.7   NO 3 2 2 3.1
NO 4 1 1 3.5   NO 4 1 2 3.2   NO 4 2 1 3.9   NO 4 2 2 3.8
;
Proc GLM;
class FIRE L T Block;
model Diversity= FIRE L(FIRE) T T*FIRE T*L(FIRE) Block(L FIRE);
   random L(FIRE) Block(L FIRE) T*L(FIRE);
   test h= T       e= T*L(FIRE);
   test h= T*FIRE  e= T*L(FIRE);
   test h= FIRE e= L(FIRE);
run; quit;
```

Source	Type III Expected Mean Square
Fire	Var(Error) + 2 Var(block(L Fire)) + 2 Var(T*L(Fire)) + 4 Var(L(Fire)) + Q(Fire)
L(Fire)	Var(Error) + 2 Var(block(L Fire)) + 2 Var(T*L(Fire)) + 4 Var(L(Fire))
T	Var(Error) + 2 Var(T*L(Fire)) + Q(T)
Fire*T	Var(Error) + 2 Var(T*L(Fire)) + Q(Fire*T)
T*L(Fire)	Var(Error) + 2 Var(T*L(Fire))
block(L Fire)	Var(Error) + 2 Var(block(L

RCBD with 1 observation
The error is automatically the **Trt*Block**
If Trt NS -> Tukey test to test multiplicative effects: if yes ➔ transform data

RCBD with more than 1 observation

	Trt1	Trt2	Trt3
Block 1	X X	X X	X X
Block 2	X X	X X	X X

To test the `block*Trt` we include it in the model

```
proc GLM;
  class block Trt ;
  model yield = block Trt block*Trt;
  random block Trt*block ;
run; quit;
```

Source	Type III Expected Mean Square
block	Var(Error) + 4 Var(block*Trt) + 12 Var(block)
Trt	Var(Error) + 4 Var(block*Trt) + Q(Trt)
block*Trt	Var(Error) + 4 Var(block*Trt)

A) If **Block*Trt** is NOT significant: **eliminate it from the model**

- This is equivalent to declare block and block*trt as random)

B) If **Block*Trt** interaction is significant: we have a problem.

- If it becomes NS by transformation, it was a multiplicative effect

- If the interaction is still significant indicates that the effect of the Trt is different in each block!

- It does not make much sense to describe the Trt effects by block (this is a **general situation for any random factor**).

- If you still want to make a conclusion for Trt that is valid across all blocks, the correct error term is the interaction. See the result of the random statement.

- So the advice will be the same: **eliminate Block*Trt from the model**

Remember: the significance of the interaction may be pointing to an interesting biological or ecological problem!

RCBD: Power Calculation

(ST&D 207) Percent oil in flax seed harvested from plants inoculated with rust disease at different stages.

| Treatment	Block 1	Block 2	Block 3	Block 4	Mean	τ_i^2
Seedling	34.4	35.9	36.0	34.1	35.100	0.188
Early bloom	33.3	31.9	34.9	37.1	34.300	1.521
Late bloom	34.4	34.0	34.5	33.1	34.000	2.351
Full bloom	36.8	36.6	37.0	36.4	36.700	1.361
Ripening	36.3	34.9	35.9	37.2	36.075	0.293
Control	36.4	37.3	37.7	36.7	37.025	2.225
Mean	35.267	35.100	36.000	35.767	35.533	$\Sigma = 7.940$
β_i^2	0.071	0.188	0.218	0.054	$\Sigma = 0.531$	

Source	df	SS	MS	F
Treatment	5	31.758	6.352	4.790
Block	3	3.187	1.062	0.801
Error	15	19.888	1.326	

To calculate the power of an RCBD, use Pearson and Hartley's power function charts (1953, Biometrika 38:112-130). To begin, calculate ϕ:

$$\phi = \sqrt{\frac{r}{MSE}\sum\frac{\tau_i^2}{t}} = \sqrt{\frac{4}{1.326}\left(\frac{7.940}{6}\right)} = 2.00$$

Looking in the power charts for $df_t = 5$, $df_e = 15$, and $\alpha = 0.05$, we find a power of $\sim 90\%$.

Power and sample size in analysis of variance: $\nu_1 = 5$.
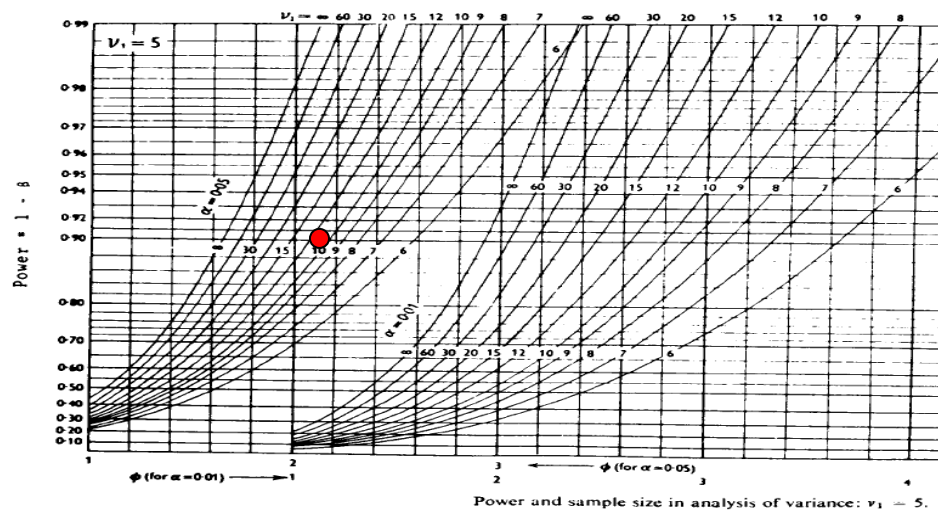
Other models: In a split plot use the correct MSE depending on the question

Power analysis in SAS is also straightforward:

```
Data Flax;
Do Trtmt = 1 to 6;
     Do Block = 1 to 4;
           Input Oil @@;
           Output;
     End;
End;
Cards;
34.4 35.9 36.0 34.1
33.3 31.9 34.9 37.1
34.4 34.0 34.5 33.1
36.8 36.6 37.0 36.4
36.3 34.9 35.9 37.2
36.4 37.3 37.7 36.7
;
Proc GLM Data = Flax;
     Class Trtmt Block;
     Model Oil = Trtmt Block;

Proc GLMPower Data = Flax;
     Class Trtmt Block;
     Model Oil = Trtmt Block;
     Power
        Stddev = 1.151472     ← The Root MSE from ANOVA table
        Alpha  = 0.05         ← Specify the alpha
        ntotal = 24           ← Specify n, the No. of EUs in the exp.
        Power  = .;           ← A period here tells SAS to calculate  power
Run;
Quit;
```

The output:

Fixed Scenario Elements

Dependent Variable	Oil
Alpha	0.05
Error Standard Deviation	1.151472
Total Sample Size	24
Error Degrees of Freedom	15

Computed Power

Index	Source	Test DF	Power
1	Trtmt	5	0.904