

Topic 6. Two-way designs: Randomized Complete Block Design

[ST&D Chapter 9 sections 9.1 to 9.7 (except 9.6) and Chapter 15: section 15.8]

6. 1. Variability in the completely randomized design (CRD)

In the CRD it is assumed that the experimental units are uniform. This is not always true in practice and it is necessary to develop methods to deal with variability. If in comparing two methods of fertilization one region of the field has much greater fertility than the others, then a treatment effect might be incorrectly ascribed to the treatment applied to this part of the field, making a Type I error. For this reason in CRD it is always advocated to include as much of the native variability of the experiment as possible *within* each plot, making each plot as representative of the whole experiment, and the whole experiment as uniform, as possible. In actual field studies plots are designed long and narrow to achieve this effect. However, if the plots are more variable, experimental error (MSE) is larger, F (MST/MSE) is smaller, and the experiment is less sensitive. Finally, if the experiment is replicated in a variety of situations to increase the scope of the experiment, this additional variability needs to be removed from the analysis to focus on the treatment effect. This is the purpose of blocking.

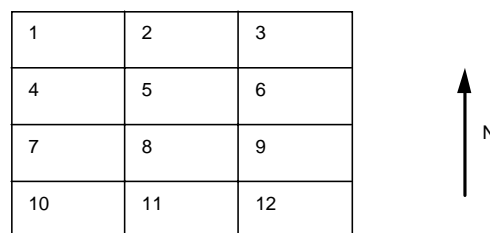
6. 2. Randomized complete block design (RCBD)

6. 2. 1. Definition

The RCBD assumes that a population of experimental units can be divided into a number of relatively homogeneous subpopulations or *blocks*. The treatments are then randomly assigned to experimental units such that each treatment occurs equally often (usually once) in each block –each block contains all treatments. Blocks usually represent naturally occurring differences not related to the treatments. In the analysis the variation among blocks can be partitioned out, usually reducing the experimental error (MSE).

6. 2. 2. Example

Consider a field trial comparing three cultivars with four replications. Suppose the native N level of the soil in the field site varies from high at the north end to low at the south end. Yield is expected to vary from one end to another, whether or not the difference is due to cultivar differences. This violates the assumption that the error terms are identically distributed since the error terms will tend to be negative at one end of the field and positive at the other.



The field can be divided into four blocks of three plots each, and the soil in each of these blocks will be uniform. This is the basic idea of the *randomized complete block* design. In the *completely randomized design*, each plot has an equal chance of being chosen for each treatment. In the *randomized complete block design*, within each block

every plot has the same chance of being chosen for each treatment, but within each block a fixed number (often 1) of plots will be chosen for each treatment. The term *complete* refers to the fact that each block gets every treatment.

Block			
1	1 (B)	2 (A)	3 (C)
2	4 (A)	5 (B)	6 (C)
3	7 (A)	8 (C)	9 (B)
4	10 (A)	11 (C)	12 (B)

↑
N

6. 2. 3. Statistical model

The new model is $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$. Here τ_i represents the average value associated with treatment i , $i = 1, \dots, t$; and β_j represents the error associated with block j , $j = 1, \dots, r$ (this assumes one treatment per block). The data represent the model as:

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})$$

The sum of squares equation becomes:

$$\sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_{..})^2 = r \sum_{i=1}^t (\bar{Y}_{i.} - \bar{Y}_{..})^2 + t \sum_{j=1}^r (\bar{Y}_{.j} - \bar{Y}_{..})^2 + \sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$$

$$\text{or, } SS = SST + SSB + SSE.$$

Since the variance of means of n observations is σ^2/n , multipliers r and t shown above in the SSB and SST result in all mean squares being estimates of the same σ^2 when there are no block or treatment effects. This is another example of *partitioning* of variance. This partitioning is possible because the sums of squares of blocks and treatments are *orthogonal*.

6. 2. 4. ANOVA

ANOVA table for the **RCBD**

Source	df	SS	MS	F
Blocks	$r - 1$	SSB	$SSB/(r-1)$	
Treatments	$t - 1$	SST	$SST/(t-1)$	MST/MSE
Error	$(r-1)(t-1)$	$SS-SST-SSB$	$SSE/(r-1)(t-1)$	
Total	$rt - 1$	SS		

ANOVA table for the **CRD**

Source	df	SS	MS	F
Treatments	$t - 1$	SST	$SST/(t-1)$	MST/MSE
Error	$t(r - 1)$	$SS - SST$	$SSE/r(t-1)$	
Total	$rt - 1$	SS		

Notice that there are fewer degrees of freedom for error in the RCBD design than in the CRD design, $(r-1)(t-1)$ vs. $t(r-1)$, or $(r-1)$ fewer degrees of freedom. In the RCBD, these $r-1$ degrees of freedom have been partitioned from the error and assigned to the blocks.

No difference among blocks: if the RCBD design were applied to an experiment in which the blocks were really no different (i.e. no significant block effect), the MSE for the CRD would be smaller than the MSE for the RCBD simply due to degrees of freedom. For example, if $t=3$ and $r=4$, $MSE_{CRD} = SSE/9$, and $MSE_{RCBD} = SSE/6$. Therefore, the F statistic for the CRD would be larger.

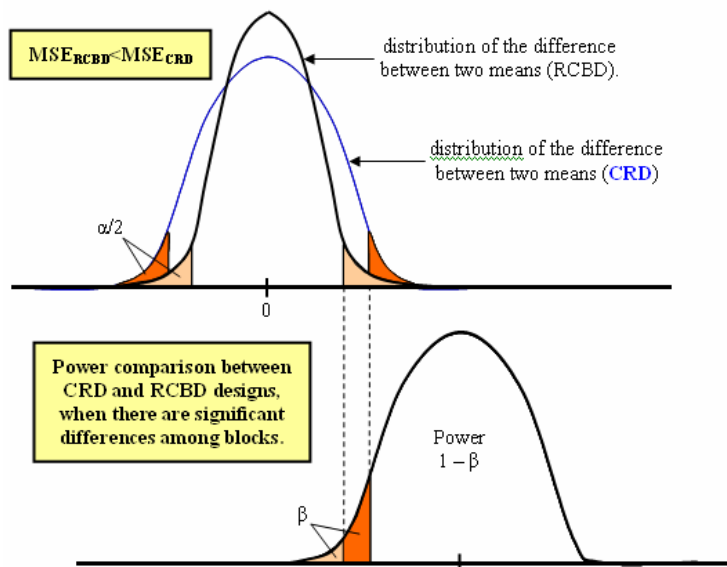
Consider a confidence interval for the differences between two means

$$\bar{Y}_A - \bar{Y}_B \pm \sqrt{\text{Critical}F_{(1, MSE df), \alpha} * MSE \frac{2}{r}}$$

Under $H_0 = \bar{Y}_A - \bar{Y}_B = 0$

The CRD has a smaller critical F value than the RCBD because of its larger df. In addition if there are no differences among blocks then $MSE_{CRD} = MSE_{RCBD}$. Therefore, the larger critical F value in the RCBD moves the threshold of the rejection further from the mean (0) than in the CRD. This change in the position of the rejection threshold affects the Type II error (β) and the power of the test ($1-\beta$). Under this scenario, the probability of accepting a false null hypothesis (β) will be smaller in the CRD than in the RCBD. In other words, the CRD would in this situation be more powerful (larger $1-\beta$).

Significant difference among blocks: On the other hand, suppose that there really were a substantial difference among blocks as well as among the treatments (H_0 false). If the CRD were used, this difference among blocks would be allocated to the error, so the F statistic for the CRD would be smaller than the F statistic of the RCBD.



Under this scenario, the RCBD would still have a larger critical F value because of the lost degrees of freedom, but this may be more than compensated by the smaller MSE. If the effect of the reduced MSE (threshold closer to 0) is larger than effect of the larger critical value (threshold further from 0) the net result will be a smaller β , and a larger power ($1-\beta$) in the RCBD relative to the CRD.

Obviously one should only use the RCBD when the variation explained by the blocks more than offsets the degrees of freedom they consume. So how can one determined when an RCBD is appropriate?

The concept of *efficiency*, discussed in section 6. 3., answers this.

6. 2. 5. Example

The example is from Little and Hills, and involves the response of sheep to estrogen. The sheep are blocked by ranch, with four treatments per block. The treatments are combinations of sex of the sheep (M or F) and level of estrogen treatment (S0 or S3). Although these data could be analyzed as a factorial experiment, in this example they are treated as four separate treatments.

Table 6.1 RCBD. Effect of estrogen on weight gains. Blocks are 4 different ranches.

Treatment	Block				Treatment	
	I	II	III	IV	Total	Mean
F-S0	47	52	62	51	212	53
M-S0	50	54	67	57	228	57
F-S3	57	53	69	57	236	59
M-S3	54	65	74	59	252	63
Block Total	208	224	272	224	928	
Block Mean	52	56	68	56		58

Table 6.2 RCBD ANOVA

Source of Variation	df	SS	MS	F
Totals	15	854		
Blocks	3	576	192.00	24.69**
Treatments	3	208	69.33	8.91**
Error	9	70	7.78	

Table 6.3 CRD ANOVA

Source of Variation	df	SS	MS	F
Totals	15	854		
Treatments	3	208	69.33	1.29 NS
Error	12	646	53.83	

Since each treatment occurs the same number of times in each block, differences among blocks do not result from treatments but from other differences associated with the blocks. This component of the total sum of squares can be removed and the experimental error reduced accordingly. Compare the SS_{error} in Tables 6.2 and 6.3

6. 2. 5. 2. SAS Program

There are now two-classification variables block and sex_est. The response variable is gain. There are two independent effects: block and treatment (sex_est). SAS

does not know that the scientific interpretation is different for these two effects - it simply computes an F statistic for both of them.

SAS PROGRAM

```

data lambs;
  input sex_est $ @;
  do block = 1 to 4;
    input gain @;
    output;
  end;
cards;
f0 47 52 62 51
m0 50 54 67 57
f3 57 53 69 57
m3 54 65 74 59
;
proc glm;
  class block sex_est;
  model gain=block sex_est;

run; quit;

```

6. 3. Relative efficiency [ST&D p. 221, and Topic 1 section 1.4.4.6]

We saw earlier that if the variation among blocks is large then we can expect the RCBD method to work better than the CRD while if this variation is small it may not. The concept of *relative efficiency* formalizes the comparison between two experimental methods. Recall that the F statistic is defined by the formula $F = MST/MSE$. The experimental design affects primarily the MSE since the degrees of freedom for treatments is always $t - 1$. The information in the design is $1/MSE$, so the relative efficiency of design to design to is $(1/MSE_1)/(1/MSE_2) = MSE_2/MSE_1$. When the degrees of freedom of the mean squares from which the relative efficiency is calculated are less than 20 a correction factor is used (Cochran and Cox, 1957). The following formula includes the correction factor and gives an estimate of the relative amount of information provided by two designs:

$$RE_{1 \text{ to } 2} = \frac{(n_1+1)/[(n_1+3)MSE_1]}{(n_2+1)/[(n_2+3)MSE_2]} = \frac{(n_1+1)(n_2+3)MSE_2}{(n_2+1)(n_1+3)MSE_1}$$

where n_i is the total degrees freedom associated with error for design i . If this ratio is >1 , design 1 provides more information and is more efficient than design 2. If $RE_{1 \text{ to } 2} = 2.0$, for example, then each rep of design 1 gives as much information as two reps of design 2.

The problem is to estimate MSE for the alternative design. This MSE can be estimated, however, by scaling the MSE of partitioned sums of squares by their degrees of freedom. For example, in comparing the CRD to the RCBD when the experiment has been run as an RCBD (ST&D p.222), we can use the formula

$$MSE_{CR} \cong \frac{f_b MSB_{RCB} + (f_t + f_e) MSE_{RCB}}{f_b + f_t + f_e}$$

where MSB and MSE are the block and error mean squares, and f_b , f_t , and f_e are the block, treatment, and error degrees of freedom. To obtain this formula the total SS of the two designs are assumed equal, and then the MS are replaced by the variance components of the expected MS. This concept will be studied later in the course (for a complete derivation of this equation see Sokal & Rohlf 1995, Biometry p.838-839)

From the sheep experiment, we have $MSE_{RCBD} = 7.78$, and $MSB_{RCBD} = 192.0$. Therefore,

$$MSE_{CRD} \cong \frac{3 * 192.0 + (3 + 9) 7.78}{3 + 3 + 9} = 44.62$$

$$RE_{RCBD \text{ to } CRD} = \frac{(f_{rcbd} + 1)(f_{crd} + 3)MSE_{crd}}{(f_{crd} + 1)(f_{RCBD} + 3)MSE_{RCBD}} = \frac{(9 + 1)(12 + 3)44.62}{(12 + 1)(9 + 3)7.78} = 5.51$$

which says that it takes 5.51 replications of the CRD to produce the same amount of information as one replication of the RCBD.

6. 4. Assumptions of the model

The model for the RCBD is $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$. As in the CRD it is assumed that the ε_{ij} are independent, homogeneous and normally distributed. In two-way or higher-order ANOVA **without replication** it is necessary to assume that interaction is not present if one is to make tests of the main effect using the MSE. This assumption of no interaction in a two-way ANOVA is sometimes also referred to as the assumption of **additivity** of the main effects. When treatments and block effects are additive there is no significant $\tau_i * \beta_j$ effect in the model. If interaction is present, the F-test will be very inefficient and possibly misleading if the effect of the interaction is very large. An interaction term will result if the effect of the two factors A and B on the response variable Y is multiplicative rather than additive. An example will make this clear

Table 6.4. Additive and multiplicative effects

	Factor A			
Factor B	$\tau_1 = +1$	$\tau_2 = +2$	$\tau_3 = +3$	
$\beta_1 = +1$	2	3	4	Additive effects
	1	2	3	Multiplicative effects
	0	0.30	0.48	Log of multiplicative effects
$\beta_2 = +5$	6	7	8	Additive effects
	5	10	15	Multiplicative effects
	0.70	1.00	1.18	Log of multiplicative effects

In Table 6.4 additive and multiplicative treatment effects are shown in a hypothetical two-way ANOVA. Let us assume that the population mean is $\mu=0$. Then the mean of the sample subjected to treatment 1 of factor A and treatment one of factor B should be 2 by the conventional additive model. Similarly, the expected subgroup mean subjected to level 3 of factor A and level 2 of factor B is 8, since the respective contributions to the mean are 3 and 5. If the process is multiplicative rather than additive, however, as occurs in a variety of physicochemical and biological phenomena, the expected values are quite different. For treatment A_3B_2 , the expected value is 15, the product of 3 and 5.

If multiplicative data of this sort is analyzed by a conventional ANOVA, the interaction SS is greatly augmented because of the nonadditivity of the treatment effects. In this case there is a simple remedy. By transforming the variable into logarithms the additivity of the data is restored. The third line in each cell gives the logarithm of the expected value, assuming multiplicative relations. After the transformation the increments are strictly additive again ($\tau_1=0$, $\tau_2=0.30$, $\tau_3=0.48$, $\beta_1=0$, $\beta_2=0.70$). This is a good illustration of how transformations of scale can be used to meet the assumptions of analysis of variance.

6. 4. 1. Tukey's test for nonadditivity

A test devised by Tukey (ST&D p395) can be used to test whether the interaction found in a given set of data can be explained in terms of non-additive main effects. Tukey's test partitions the interaction sum of squares into one degree of freedom due to non-additive effects of the main effects and a residual sum of squares to represent the other possible interactions that serve as error in ANOVA with no replications within cells. If the single-degree-of-freedom SS for nonadditivity is not significant the assumptions for the ANOVA model are satisfied. If the test is significant, then an analysis of the log of the variable may be more appropriate.

Tukey's method is shown and explained below.

$$SS \text{ (nonadditivity)} = (\sum_i \sum_j \tau_i \beta_j y_{ij})^2 / (\sum_i \tau_i^2) (\sum_j \beta_j^2)$$

$$SS \text{ (nonadditivity)} = Q^2 / \text{Cross Product}, \quad df = 1, \quad \text{where}$$

$$Q^2 = \left\{ \sum_i (\bar{y}_{i.} - \bar{y}_{..}) (\bar{y}_{.j} - \bar{y}_{..}) y_{ij} \right\}^2$$

$$\text{Cross Product} = \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2$$

These calculations can be implemented using SAS (ST&D p397) and will be discussed in the laboratory class.

SAS Program

```

Data lambs;
  Input sex_est $ block gain @@;
Cards;
xx   xx   xx
;
Proc GLM Data =lambs;
  Class block sex_est;
  Model gain= block sex_est;
  Output out= lambs1 p= pgain r= resigain;

Proc GLM Data = lambs1;
  Class block sex_est;
  Model gain= block sex_est pgain*pgain;
run; quit;

```

Output

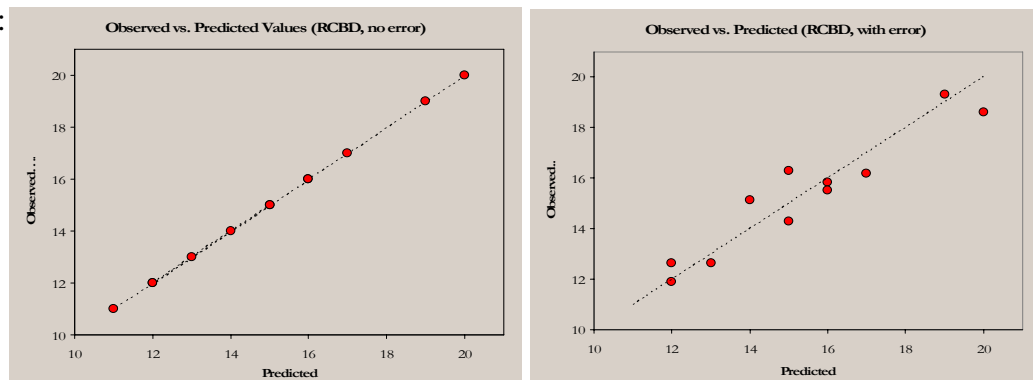
Source	DF	SS	MS	F	Pr > F
block	3	0.74	0.25	0.03	0.9927
sex_est	3	0.30	0.10	0.01	0.9981
pgain*pgain	1	3.42	3.42	0.41	0.5395 NS

The Tukey test is NS and therefore we do not reject the null hypothesis of additivity and we can use the interaction as a correct estimate of the error term. This test is necessary **ONLY** when there is **one observation** per block / treatment combination. If there are more replications, the block*tratement interaction can be added to the Model, and the variation among the e.u. within each block / treatment combination will be used as error term.

How does a regression of the observed data against the squares of its predicted values tell you anything about the existence of nonadditive effects?

Under the linear model, each observation is characterized as: $y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$

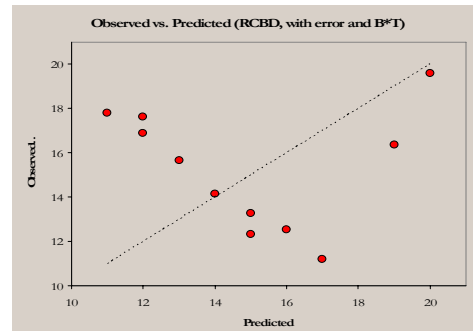
Therefore:



If the errors in the experiment are in fact random and independent then ε_{ij} will be a *random variable* that causes *no systematic deviation* from this linear relationship.

If there is an interaction $\varepsilon_{ij} = \varepsilon_{RANDOMij} + \mathbf{B*T Interaction Effects}$

The following plot illustrates the deviation from linearity that results from significant multiplicative effects. If the quadratic effect **pred*pred** is significant, it indicates a significant departure from the linearity expected from the additive effects.



6. 4. 2. Diagnostic checking of the Model

Table 6.5. Yield of penicillin in four different treatments A, B, C, D. Blocks are different stocks of an important raw material. The numbers below each observation are the predicted (P: Grand Mean + Treatment effect + Block effect) and residual (R) values.

Discrepancies of many different kinds between the tentative model and the data can be detected by studying residuals. These residuals, the third value in each cell, are the quantities remaining after the systematic contributions associated with the assumed model (in this case treatments and blocks) are removed (See ST&D p.213-214).

Block	Treatment				Block Mean	Block Effect
	A	B	C	D		
Stock 1	O: 89 P: 90 R: -1	O: 88 P: 91 R: -3	O: 97 P: 95 R: 2	O: 94 P: 92 R: 2	92	+6
Stock 2	O: 84 P: 81 R: 3	O: 77 P: 82 R: -5	O: 92 P: 86 R: 6	O: 79 P: 83 R: -4		
Stock 3	O: 81 P: 83 R: -2	O: 87 P: 84 R: 3	O: 87 P: 88 R: -1	O: 85 P: 85 R: 0	85	-1
Stock 4	O: 87 P: 86 R: 1	O: 92 P: 87 R: 5	O: 89 P: 91 R: -2	O: 84 P: 88 R: -4		
Stock 5	O: 79 P: 80 R: -1	O: 81 P: 81 R: 0	O: 80 P: 85 R: -5	O: 88 P: 82 R: 6	82	-4
Treatment mean	84	85	89	86		
Treatment effect	-2	-1	3	0		

SAS program

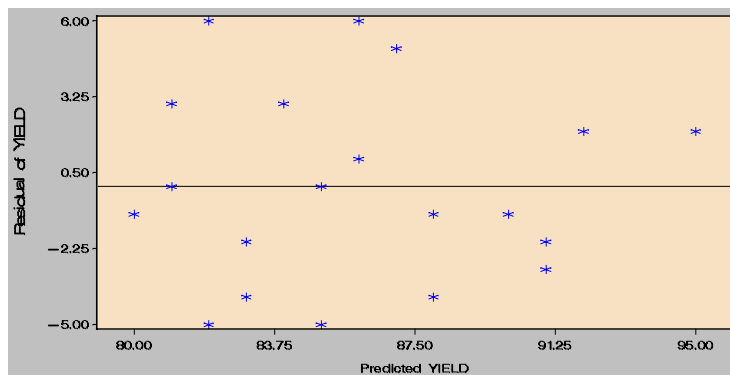
The OUT option after MODEL creates a new data set, named 'resplot' in this example, that includes the original variables plus the predicted (PREDYIEL) and residual (RESIYIEL) variables.

SAS program

```
data penicil;
  input block trtmnt yield;
cards;
1 1 89
. . .
5 4 88
;
proc GLM;
  class block trtmnt;
  model yield=block trtmnt;
  output out=resplot p=predyiel r=resiyiel;

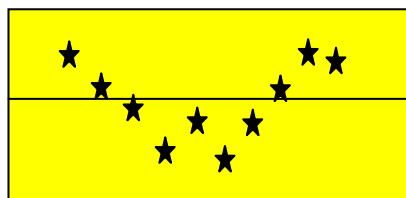
proc plot data=resplot;
  plot resiyiel*predyiel=trtmnt;

proc univariate data=resplot normal;
  var resiyiel;
run; quit;
```

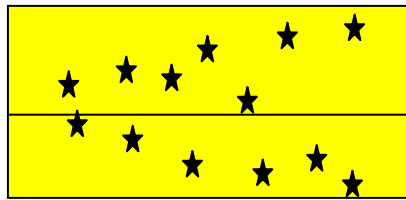


In this example no particular pattern is observed in the residuals. This observation parallels the normal distribution of the residuals, the homogeneous variances of the residuals among treatments and a non-significant Tukey's test for nonadditivity.

However, sometimes the plot of the residuals versus the predicted values shows a curvilinear relationship with positive residuals for low and high values of Y and negative residuals for intermediate values.



This appearance suggests nonadditivity between the block and the treatment effects that might be eliminated by a suitable transformation of the response. Other times the plot of the residuals versus the predicted shows a funnel-like appearance.



This indicates that the variance increases as the value of the response increases (a situation that is common when the variance is a constant percentage of the mean). Other useful plot can be the plot of the residuals versus the sequence in which the data was obtained. This can show some systematic variation during the collection of the data.

6.5 Example of a nested design within a RCBD

The next data set was created assuming two subsamples within each experimental unit. Values were created in a way that their average is the previous value in the lamb dataset

```

data lambs;
input sex_est $ block animal gain @@;
cards;
f0 1 1 46    f0 2 1 51    f0 3 1 61    f0 4 1 50
m0 1 1 49    m0 2 1 53    m0 3 1 66    m0 4 1 56
f3 1 1 56    f3 2 1 52    f3 3 1 68    f3 4 1 56
m3 1 1 53    m3 2 1 64    m3 3 1 73    m3 4 1 58

f0 1 1 48    f0 2 1 53    f0 3 1 62    f0 4 1 52
m0 1 1 51    m0 2 1 55    m0 3 1 68    m0 4 1 58
f3 1 1 58    f3 2 1 54    f3 3 1 70    f3 4 1 58
m3 1 1 55    m3 2 1 66    m3 3 1 75    m3 4 1 60

proc glm data=lambs order=data;
*With order=data, SAS reads f0 m0 f3 m3: Contrast are different!;
class block sex_est animal;
model gain= block sex_est animal(block*sex_est);
random animal(block*sex_est);
test h=sex_est e=animal(block*sex_est);

contrast 'sex'          sex_est 1 -1 1 -1 / e=animal(block*sex_est);
contrast 'estrogen'     sex_est 1 1 -1 -1 / e=animal(block*sex_est);
contrast 'interaction'  sex_est 1 -1 -1 1 / e=animal(block*sex_est);

means sex_est/tukey e=animal(block*sex_est);

*In nested models specify the correct error term in all mean comparisons;
*Next is Tukey test with incorrect error for comparison;

means sex_est/tukey;

proc varcomp Method= Type1;
class block sex_est animal;
model gain= block sex_est animal(block*sex_est);

run; quit;

```

To indicate that there is a subsample within each Treatment x block combination a nested factor is included `animal(block*sex_est)` and is defined as RANDOM. It is

important to note that the error term to test differences between treatment means (`animal(block*sex_est)`) should be defined for each hypothesis tested (e.g. $h = \text{sex_est}$) but also for each contrast or mean comparison. If the error term is not specified SAS will test every hypothesis and every contrast using the subsampling error.

Output LAMB NESTED

ANOVA Dependent Variable: gain

Source	DF	SS	MS	F Value	Pr > F
Model	15	1700.5	113.4	59.47	<.0001
Error	16	30.5	1.9		
Corrected Total	31	1731.0			

Source	DF	SS	MS	F Value	Pr > F
block	3	1132.1	377.4	198.0	<.0001
sex_est	3	426.1	142.0	74.5	<.0001
animal(block*sex_est)	9	142.3	14.8	8.3	0.0002

Tests of Hypotheses Using the Type III MS for animal(block*sex_est) as an Error Term

Source	DF	SS	MS	F Value	Pr > F
sex_est	3	426.1	142.0	8.98	0.0045

Contrast	DF	SS	MS	F Value	Pr > F
sex	1	132.0	132.0	8.35	0.0179
estrogen	1	294.0	294.0	18.60	0.0020
interaction	1	0.03	0.03	0.00	0.9655

Tukey's Studentized Range (HSD) Test for gain

	Correct	Incorrect
Minimum Significant Diff.	6.2	2.0

Correct Grouping	Incorrect Grouping	Mean	N	sex_est
	A	63.000	8	m3
B	A	59.000	8	f3
B	A	57.000	8	m0
B	D	52.875	8	f0

Variance Components

Source	Expected Mean Square
block	$\text{Var}(\text{Error}) + 2 \text{Var}(\text{animal}(\text{block*sex_est})) + 8 \text{var}(\text{block})$
sex_est	$\text{Var}(\text{Error}) + 2 \text{Var}(\text{animal}(\text{block*sex_est})) + 8 \text{var}(\text{sex_est})$
animal(block*sex_est)	$\text{Var}(\text{Error}) + 2 \text{Var}(\text{animal}(\text{block*sex_est}))$
Error	$\text{Var}(\text{Error})$

Variance Component	Estimate	%
Var(block)	45.3	64.7
Var(sex_est)	15.8	22.6
Var(animal(block*sex_est))	7.0	10.0
Var(Error)	1.9	2.7

The calculation of the variance comp. is the objective of the nested design

This information about experimental unit variation and subsample variation can be used together with cost information to calculate the optimum allocation of resources to subsamples and experimental units.