



---

**SPLIT-PLOT DESIGN: ANALYSIS DOLDRUMS COMPARING GENSTAT, SPSS AND R**

Hypolite Bayor<sup>1</sup>, George Nyarko<sup>1\*</sup> and Damian Tom-Dery<sup>2</sup>

<sup>1</sup> Department of Horticulture, Faculty of Agriculture, University for Development Studies  
Box TL 1882, Tamale, Ghana

<sup>2</sup> Department of Forestry and Forest Resources, Faculty of Renewable Natural Resources,  
University for Development Studies, Box TL 1882, Tamale, Ghana

Corresponding author email: [nyarko2uds@yahoo.com](mailto:nyarko2uds@yahoo.com)/[gnyarko@uds.edu.gh](mailto:gnyarko@uds.edu.gh)

---

**Abstract**

A modelled split plot experiment involving Completely Randomized Design of the main (whole) plots replicated three times was analysed using three statistical software packages namely Genstat, SPSS and R. The results were compared in terms of accuracy of output and ease of analysis. The data set was also analysed semi-manually to validate the results of the software outputs. Genstat and R produced identical result to the semi-manual analysis indicating they were correct. SPSS however, produced erroneous test results even when the correct linear model was specified. The correct output could however be obtained by specifically instructing SPSS on how to carry out the test. This experiment demonstrates the need for experimenters to be in charge of every aspect of an experiment from design to analysis leaving no part to the whims of statistical software. It also demonstrates that SPSS on its own is incapable of producing the correct result based on the linear model for this experiment. SPSS and Genstat are menu-driven and may be easier to learn than R which is mostly command based.

---

**INTRODUCTION**

Split plot designs are common field plot techniques in agriculture and other research areas. A Split plot is considered as a “generalization of Randomized Complete Block designs (RCBD)” (Montgomery, 1991) to accommodate treatment combinations that are too many to be accommodated in conventional RCBDs (Gomez & Gomez, 1984) or a special case of incomplete block design involving several sizes of experimental units (Milliken & Johnson, 1996). One

implementation of Split-plot (for a two factor experiment) involves creating a number of blocks which are divided into plots corresponding to the number of treatments in the main plot treatment (Factor A); then subdividing each experimental unit of Factor A into sub-plots corresponding to the number of treatments in the subplot treatment (Factor B). Treatments of Factor A are randomized at the level of the blocks (one block at a time) such that each main (whole) plot is allocated

only one treatment of Factor A. Factor B is also randomized but at the level of the whole plots (one whole plot at a time) such that one sub-plot in each whole plot is allocated one of the treatments of Factor B. With respect to Factor B, each experimental unit of Factor A becomes a block. A third or fourth factor could be added using the same procedure and further subdividing the subplots. At the end, the experiment consists of a hierarchical series of RCBDs, each experimental unit serving as a block for the sub-plots immediately below it. The main plots serve as “sub-plots” for the topmost blocks.

However, this is not the only way to implement split-plots. According to Steel & Torrie (1980) split-plot designs “incorporate one or more of the completely randomized, randomized complete block or Latin square designs” implying a split-plot could start with a CRD rather than an RCBD. Split-plots are notorious for lacking sufficient degrees of freedom at the main plot level and this has led some writers to suggest that “precision of the main plots are sacrificed for the sub-plots” (Gomez & Gomez, 1984). This does not always have to be the case. In fact better sensitivity can be achieved if we use CRD rather than RCBD at the main plot level due to gain in degrees of freedom for testing the main plots and that improves the sensitivity of the experiment.

When the main plot is CRD, there is a mis-match between the main plot design and the sub-plot design. While the main-plot design is CRD, the sub-plot design is an RCBD with respect to the experimental units of the main plots. This can become a source of confusion especially for some statistics software programmes and may lead to erroneous result. This therefore calls for clarity of thinking on the part of the experimenter and due process should be followed.

GENSTAT is a general statistical package suitable for agricultural field experiments. It has both commercial and free versions for some countries. It is noted

for its rigour in statistical approach. SPSS is a general statistical package but geared towards the social sciences. It is widely used for analysis of agricultural field experiments partly because of its ease of use and wide availability. R is free software and provided without guarantee. It is rigorous in its approach but is limited in the provision of menu-driven graphical interphase. It is therefore largely command base which makes it difficult to grasp initially.

The aim of this paper is therefore to compare these three software programmes in terms of accuracy in analysing a simple split-plot design in which the main plots are completely randomized. Accuracy of result is the most important criterion although ease of use (especially menu-driven interphase) is also a criterion.

## MATERIALS AND METHODS

### The experiment

Let us assume that a mango farmer is interested in investigating the effects of two land preparation methods (burning and non-burning) and three fertilizers (nitrogen, phosphate and NPK (15-15-15)) on the yield of mango trees. He wishes to replicate each treatment combination three times. Implementing this as a CRD 2X3X3 factorial would involve randomly selecting one of the 18 burning-fertilizer treatment combinations, implementing it and then go to the next. This would involve burning 9 small patches (in some random order) each representing one experimental unit. The unit to be burnt might be a small area for planting one tree or a few trees planted in a row. This is obviously not the most desirable thing.

On the other hand, having just two areas, one burnt the other not burnt will result in so severe a restriction that there would be no “pure” error to test for burning. It might be expedient to divide the experimental plot into sizes that are reasonably large for proper burning to take place. The two land preparation methods (burn and no burn) would then be randomly assigned to the plots. Let us assume the area

is divided into six plots, three plots would randomly be assigned to the burning and therefore burnt and the other three left unburnt. Each of the six large plots would be divided into three sub-plots and the fertilizer would then be randomized on these plots. This arrangement leads to a split-plot design. The main plot is a completely randomized experiment with three replicates. One realization of this arrangement may be as shown in figure 1.

### Linear model and expected mean squares

One way of specifying the linear model for the analysis of this experiment is as follows;

$$Y_{ijk} = \mu + B_i + \omega_{(ij)} + \partial_{(ij)} + F_k + BF_{ik} + \omega F_{(ijk)} + \epsilon_{(ijk)} \dots \dots \dots \text{Eqn 1}$$

$$i=1, 2; \quad j=1, 2, 3; \quad k=1, 2, 3.$$

Where

$Y_{ijk}$  = the yield (quantity) of fruits in a sub-plot belonging to the  $k^{\text{th}}$  fertilizer treatment (fixed) in one of the three ( $j^{\text{th}}$ ) replicates of the  $i^{\text{th}}$  burning (fixed) treatment;

$\mu$  = the overall mean

$B_i$  = the  $i^{\text{th}}$  Burning effect (fixed)

$\omega_{(ij)}$  = the random error effect due to the  $j^{\text{th}}$  replicate of the  $i^{\text{th}}$  burning effect assumed IID  $\sim N(0, \sigma^2)$

$\partial_{(ij)}$  = the  $ij^{\text{th}}$  random restriction error on the randomization of the fertilizer on each of the six plots assumed IID  $\sim N(0, \sigma^2)$ . This error cannot be estimated.

$F_k$  = the  $k^{\text{th}}$  fertilizer effect (fixed)

$BF_{ik}$  = the fixed interaction effect of the  $i^{\text{th}}$  burning effect and the  $k^{\text{th}}$  fertilizer effect

$\omega F_{(ijk)}$  = the mixed interaction effect of the  $k^{\text{th}}$  fertilizer and the  $j^{\text{th}}$  replicate within the  $i^{\text{th}}$  burning

$\epsilon_{(ijk)}$  = the random error due to the  $ijk^{\text{th}}$  measurement assumed to be IID  $\sim N(0, \sigma^2)$ .

This error term cannot be estimated in the experiment.

To determine whether our favourite tests are testable we developed the expected mean squares (EMS) table. The expected mean squares table for this linear model is shown in Table 1. From this table we can write down the correct tests of our hypothesis of interest for which we started the experiment. For purposes of comparison, analysing the experiment as CRD and RBCD are added to show that these would give us the wrong tests.

Also note that if the main plot part of the experiment were designed as RCBD with three blocks rather than three (genuine) replicates, the test for burning ( $(B)=0$ ) would have a reference distribution of  $F_\alpha(1,2)$  which is less sensitive than the design above with a reference distribution of  $F_\alpha(1,4)$  provided blocking was just for convenience but not essential.

### Analysis

The experiment was analysed using the linear model specified in Eqn 1. All three packages were used one at a time for the analysis. Genstat Discovery 4 (version 10.3.806) (Payne, R.W., Murray, D.A., Harding, S.A., Baird, D.B. & Soutar, 2007), SPSS (v18.0) (SPSS, 2009) and R (v 3.0) (R Development Core Team, 2013) were all running on a Sony Vaio laptop. The experiment was also analysed semi-manually in excel by first calculating the means of the six experimental units in each of the burn/no burn plots and using the means to run a one-way ANOVA. This should be the correct analysis of the land preparation experiment with or without the fertilizer experiment. From the ANOVA output of SPSS, the mean square for burning was also calculated using  $MS(B)/MS(\omega)$  and a Reference distribution of  $F_\alpha(1,4)$  as dictated by the expected mean squares (Table 2).

### RESULTS

The result of the semi-manual analysis of the main plot as a CRD is shown in Table 3. The result shows the correct tests for the

land preparation experiment. The result shows that the Burning treatments were not significantly different at the 5% level of probability. The results of the other statistical packages are compared with this. It is therefore clear that SPSS output appeared to be wrong but not that of Genstat or R.

Genstat results are shown in Table 4 and are basically correct. Genstat recognises the model structure and tests our three hypotheses stated in Table 2 with the correct mean squares and degrees of freedom.

SPSS analysis output is shown in Table 5a. This output has the correct sums of squares and degrees of freedom but the test for the land preparation (Burning) appears to be wrong. The test for the land preparation is highly significant in Table 5a when indeed it is not at 5% level of probability (Table 4). However, if we manually calculate the F-value as dictated by the expected mean squares (Table 2), we get  $F_{\text{calculated}} = 5.04$ , implying we fitted the correct linear model but SPSS does not “understand” the data structure. By including the command “/test = MS(B) vs MS( $\omega$ )”, Table 5b is produced in addition and this contains the correct test for the land preparation experiment.

## DISCUSSION

The results indicate that Genstat gets the result right. However, in the analysis we had to create a factor “block” which did not exist in the linear model. This can be a source of confusion because the linear model does not have a block also if the experiment is analysed as RCBD we get it wrong. The blocks here refer to the experimental units of the land preparation experiment. Each experimental unit becomes a block relative to the fertilizer experiment because fertilizers were randomized in each unit of the six burning plots one at a time. Indeed Genstat requires a block to be specified in order to analyse any split plot experiment and will not proceed without a block being specified.

In SPSS, when the model is specified as shown in Eqn 1, the result is as in Table 5a. The test for the land preparation is highly significant when indeed it is not at 5% level of probability. Although the sums of squares are right the test for the land preparation experiment is wrong. This is very dangerous for those who may not have clear thought about the experiment and hope that SPSS might save the situation. It is therefore important to develop the tests for any proposed experimental design before implementing it. It is also important that the computer output should be checked against the result expected before accepting it as correct. In this situation the garbage out of the computer is as a result of “mis-understanding of the model structure” by the software but not due to a mis-specification of the linear model. To force SPSS to do the right thing it is important to click on “Paste” which will convert the windows model to SPSS code. The line “/Test = burning vs burning\*block” is added to the code. This tells SPSS to test burning effects with burning\*blocks which is the error term for the burning effects. SPSS prints an additional table shown in Table 5b which is the result we expected.

The result produce by R is correct and conforms to that produced by Genstat. R can therefore be used easily for this situation. However, R will require typing the code unlike Genstat which has a fully developed menu-driven interphase.

With regard to ease of use Genstat and SPSS have better developed menu-driven interphase than R. They are therefore easier to learn but this can also lead to mis-use by obtaining meaningless statistics as demonstrated here for SPSS. Genstat incorporates better features to guard against abuse.

## CONCLUSION

This experiment shows that it is important to develop tests for real experiments before they are carried out. This has important implications. First any flaw in the design of the experiment would be noticed and corrected. This experiment

also demonstrates that what may seem straight forward is sometimes not. CRD and probably RCBD involving only fixed factors have tests that are almost always apparent but real field experiments are usually more sophisticated and the tests may not be common senses but have to be worked out. Needless to say textbooks cannot be used as cookbooks for real field experiments. Experiments must be designed based on efficiency, resources available including the sets of treatments, the nature

of experimental material and the precision required. The linear model, expected mean squares and tests of any hypothesis of interest must be developed. When the experimenter is satisfied that his requirements are met based on any constraints encountered should money, effort and time be spent on the actual field implementation. This experiment cannot be assumed to be an isolated case and therefore care should be taken in carrying out real field experiments.

## REFERENCES

- Gomez, K. A., & Gomez, A. A. (1984). *Statistical procedures for agricultural Research* 2<sup>nd</sup> Ed.,. New York: John Wiley & Sons Inc. 680pp
- Milliken, G. A., & Johnson, D. E. (1996). *Analysis of Messy Data*. London: Chapman and Hall. 473pp
- Montgomery, D. C. (1991). *Design and Analysis of Experiments*. 3<sup>rd</sup> Ed. New York: John Wiley & Sons Inc. 649pp
- Payne, R. W., Murray, D. A., Harding, S. A., Baird, D. B. & Soutar, D. M. (2007). *GenStat for Windows Introduction* 10th ed.,. Hemel Hempstead: VSN International. 810pp
- R Development Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing. Retrieved from URL <http://www.R-project.org>
- SPSS. (2009). *PASW® Statistics 18 Command Syntax Reference* 18th ed. Chicago: SPSS Inc. 2324pp
- Steel, R. D. G., & Torrie, J. H. (1980). *Principles and procedures of statistics: A biometrical Approach*. New York: McGraw-Hill Inc. 633pp

TABLE 1  
Expected mean squares table of a split-plot design with a CRD main plots

Source of Variation	EMS	df
B <sub>i</sub>	$\sigma^2 + 3\sigma_{\theta}^2 + 3\sigma_{\omega}^2 + 9\varphi(B)^*$	1
$\omega_{(i)j}$	$\sigma^2 + 3\sigma_{\theta}^2 + 3\sigma_{\omega}^2$	4
$\theta_{(ij)}$	$\sigma^2 + 3\sigma_{\theta}^2$	0
F <sub>k</sub>	$\sigma^2 + \sigma_{\omega F}^2 + 3\varphi(F)^*$	2
BF <sub>ik</sub>	$\sigma^2 + \sigma_{\omega F}^2 + 3\varphi(BF)^*$	2
$\omega F_{(i)jk}$	$\sigma^2 + \sigma_{\omega F}^2$	8
$\epsilon_{(ijk)}$	$\sigma^2$	0
Total		17

\* $\varphi(F)$ ,  $\varphi(BF)$  and  $\varphi(B)$  are summarized forms of writing the EMS of the fixed factor effects.

TABLE 2  
Tests of Hypotheses with reference Distribution of a split-plot also analysed as a CRD

Split-plot Design (Correct Design)			CRD (Wrong Design)		RCBD (Wrong Design)	
Hypothesis	Test statistic	Ref Dist.	Test statistic	Ref Dist.	Test statistic	Ref Dist.
$\varphi(B)=0$	MS(B)/MS( $\omega$ )	F $_{\alpha}(1, 4)$	MS(B)/MS( $\epsilon$ )	F $_{\alpha}(1,12)$	No test	-
$\varphi(F)= 0$	MS(F)/MS( $\omega F$ )	F $_{\alpha}(2, 8)$	MS(F)/MS( $\epsilon$ )	F $_{\alpha}(2,12)$	No test	-
$\varphi(BF)=0$	MS(BF)/MS( $\omega F$ )	F $_{\alpha}(2, 8)$	MS(BF)/MS( $\epsilon$ )	F $_{\alpha}(2,12)$	No test	-

TABLE 3  
One-way ANOVA analysis of means of the main plots (semi-manual analysis). This result shows the correct analysis of the land preparation experiment

Source of Variation	Sum of Squares	Degrees of freedom		Mean Squares	F-calculated	P-value	F-critical
Burning	15.57	1		15.57	5.036	0.088	7.709
Error	12.37	4		3.09			
Total	27.94	5					

TABLE 4  
Analysis of a split-plot with a CRD main-plot; Genstat produces correct result

Source of variation	Degrees of freedom		Sum of squares	Mean squares	F-calculated	p-value.
burning	1		46.722	46.722	5.04	0.088
Residual	4		37.111	9.278	4.39	
fertilizer	2		36.000	18.000	8.53	0.010
burning.fertilizer	2		5.778	2.889	1.37	0.308
Residual	8		16.889	2.111		
Total	17		142.5			

TABLE 5A

SPSS analysis showing that the test for burning is significant when indeed it is not at 5% level of probability. This occurs because the variance ratio (F) and the reference distribution are both wrong.

Source	Type III Sum of Squares	Degrees of freedom	Mean Square	F-calculated	Sig.
Intercept	10224.500	1	10224.500	4843.184	.000
burning	46.722	1	46.722	22.132	.002
block * burning	37.111	4	9.278	4.395	.036
fertilizer	36.000	2	18.000	8.526	.010
burning * fertilizer	5.778	2	2.889	1.368	.308
Error	16.889	8	2.111		
Total	10367.000	18			

TABLE 5B

SPSS output showing the correct test for the burning experiment to be non-significant at 5% level of probability. This test has to be specifically requested in the syntax.

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	46.722	1	46.722	5.036	0.088
Error	37.111	4	9.278		

TABLE 6

Analysis with R gives correct result. Again block have to be specified to get it correct

```

Error: block
      Df Sum Sq Mean Sq F value Pr(>F)
burning  1  46.72   46.72   5.036 0.0882 .
Residuals 4  37.11    9.28
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: block:fertilizer
      Df Sum Sq Mean Sq F value Pr(>F)
fertilizer  2  36.00   18.000   8.526 0.0104 *
burning:fertilizer  2   5.78    2.889   1.368 0.3082
Residuals  8  16.89    2.111
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figure 1: A split-plot design in which the main (burning) plots are completely randomized and replicated three times with fertilizer (nitrogen, phosphate and NPK) randomized in each of the six experimental units of the burning experiment

	<b>Burn</b>			<b>Noburn</b>			<b>Burn</b>		
Nitrogen	NPK	Phosphate	NPK	Nitrogen	Phosphate	NPK	Phosphate	Nitrogen	
Phosphate	Nitrogen	Phosphate	NPK	Phosphate	Nitrogen	NPK	Nitrogen	Phosphate	
	<b>Burn</b>			<b>Noburn</b>			<b>Noburn</b>		

## APPENDICES

### Appendix 1: Data used for the analysis (artificial)

plot no	block	burning	fertilizer	Yield of mango (kg/plot)
1	1	burn	Nitrogen	25
2	1	burn	Phosphate	27
3	1	burn	NPK	29
4	2	no_burn	Nitrogen	19
5	2	no_burn	Phosphate	20
6	2	no_burn	NPK	24
7	3	burn	Nitrogen	25
8	3	burn	Phosphate	21
9	3	burn	NPK	25
10	4	no_burn	Nitrogen	24
11	4	no_burn	Phosphate	22
12	4	no_burn	NPK	27
13	5	burn	Nitrogen	25
14	5	burn	Phosphate	26
15	5	burn	NPK	26
16	6	no_burn	Nitrogen	19
17	6	no_burn	Phosphate	21
18	6	no_burn	NPK	24



## Appendix 2: SPSS AND R COMMANDS FOR ANALYSIS OF DATA

SPSS Syntax (code) used to analyse the experiment

```
UNIANOVA yield BY block burning fertilizer  
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/CRITERIA=ALPHA(0.05)  
/DESIGN=burning block*burning fertilizer burning*fertilizer  
/Test =burning vs block*burning.
```

Appendix 3: R code used to analyse experiment

```
AnovaModel.1 <- (aov(Yield ~ burning*fertilizer + Error(block/fertilizer), data=splitplot))  
summary(AnovaModel.1)
```