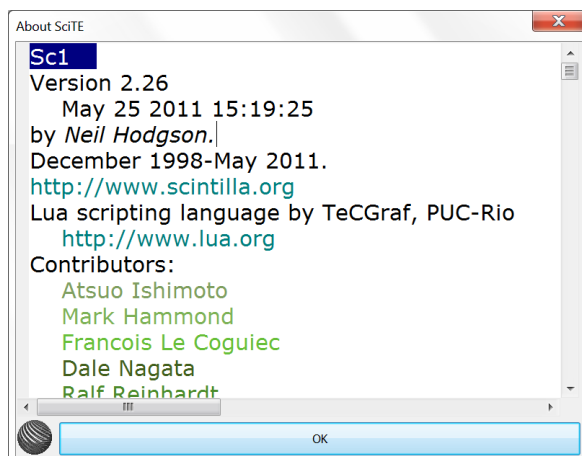


Wolfgang Link wlink@gwdg.de
2014

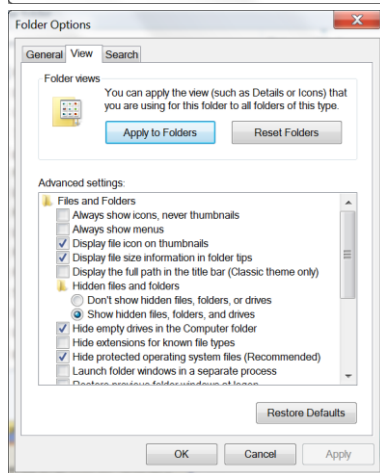
Beginners course on PLABSTAT (Software to analyse plant breeding experiments; programmed and cared for by F. Utz, Hohenheim)
<https://www.uni-hohenheim.de/plantbreeding/software>

Plabstat is working with ASCII-data only. Thus, whatever we do, we have to save data as “text only”, as “*.txt”, as “Text (MS Dos, *.txt)” etc., depending on whether you save from MS Word or from EXCEL etc.



I add the SC1.exe file; the SC1 is an editor which allows to easily take data from EXCEL into a Plabstat-input file without the usual “ASCII-versus-*.xlsc calamities”.

We employ “PLABSTAT VERSION 3Awin.exe of 2005” with 988KB”. And we have the to-be-analyzed data files as well in the very same folder where the Plabstat-software



is located. We make sure that the path to this folder is as short as possible, such as d:\myplabstat\; no gaps or other peculiarities in file names.

Have your Microsoft Explorer adjusted so that it shows the extension of file names (*.xyz, such as *.xls or *.xlsx). This may be done via ‘Folder Options’ and not tick at ‘Hide extensions for known file types’



If you open and save an ASCII-file with the mentioned SC1-Editor or with your Microsoft Editor, then your file will stay to be ASCII and you will (normally) have no problems. Names of ASCII files must not be longer than eight digits!

The first file that we open (with EXCEL) is called: FCLD1Y1.xls (Frost Chamber Load1 Year1); it holds the data and the randomization of the experiment.

We save the data as ASCII-File (Text MS-Dos *.txt); FCLD1Y1.txt
 We add the **BASIC** commands and save it as FCLD1Y1.dat (still ASCII).
 We make Plabstat run and find e.g. a Pearson correlation between the two traits “DS-P1” and “DS-P2” of $r=0.816^{**}$, compared to Spearman $r=0.698^{**}$ (DS = Disposition to survive).

We open again the FCLD1Y1.dat file and prepare now the honest “lattice analysis”. We add the necessary **LATTICE** commands (take them from the file ‘head for lattice.txt’) and save it then as LD1Y1LAT.dat - and make PLABSTAT run this file.

The LD1Y1LAT.prt (‘prt’ means ‘print’ means ‘output’, ‘result’) file shows, among many other results, that the Pearson correlation between DS-P1 and DS-P2 became $r=0.852^{**}$; different from what we have seen before. Why, what is (are) the difference(s)?

We open again the LD1Y1LAT.dat file, and substitute the LATTICE commands by **ANOVA** commands (‘head for anova.txt’); and save it again, now as LD1Y1RBD.dat (RBD=Randomized complete Block Design); we see that the Pearson correlation between DS-P1 and DS-P2 became $r=0.876^{**}$; why, what is the difference to the $r=0.852^{**}$?

We compare the prt-file from LATTICE with that one from ANOVA, e.g. for trait 10 (DSURV).

ANOVA of RCBD

Source	DF	MS	Var.cp	F	DF-NM	DF-DN	LSD5
R	1	1258.9320	24.8735	5.88*	1.00	40.00	6.46
G	41	631.3833	208.5689	2.95**	41.00	40.00	29.58
RG	40	214.2456	214.2456				

LATTICE ANALYSIS

Source of var.	DF	MS	Var.cp.	F-value
Replications	1	1232.4650	15.2463	5.75 *
Treatm(unadj)	41	636.9651	236.4380	3.74 **
Blocks (adj)	12	369.9257	74.0648	2.50 *
Intrabl.error	28	147.7314	147.7314	

Character	SD(tr)	C.V.%	Repeat1%	Effic.	Eff.Err.MS
10 DSURV	17.33	19.3	57.6	123.1	174.1357

“Repeat1%” means repeatability (heritability) based on $r=1$.

$$236.4380 / (236.4380 + 174.1357) = 0.5759 \quad (57.6\%)$$

If you want to know the repeatability based on the actual $r=2$, then:

$$\text{Repeatability} = 236.4380 / [236.4380 + (\frac{1}{2})(174.1357)] = 0.7309 \quad (73.1\%)$$

From ANOVA we get the repeatability as:

$$\text{Repeatability} = 208.5689 / [208.5689 + (\frac{1}{2})(214.2456)] = 0.6607$$

Thus, it was really better to randomize as lattice and reduce the error term; this is why - for this trait - the lattice efficiency is shown to be $>100\%$ (123.1%).

There was, in the next winter season of 2010/11, a test of the same genotypes, just like in 2009/10. The *.mzv file pertaining to this experiment is available, the name is LD1Y2LAT.mzv.

*.mzv files (mzv = MeaN Values) hold the mean values as specified in the ANOVA commands.

So why not analyzing these two experiments together – as one experiment that ran across two seasons? So indeed a joint file was created. It contains the *.mzv files of the first year’s analysis in the upper part and that of the second year in its lower part. This joint file needs now some editing. The “head for anova” was added on top; yet, it now did not fit to our needs. It was necessary to adapt it ... ;

And we deleted the lines addressed as ‘C.V’, ‘REPY’, ‘EFFY’, ‘CHEC’ and ‘MEAN’ that are given with such *.mzv files; they are dispensable for our current purpose.

This edited, joint file is available under the name LD1Y1Y2.dat. Analyze it by Plabstat. It is a file that shows how a series (here of two seasons) of experiments can be analyzed without losing the information about the single experiments’ errors.

We use the same file again, yet add the GENOT statement (cf. chapter 6.1.9 in the PLABSTAT manual), and save it und name of L1-2YCOV.dat, meaning Load1-2yearsCOVARIANCE-ANALSIS, and analyze it again by Plabstat.

ANOVA of L1-2YCOV.dat

```
----- 10*10 DSURV*DSURV -----
Source DF  MS      Var.cp  V.cp%  F        DF-NM  DF-DN
Y       1   7070.4423  166.3403  98.81  84.02**  1.00  39.00
G      41   284.0668   99.9582  70.38   3.38**  41.00  39.00
GY     39    84.1505   84.1505
-----
```

ANOVA of LD1Y1Y2.dat

Source	DF	MS	Var. cp	F	DF-NM	DF-DN	LSD5
Y	1	7182.20	168.86	79.75**	1.00	39.00	4.19
G	41	307.11	<u>108.53</u>	3.41**	41.00	39.00	19.19
GY	39	90.05	40.61	1.82*	39.00	52.00	19.95
Error	52	49.44	49.44				

HERITAB **70.68** (45.13 84.41)

We find that now the ANOVA, e.g. for trait 10, differs from the ANOVA of LD1Y1Y2 (where we had the ERRORS of the single experiments, analyzed as LATTICE, included). N.B.: Missing data are treated differently when it comes to covariance analysis, compared to variance analysis.

We find all the covariance analyses between all pairs of traits, such as between

```
<9*10 REGRW*DSURV>,
<9*11 REGRW*TL+CL>
<10*11 DSURV*TL+CL>.
```

```
*****      9*10      REGRW*DSURV      *****
Source DF      MP      Cov. cp      r-phen      r-gen
Y      1      1133.777      27.0887      0.000      0.000
G      41      13.929      8.9392      0.627**      1.333++
GY     39      -3.949      -3.9489      -0.470**      -0.470++
*****
```

What could be the purpose to run a Covariance Analysis (anyway only meaningful if you have data across more than one 'environment')?

We may wish to improve our main trait

- (1) "**Disposition to Survive**" by adding information from
- (2) "Sum of Turgor and Colour Losses".

Why not creating an **index** with DSURV as main trait and TL+CL as auxiliary traits? Admittedly, this strategy would only be meaningful if having more than just two seasons, because with few seasons the pertinent parameters suffer from large error terms.

To get the necessary weighing factors b_1 , and b_2 , we have to solve these equations (cf. Smith and Hazel Optimum Index)

$b_1 \sigma^2_{P(1.1)}$	+	$b_2 \omega_{P(1.2)}$	+	=	$\sigma^2_{G(1.1)}$
$b_1 \omega_{P(1.2)}$	+	$b_2 \sigma^2_{P(2.2)}$	+	=	$\omega_{G(1.2)}$
$\sigma^2_{P(1.1)}$		phenotypic variance of trait 1 (main trait, DSURV),			
$\sigma^2_{P(2.2)}$		phenotypic variance of trait 2 (auxiliary trait, TL+CL),			
$\omega_{P(1.2)}$		phenotypic covariance between traits 1 and 2,			
$\sigma^2_{G(1.1)}$		genotypic variance of trait 1,			
$\omega_{G(1.2)}$		genotypic covariance between traits 1 and 2.			

From a due inspection of the L1-2YCOV.prt file we find:

142.0334	b_1	-26.3401	b_2	=	99.9582
-26.3401	b_1	+ 9.5143	b_2	=	-24.6874

// 9.5143 comes from 19.0286/2; -26.341 from -52.6802/2 //

Solve the two equations, find:

$b_1 = 0,4574$; and $b_2 = -1,3285$. Thus, we may add this index as a further trait.

Index = $0.4574 * \text{"DSURV"} + (-1.3285) * \text{"TL+CL"}$

We open again file LD1Y1Y2.dat and add, via the CALCULATE statement, this index. Save under new name: L12YINDX.dat and run Plabstat.

From inspecting the PLOTS of the index versus its component traits and versus REGRW we may decide whether what we did was useful or not.

As next step we are going to analyse a factorial experiment.

The diagram shows two replicates of a split-plot experiment on alfalfa yield, with main treatments (varieties) and sub-treatments (four dates of final cutting in the year before the actual yield assessment; SNEDECOR and COCHRAN, 7.ed., p. 327). The final cuttings in the preceding year were: none, Sept. 1 (S1), Sept. 20 (S20) and Oct. 7 (O7). Six replicates were grown, three varieties (Ladak, Cossak and Ranger) were tested. A late final cutting last year may reduce alfalfa yield in this year.

From the yield data we create a second trait X2 by logarithmic transformation (LN(x) *10). The split-plot model is

$$B + V + BV/ + D + VD + DBV$$

The “/” after BV forces BV to be error term for the left side of it (B, V)

Nota bene: This is Plabstat-Jargon. In a scientific paper, the model would rather be written as: $Y_{ijkl} = \mu + b_j + v_i + vb_{ji} + d_k + vd_{ik} + bvd_{jik}$

Study the table of Expected Composition of Mean Squares

MS	Components of variance & their coefficients					
	BDV	DV	D	BV	V	B
	1	B	BV	D	BD	DV
B	1			1		1
V	1			1	1	
BV	1			1		
D	1		1			
DV	1	1				
BDV	1					

It says, e.g., that the MS(V) is composed from $(1) \sigma^2(\text{BDV}) + (D=4) \sigma^2(\text{BD}) + (BD=24) \sigma^2(\text{V})$

And it says that the MS(BV) is composed from $(1) \sigma^2(\text{BDV}) + (D=4) \sigma^2(\text{BD})$

Thus, if there is any true V-Variance contained in the MS(V), then MS(V) should be larger than MS(BV). This is why MS(V)/MS(BV) is calculated, and if >1.0 then (via 5% error threshold and degrees of freedom) the **F-test** for ‘Variety’ as a source of variation is conducted.

BLOCK I								
Cossack			Ranger			Ladak		
S1	07	S20	NONE	S20	07	S1	S1	07

BLOCK II								
Ladak			Cossack			Ranger		
07	S1	S20	NONE	S20	07	S1	NONE	S20

Fig. 16.15.1—First two blocks of a split-plot experiment on alfalfa, illustrating random arrangement of main and subplot treatments.

If we calculate MS(V) minus MS(BV), the result is $(24) \sigma^2(\mathbf{v})$; thus, divide by 24 and you come to the estimate of $\sigma^2(\mathbf{v})$, which is the Varieties' Variance Component (24 is number of DATES times number of BLOCKS).

Obviously, the Mean Square BV holds the error term for the varieties, whereas the Mean Square DV holds the error for the dates.