THE AMMI ANALYSIS AND GRAPHING THE BIPLOT

Mateo Vargas Hernández
Universidad Autónoma Chapingo, Mexico
Biometrics and Statistics Unit, CIMMYT, Mexico

José Crossa
Biometrics and Statistics Unit, CIMMYT, Mexico

CIMMYT, INT.
Apdo. Postal 6-641
06600 Mexico DF
México

March 2000

Contents

Examples

COMPUTING AMMI ANALYSIS AND GRAPHING THE AMMI BIPLOT

Mateo Vargas Hernández and José Crossa

Introduction

With the objective of facilitating researchers of CIMMYT and National Programs to compute the AMMI model and the biplot of multi-environment trials, three programs were written in SAS for computing the AMMI analysis and using the Gollob test (Gollob, 1967) for determining the significant number of multiplicative terms. For each of these programs the graphic of the biplot is obtained.

The three cases are

1) when only adjusted means are available. In this case, the AMMI analysis and its biplot can be computed. For the Gollob test, one must have an estimate of the combined error variance, its degrees of freedom (df), and the number of replicates; this information must be provided to the SAS program.

2) when the information for each replicate in each genotype-environment combination is available and the design for each site is randomized complete block (RCB). In this case, the PROC GLM of SAS is used and the combined error variance with the appropriate degrees of freedom are automatically given.

3) Similar to case 2, but the design for each site is an incomplete block design (i.e., lattice). In this case, the PROC MIXED of SAS is used to recover the inter-block information.

In the following sections we present detailed descriptions of each of these three programs. The three programs share several routines; these shall be explained in detail only for Case 1. For Cases 2 and 3, only new codes will be described.

Note that ***italic letters*** will be used to identify the actual codes of the program, whereas "**normal letters**" are used to identify the descriptive text.

Users who need to access data, programs, outputs or graphics contained in this document, please visit our Website www.cimmyt.cgiar.org/biometrics, and download the file AMMI.EXE.

Case 1: Computing AMMI analysis using adjusted means

In this case, one must introduce into the SAS codes the value of the combined error variance (mean square error), degree of freedom (df), and the number of replicates.

The first line determines the size of the page (5000 lines), the width of the line (78 characters) and no date for the analysis is required.

options ps = 5000 ls=78 nodate;

The following two lines are used to generate the biplot graph as a CGM (Graphics Metafile) file such that it can be exported to Power Point with the name "*example1.cgm*" and emplying the option *device=cgmmwwc*. To replace an old graph with a new one use *gsfmode=replace*. Another option for adding to the graph is to use *gsfmode=append*.

filename biplot 'example1.cgm';

goptions device=cgmmwwc gsfname=biplot gsfmode=replace;

(Note: To see the graph on the screen, cancel these two lines by adding an asterick at the beginning of the lines)

The following section shows the clasical way of reading an external file called "***example1.dat***" in which the order and the type of the variables is established. It is important that the variables that identify the environments (*env*) and genotypes (*gen*) are declared as alphanumeric variables by using the sign $ after the name of the variable.

data raw;

infile'example1.dat';

input env $ gen $ yld;

cards;

Once the data have been read, the **Analysis of Variance** (AOV) is obtained using the adjusted means. Remember that this analysis can't estimate a combined error variance thus F and probability values are not shown. To obtain the correct sum of squares (corrected by the number of replicates) and their significance, it is necessary to

introduce a value for the combined error variance, error df, and the number of replicates (*mse, dfe, y nrep*, respectively).

```
title1  'Analysis of Variance with the adjusted means ';
title2  'no corrected by the number of replicates';
proc glm data=raw outstat=stats ;
 class env gen;
 model yld = env gen env*gen/ss4;
```

(Note: For Example 1, results are shown in Table 1.1)

```
title1  'Analysis of Variance corrected by the number of replicates';
data stats2;
  set stats ;
  drop _name_ _type_;
  if _source_ = 'ERROR' then delete;
  mse=251943;      * combined error variance;
  dfe=478;         * error df ;
  nrep=3;          * number of replicates ;
  ss=ss*nrep;
  ms=ss/df;
  f=ms/mse;
  prob=1-probf(f,df,dfe);
proc print data=stats2 noobs;
  var _source_ df ss ms f prob;
```

(Note: For Example 1, results are shown in Table 1.2)

The values of the combined error variance, error df, and number of replicates are stored and then used for the Gollob test.

The next section obtains the two-way table of residuals corresponding to the values of the genotype x enviroment interaction (GEI).

```
proc glm data=raw noprint;
  class env gen;
  model yld = env gen / ss4 ;
  output out=outres r=resid;
proc sort data=outres;
  by gen env;
proc transpose data=outres out=outres2;
  by gen;
  id env;
  var resid;
```

The matrix of residuals (GEI), stored in file *outres2*, is subjected to a Singular Value Decomposition (SVD) using SAS IML. The sum of squares for each AMMI multiplicative term is the squared of each singular value, and the proportion of variability explained for each AMMI term (multiplicative component) is the ratio of variability explained for each AMMI term devided by the total GEI variability.

```
proc iml;
use outres2;
read all into resid;
ngen=nrow(resid);
nenv=ncol(resid);

use stats2;
read var {mse} into msem;
read var {dfe} into dfem;
read var {nrep} into nrep;

call SVD (U,L,V,resid);        * SVD of the residuall (GEI) matrix;
ss=( L ## 2 )*nrep;            * Sum of squared for the AMMI terms;
suma=sum(ss);
```

```
porcent=((1/suma) # ss)*100;    * Computing variability explained for each AMMI
                                  term ;
minimo=min(ngen,nenv);
```

The following lines calculate the df of each AMMI multiplicative term and the proportion of variability accumulated up to the i^{th} AMMI term.

```
porcenta=0;
  do i = 1 to minimo;
   df=(ngen-1)+(nenv-1)-(2*i-1);
   dfa=dfa//df;
   porceacu=porcent[i,];
   porcenta=porcenta+porceacu;
   porcenac=porcenac//porcenta;
  end;
```

The following codes are used to generate matrices with the error df (*dfem*) and the error mean squares (*msem*) such that they are compatible with the results obtained for the sum of squares (*ss*), proportion of variability explained for each AMMI multiplicative term (*porcent*), proportion of variablility accumulated (*porcenac*), and the df of each AMMI term (*dfa*). All is information is stored in one matrix called *ssdf*.

```
dfe=J(minimo,1,dfem);
mse=J(minimo,1,msem);
ssdf=ss||porcent||porcenac||dfa||dfe||mse;
```

The folowing codes obtain the genotypic and environmental ***scores*** (eigenvectors). They are weighted by the squared root of the singular values. Only eigenvectors associated to the first three multiplicative components are computed, but this can be easily extended to to any number of multiplicative components. The scores are stored in a matrix *scores*, that will be used for graphing the biplot.

```
L12 = L ## 0.5;
scoreg1=U[,1]# L12[1,];     * Genotypic scores of the first AMMI term;
scoreg2=U[,2]# L12[2,];
scoreg3=U[,3]# L12[3,];
scoree1=V[,1]# L12[1,];     * Environmental scores for the first AMMI term;
scoree2=V[,2]# L12[2,];
scoree3=V[,3]# L12[3,];
scoreg=scoreg1||scoreg2||scoreg3;
scoree=scoree1||scoree2||scoree3;
scores=scoreg//scoree;
```

Once the results of the AMMI are obtained using SAS IML, the information is stored in files created using a SAS Date Step so that it can be used for computing the Gollob test and for obtaining the biplot.

```
create sumas from ssdf;      * Craetes files to be used ;
append from ssdf;            * in the Gollob test ;
close sumas;                 * and in the biplot ;
create scores from scores;
append from scores ;
close scores;
```

(Note: This is the end of SAS IML)

The following section shows the codes for computing the Gollob test. Results are given in the following order: Sum of squares for each AMMI term; proportion of variability explained for each term; proportion of variability accumulated; df for each AMMI term; F value and probability level for each AMMI term. The order of the output can be modified changing instructions in the *Proc Print*.

```
data ssammi;
 set sumas;
 ssammi =col1;
 porcent =col2;
 porcenac=col3;
 dfammi =col4;
 dfe   =col5;
 mse   =col6;
 drop col1 - col6;
 msammi=ssammi/dfammi;
 f_ammi=msammi/mse;
 probf=1-probf(f_ammi,dfammi,dfe);
title1   'Sum of Squares of the AMMI terms';
proc print data=ssammi noobs;
  var ssammi porcent porcenac dfammi msammi f_ammi probf;
```

(Note: For example 1, results are shown in Table 1.3)

In order to get the biplot, first it is required to assign names (or numbers) to each of the genotypes and environments, as well the identification of the variables, environments (*ENV*) and genotypes (*GEN*). These values are stored in the files *namegen* y *nameenv* that later are merged in one file called *typename*.

```
* Assigning type and names to the genotypes;
    proc sort data=raw;
     by gen;
    proc means data = raw noprint;
     by gen ;
     var yld;
     output out = mediag mean=yld;
    data namegen;
```

```
  set mediag;
  type = 'GEN';
  name = gen;
  keep type name yld;

* Assigning type and names to the environments;
  proc sort data=raw;
   by env;
  proc means data = raw noprint;
   by env ;
   var yld;
   output out = mediae mean=yld;
  data nameenv;
   set mediae;
   type = 'ENV';
   name1 = 'S' || env;
   name = compress(name1);
   keep type name yld;

  data typename;     * It contains type and name for genotypes and environments;
   set namegen nameenv;
```

With the *type* and *name* of the genotypes and environents and the file with the values of the genotypic and environmental scores (stored in file *scores*) a file is generated for graphing the biplots.

```
  data biplot ;
    merge typename scores;
    dim1=col1;
    dim2=col2;
```

```
dim3=col3;
drop col1-col3;

title1  'Genotype and environmental scores for the biplot' ;
proc print data=biplot noobs;
 var type name yld dim1 dim2 dim3;
```

(Note: For Example 1, results are shown in Table 1.4)

The next step is to generate a file of the type "*Annotate.*" In SAS GRAPH this is useful because it allows us to assign variables to the coordinates, fonts (type, size, color) for the labels, titles of the coordinates, etc. The functions *MOVE* y *DRAW* are used to draw vectors from the origin of the biplot up to the values of the scores and the function *LABEL* is used to create labels.

```
Data labels;
 set biplot;
 retain xsys '2' ysys '2' ;
 length function text $8 ;
 text = name ;

  if type = 'GEN' then do;
   color='black ';
   size = 0.6;
   style = 'hwcgm001';
   x = dim1;
   y = dim2;
   if dim1 >=0
    then position='5';
    else position='5';
   function = 'LABEL';
```

```
 output;
end;

if type = 'ENV' then DO;
 color='black ';
 size = 0.6;
 style = 'hwcgm001';
 x = 0.0;
 y = 0.0;
 function='MOVE';
 output;
 x = dim1;
 y = dim2;
 function='DRAW' ;
 output;
 if dim1 >=0
  then position='5';
  else position='5';
 function='LABEL';
 output;
end;
```

[Note: the instruction *'if dim1 >=0 then position='5'; else position='5'* must be modified when, apart from assigning a name to the coordinate, it is required to use different symbols for environments and genotypes. In this case the symbol is drawn in the middle of the coordinate and the label should be moved to the right for values >0.0, and to the left for values <0.0. Thus, modification is recommended for the instruction to *if dim1 >=0 then position='6'; else position='4';*. For more information about possible positions of the labels, see the user's manual for SAS GRAPH (SAS / GRAPH, User's Guide)].

Now all the required information for obtaining the biplot is ready. Thus, the SAS GPLOT procedure is used. In this case, the second AMMI multiplicative term (*dim2*) is plotted in the y-axis and the first AMMI multiplicative term (*dim1*) is plotted in the x-axis. This can be modified to plot the third AMMI term against the first and so on. The file created in the previous step (***labels***) is used as "*annotate* " and the following items are required: 1) a frame for the biplot (*frame*), 2) vertical and horizontal reference lines through the origin (*vref=0.0* and *href=0.0*, respectively), 3) the color (*cvref, chref*) and type of reference lines (*lvref, lhref*), 4) vertical axis (*vaxis*) and horizontal axis (*haxis*), and 5) legends not included (*nolegend*).

```
Proc gplot data=biplot;
Plot dim2*dim1 / Annotate=labels frame
        Vref=0.0 Href = 0.0
        cvref=black chref=black
        lvref=3 lhref=3
        vaxis=axis2 haxis=axis1
        vminor=1 hminor=1 nolegend;

symbol1  v=none  c=black  h=0.7   ;
symbol2  v=none  c=black  h=0.7   ;

axis2
  length = 6.0 in order = (-40 to 40 by 10)
  label=(f=hwcgm001  h=1.2  a=90  r=0  'Factor 2')
  value=(h=0.8)
  minor=none;
axis1
  length = 6.0 in order = (-40 to 40 by 10)
  label=(f=hwcgm001  h=1.2  'Factor 1')
  value=(h=0.8)
  minor=none;
```

Title1 f=hwcgm001 h=1.0 'AMMI biplot for example 1 using adjusted means';

run;

Note that *symbol* specifies not to use the option *value* (*v=none*). If symbols are to be used, then *value* can be changed to *v=square, v=circle, v=triangle, v=diamond, v='*', v='+', v='A', v='B'*, etc. The size and color of the symbols can also be modified.

The instructions *Axis1* y *Axis2* are used with the following option:

length = 6.0 in that indicates the longitude of the axes in inches.

order = (-40 to 40 by 10) indicates the maximum and the minimum values of the axes and the increments. The first time this command can be cancelled so that SAS GRAPH generates these values by default.

label = (f = hwcgm001 h = 1.2 a=90 r=0 'Factor 2') indicates font, size, angle, rotation and title of the corresponding axis.

value=(h=0.8) ,it controls the size of the marks on the axes.

minor=none indicates that no minor divisions are required on the axes.

The instruction *title1* controls the font, size, color, and label of the main title of the biplot. Two or more titles can be specified simultaneously.

More options are available in the SAS GRAPH user's manual guide.

(Note: For example 1, the biplot is depicted in Fig. 1.1)

Often one must plot the first AMMI term against the mean yield of the genotypes and sites in the x-axis. This is not a biplot, but it can provide information about the main effects of genotypes and sites. In order to do this, the instruction ***plot dim2*dim1*** should be modified for ***plot dim1*yld*** and adapt (or cancel) the options asociated with the vertical and horizontal reference lines (*vref, href, cvref, chref, lvref, lhref*, etc). An example of this biplot is depicted in Figure 1.2.

Example 1

This example contains data from a wheat variety trial with 8 genotypes tested during 6 years (1990-95) in Cd. Obregón, México. In each year the genotypes were arranged in a complete block design with 3 replicates. The 8 genotypes correspond to an historical series of varieties released from 1960 to 1980 with the numbers 1-8 representing the order of release (Sayre et. al. 1997). The mean grain yields are in file ***example1.dat.***

To obtain the AMMI analysis, run the SAS program ***example1.sas***, .Here it is necessary to intoduce the combined error variance (0.1580245), the error df (94) and the number of replicates (3). The combined error variance was previously obtained from the combined analysis of variance considering randomized complete blocks.

The results are shown in file ***example1.lst***, which includes the following sections.

Table 1.1. Analysis of variance of Example 1 using the adjusted means without correcting for the number of replicates

Analysis of Variance of the adjusted means
no corrected by the number of replicates

General Linear Models Procedure

Dependent Variable: YLD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	47	36.27979794	0.77191059	.	.
Error	0	.	.		
Corrected Total	47	36.27979794			

R-Square	C.V.	Root MSE	YLD Mean
1.000000	0	0	7.053890

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
ENV	5	16.39991745	3.27998349	.	.
GEN	7	14.25289126	2.03612732	.	.
ENV*GEN	35	5.62698924	0.16077112	.	.

Note that the total df is 47 (48 observations = 6 years x 8 genotypes, no replicates). The combined error variance can not be estimated, therefore significant values are not available. The sum of squares are not corrected by the number of replicates.

In Table 1.2, the sum of squares are corrected by the number of replicates and the combined error variance as well as the error df have been introduced into the program so that a complete AOV can now be obtained, including a probability level for each term of the model.

Table 1.2. Analysis of variance of Example 1 corrected for the number of replicates

Analysis of variance corrected for the number of replicates.

SOURCE	DF	SS	MS	F	PROB
ENV	5	49.1998	9.83995	62.2685	.0000000000
GEN	7	42.7587	6.10838	38.6547	.0000000000
ENV*GEN	35	16.8810	0.48231	3.0521	.0000096813

In Table 1.3, the Gollob F-test (Gollob, 1967) for the significance of each AMMI term are presented. The first column shows the sum of squares corrected by the number of replicates (SSAMMI); the second column shows the percent of the GEI sum of squares explained by each AMMI term (PORCENT); and the third column has the cumulative percent of the GEI sum of squares explained until the i^{th} AMMI term (PORCENAC). The other columns show the df of each AMMI term (DFAMMI), their mean squares (MSAMMI), their F values (F_AMMI), and the probability level associated to each F test for each AMMI term. In this case, it is seen that the first three AMMI multiplicative components are significant at the 1% probability level.

Computer simulations have shown that the Gollob F-test is very liberal and can result in many multiplicative terms judged significant (Cornelius et al. 1996). The Gollob F-test does not control Type I error that can run as high as 60% (Cornelius et al. 1996).

Table 1.3. Gollob F-test for the AMMI terms of Example 1

Sum of squares for the AMMI terms

SSAMMI	PORCENT	PORCENAC	DFAMMI	MSAMMI	F_AMMI	PROBF
7.24287	42.9055	42.906	11	0.65844	4.16671	0.00005
5.42327	32.1265	75.032	9	0.60259	3.81324	0.00039
2.96965	17.5917	92.624	7	0.42424	2.68462	0.01403
1.19061	7.0530	99.677	5	0.23812	1.50686	0.19509
0.05457	0.3233	100.000	3	0.01819	0.11511	0.95106
0.00000	0.0000	100.000	1	0.00000	0.00000	1.00000

Table 1.4 shows the mean yield (YLD) of genotypes and the genotypic and environmental scores of the first three AMMI components (DIM1, DIM2, DIM3) as well the values of the variables used to generate the biplot (TYPE and NAME). The variable NAME has the labels used in the biplot for each genotype and environment. The biplot for this example is shown in Figure 1.1 and it is in file **example1.cgm** (this file is automatically generated by SAS and can be exported and inserted in to any **Word** or **Power Point** document).

Table 1.4. Information for the biplot of Example 1

Genotypic and environmental scores for the first three AMMI terms (DIM1, DIM2 and DIM3)

TYPE	NAME	YLD	DIM1	DIM2	DIM3
GEN	1	6.13178	0.00956	-0.40614	0.71637
GEN	2	6.45306	-0.20785	0.11243	-0.42252
GEN	3	6.82289	0.54723	0.38688	-0.17175
GEN	4	6.97439	-0.55039	0.48510	0.32201
GEN	5	7.10122	0.79807	0.22696	0.09611
GEN	6	7.50572	-0.50527	0.31346	-0.10796
GEN	7	7.65789	-0.12223	-0.61605	-0.38304
GEN	8	7.78417	0.03088	-0.50265	-0.04923
ENV	S90	7.57075	0.26094	0.88469	-0.35836
ENV	S91	7.03321	0.14700	0.24992	0.80484
ENV	S92	6.28142	-0.39629	-0.30402	0.23481
ENV	S93	7.46617	0.91858	-0.58360	-0.19014
ENV	S94	7.69529	-0.58882	-0.25741	-0.30403
ENV	S95	6.27650	-0.34140	0.01041	-0.18713

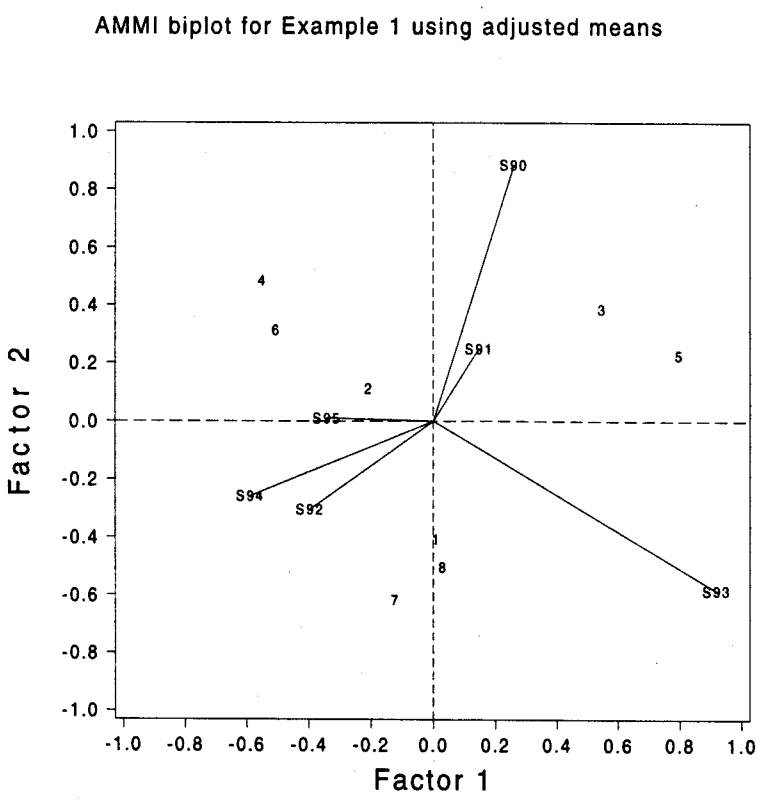

Figure 1.1. Biplot for Example 1 using adjusted means.

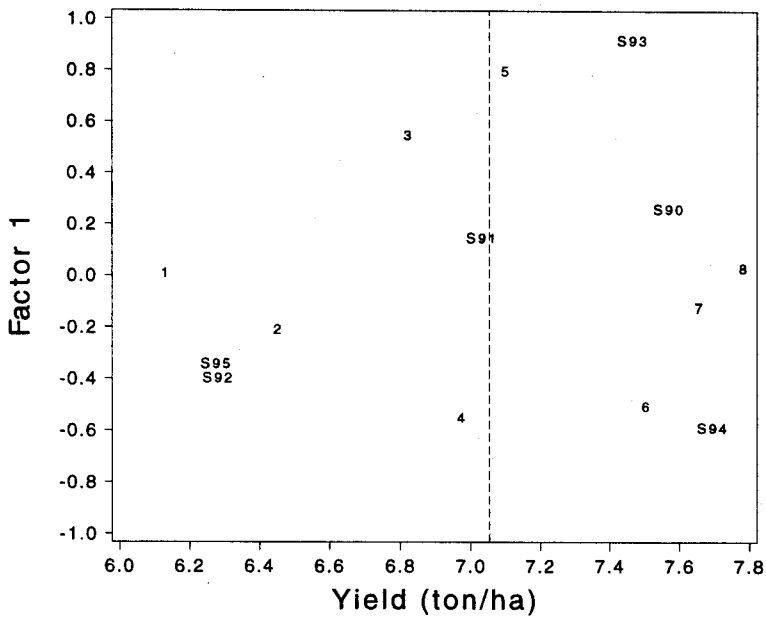

Figure 1.2. Alternative AMMI plot for Example 1.

In Fig. 1.2, the general mean of the experiment is depicted as a vertical line and all genotypes and years with grain yield larger than the general mean are shown to the right of that line. The SAS program file is called ***alterna1.sas*** and the graph is in the file ***alterna1.cgm***.

Example 2

In this example, the data set comes from a wheat trial with 24 agronomic treatments evaluated during 10 years (1988–97) en Ciudad Obregón, México. In each year, the treatments were arranged in a randomized complete block design with 3 replicates. The treatments comprised the combination of four factors at the following levels: tillage at two levels (1=without deep knife, 2=with deep knife), summer crop at two levels (1=sesbania (*Sesbania spp*), 2=soya (*Glycine max* L.)), manure at two levels (1=with chicken manure, 2=without chicken manure), and application of nitrogen at three levels (1=0 kg N ha^{-1}, 2=100 kg N ha^{-1} y 3=200 kg N ha^{-1}). This results in 2×2×2×3=24 treatments. Using the standard notation for factorial experimens, treatment 1 is denoted as [1-1-1-1], treatment 2 is [2-1-1-1], treatment 3 is [1-2-1-1], and so on until treatment 24 [2-2-2-3].

The objective for studying this data set is to investigate the treatment x year interaction and to point out that most of the statistical models used to assess genotype x environment interaction are also suitable for studying stability in agronomy trials.

The data is in file ***example2.dat*** , the program is in file ***example2.sas*** , and the results of the AMMI analysis and the biplot are in files ***example2.lst*** and ***example2.cgm***. All these files are in the public folder.

Table 2.1. Analysis of variance of Example 2 with adjusted means not corrected by the number of replicates

Analysis of variance with the adjusted means no corrected by the number of replicates.

General Linear Models Procedure

Dependent Variable: YLD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	239	475584990.3	1989895.4	.	.
Error	0	.	.		
Corrected Total	239	475584990.3			

R-Square	C.V.	Root MSE	YLD Mean
1.000000	0	0	7060.681

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
ENV	9	124421560.7	13824617.9	.	.
GEN	23	257991210.6	11217009.2	.	.
ENV*GEN	207	93172219.1	450107.3	.	.

Because replicates were not considered, the estimate of the error variance can't be computed, nor can the significance of the various sources of variation. The sum of squares are underestimated because they are not corrected by the number of replicates. Thus it is necessary to introduce an estimate of the error variance, its df, and the number of replicates. Table 2.2 shows the results of the complete and corrected AOV. Note that in this case the sum of squares and mean squares in tables 2.1, 2.2, and 2.3 have so many figures because the original data for grain yield was given in kg ha^{-1}. If we wish reduce the above, we must change the grain yield to ton ha^{-1}. This is also reflected in the values for mean yield, environmental and genotypic scores, and in the axes of the biplot.

Table 2.2. Analysis of variance corrected by the number of replicates

Analysis of variance corrected by the number of replicates 3

SOURCE	DF	SS	MS	F	PROB
ENV	9	373264681.98	41473853.55	164.616	0
GEN	23	773973631.74	33651027.47	133.566	0
ENV*GEN	207	279516657.30	1350322.02	5.360	0

In Table 2.3 the sum of squares for each of the AMMI terms corrected by the number of replicates are shown. The first five AMMI components are significant at the 1% level and the first three terms explain 81% of the GEI sum of squares. In Table 2.4, we are given the genotypic and environmental scores. Note that in this case, the letter S (denoting site) has been changed by Y (year).

Table 2.3. Results of the Gollob test

Results of the AMMI terms's sum of squares. 4

SSAMMI	PORCENT	PORCENAC	DFAMMI	MSAMMI	F_AMMI	PROBF
151129753.86	54.0682	54.068	31	4875153.35	19.3502	0.00000
39112401.00	13.9929	68.061	29	1348703.48	5.3532	0.00000
36781440.33	13.1589	81.220	27	1362275.57	5.4071	0.00000
20820728.73	7.4488	88.669	25	832829.15	3.3056	0.00000
11994973.75	4.2913	92.960	23	521520.60	2.0700	0.00269
7683775.66	2.7490	95.709	21	365894.08	1.4523	0.08908
6029538.76	2.1571	97.866	19	317344.15	1.2596	0.20535
3558974.45	1.2733	99.140	17	209351.44	0.8309	0.65727
2405070.77	0.8604	100.000	15	160338.05	0.6364	0.84530
0.00	0.0000	100.000	13	0.00	0.0000	1.00000

Table 2.4. AMMI scores for genotypes and years

Scores for graphing the biplot. 5

TYPE	NAME	YLD	DIM1	DIM2	DIM3
GEN	1	7181.17	-8.0747	-4.0615	-27.1773
GEN	10	7750.30	6.3515	-6.8882	-4.9073
GEN	11	7824.63	17.7153	7.8722	2.4007
GEN	12	7767.73	9.4554	-7.1920	-5.2177
GEN	13	7583.77	-1.7508	-6.7258	13.3027
GEN	14	7392.37	-0.4359	-15.1005	14.9585
GEN	15	7367.20	4.9272	-2.2907	7.1997
GEN	16	7139.83	-11.6736	-12.2338	21.2612
GEN	17	7845.30	29.6161	16.1221	-7.5017
GEN	18	7672.00	16.7630	-2.5758	2.6391
GEN	19	7963.03	25.5486	9.9490	-7.7752
GEN	2	6803.10	-13.0718	-24.5780	-15.8286
GEN	20	7550.43	17.6673	3.3123	1.6706
GEN	21	7894.73	10.6620	3.1322	4.3110
GEN	22	7496.83	8.5683	-2.1673	0.0746
GEN	23	7734.63	12.7495	-1.8089	3.5394
GEN	24	7587.40	0.2509	-18.3362	10.0035
GEN	3	6761.87	-22.4811	5.0950	-21.7229
GEN	4	5897.70	-34.1410	6.6128	-16.8657
GEN	5	5472.77	-17.3244	7.9189	5.3737
GEN	6	5421.17	-19.5412	-13.5552	1.5781
GEN	7	4744.47	-22.4810	17.8027	16.3609
GEN	8	4451.70	-25.6245	31.0141	11.9996
GEN	9	8152.20	16.3248	8.6826	-9.6767
ENV	Y88	7563.20	32.5808	18.5789	-23.0635
ENV	Y89	6833.94	-26.3110	-10.0547	-13.2911
ENV	Y90	6658.21	29.3690	18.8133	39.0622
ENV	Y91	7805.71	-4.7119	23.9634	-3.1391
ENV	Y92	6487.46	-4.3658	-29.6162	26.0184
ENV	Y93	7159.08	1.7859	-7.6727	-8.5071
ENV	Y94	8035.19	5.6147	-18.5471	-19.6100
ENV	Y95	5433.01	-35.5846	-7.7203	9.4413
ENV	Y96	7096.25	-38.7977	26.3959	-1.1857
ENV	Y97	7534.75	40.4207	-14.1404	-5.7255

Fig. 2.1 shows the biplot of this example using the adjusted means

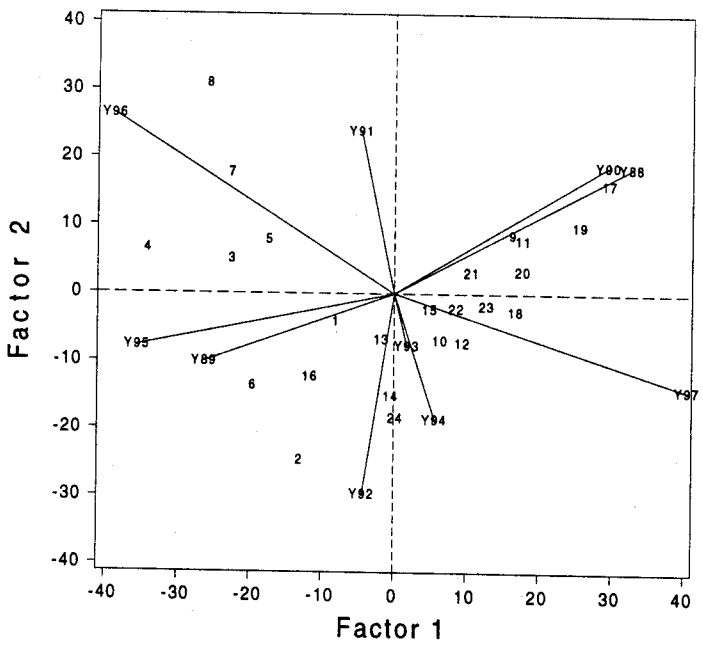

Figure 2.1. AMMI biplot of 24 treatments and 10 years using the adjusted means.

Case 2: Computing AMMI analysis with all the information and using SAS PROC GLM

For the three cases considered in this catalog, several routines of the SAS programs are the same and only limited parts of the codes are different. We will emphasize only the components of the programs that are different and specific for each case. Thus, the new commands that have been added are now shown in bold face.

First, the name of the input file is different because now all data is available and there is no need to use the codes for introducing the combined error variance, the error df, and the number of replicates because these are now directly computed.

```
options ps = 5000 ls=78 nodate;
filename biplot 'example3.cgm';
goptions device=cgmmwwc gsfname=biplot gsfmode=replace;
/*************************************************************************************
* Program to run the AMMI analysis and to use the Gollob test to determine the
significance of each AMMI term. Also the Biplot is obtained. *
 *************************************************************************************/
data raw;
 infile'example3.dat';
 input env $ rep gen $ yld;
 yld=yld/1000;
cards;

title1  'Analysis of Variance utilyzing all data ';
proc glm data=raw outstat=stats;
 class rep env gen;
 model yld = rep env gen env*gen/ss4;

*This part obtains the mse and dfe for the Gollob F-test;
```

```
data stats2;
  set stats;
  if _source_ = 'ERROR';
  sse=ss;
  dfe=df;
  mse=sse/dfe;
  keep dfe mse;
data stats3;
  set stats;
  if _source_ = 'REP';
  nrep=df+1;

* This part obtains the two-way table of residuals (GEI);
proc sort data=raw;
 by env gen;
proc means data = raw noprint;
 by env gen;
 var yld;
 output out = medias mean=yldm;
proc glm data=medias noprint;
 class env gen;
 model yldm = env gen / ss4 ;
 output out=outres r=resid;
proc sort data=outres;
 by gen env;
proc transpose data=outres out=outres2;
 by gen;
 id env;
 var resid;

* Singular value decomposition of the GEI (residual matrix) in SAS IML;
```

```
Proc IML;
use outres2;
read all into resid;
ngen=nrow(resid);
nenv=ncol(resid);
use stats2;
read var {mse} into msem;
read var {dfe} into dfem;
use stats3;
read var {nrep} into nrep;

call SVD (U,L,V,resid);      * SVD for the residual (GEI) matrix;
```

Note: From now on the codes are the same as those shown for Case 1. This specific code ends with the instruction *title1* that indicates that all information is available.

```
Title1  f=hwcgm001  h=1.0  'AMMI biplot for example 3, using all data from example
1';
run;
```

Example 3

This is the complete data set that includes 8 genotypes evaluated during 6 years in randomized complete block designs with 3 replicates ($8 \times 6 \times 3 = 144$ observations). This is in file *example3.dat.*

Results are shown in Tables 3.1-3.3. It can be seen that results are the same as those using the adjusted means (Example 1) and that the estimate combined error variance is the same as used in Example 1.

Table 3.1. Analysis of Variance for Example 3, using all data from Example 1.

General Linear Models Procedure
Class Level Information

Dependent Variable: YLD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	49	109.0396209	2.2252984	14.08	0.0001
Error	94	14.8543073	0.1580245		
Corrected Total	143	123.8939282			

R-Square	C.V.	Root MSE	YLD Mean
0.880105	5.635516	0.397523	7.053889

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
REP	2	0.20010668	0.10005334	0.63	0.5332
ENV	5	49.19976964	9.83995393	62.27	0.0001
GEN	7	42.75875311	6.10839330	38.65	0.0001
ENV*GEN	35	16.88099147	0.48231404	3.05	0.0001

Note that the sum of squares and the significances for each of the terms in the model (env, gen, envxgen) are the same as those calculated in Example 1. Similarly for the Gollob F-test for the first 3 terms.

Table 3.2 Gollob F-test for Example 3

SSAMMI	PORCENT	PORCENAC	DFAMMI	MSAMMI	F_AMMI	PROBF
7.24286	42.9054	42.905	11	0.65844	4.16671	0.00005
5.42326	32.1264	75.032	9	0.60258	3.81323	0.00039
2.96968	17.5918	92.624	7	0.42424	2.68464	0.01403
1.19062	7.0530	99.677	5	0.23812	1.50688	0.19509
0.05457	0.3233	100.000	3	0.01819	0.11511	0.95105
0.00000	0.0000	100.000	1	0.00000	0.00000	1.00000

Table 3.3 Genotypic and environmental scores for the biplot

TYPE	NAME	YLD	DIM1	DIM2	DIM3
GEN	1	6.13178	0.00956	-0.40613	0.71637
GEN	2	6.45306	-0.20785	0.11243	-0.42252
GEN	3	6.82289	0.54724	0.38688	-0.17175
GEN	4	6.97439	-0.55039	0.48510	0.32201
GEN	5	7.10122	0.79807	0.22696	0.09611
GEN	6	7.50572	-0.50527	0.31346	-0.10795
GEN	7	7.65789	-0.12223	-0.61605	-0.38304
GEN	8	7.78417	0.03088	-0.50265	-0.04922
ENV	S90	7.57075	0.26094	0.88469	-0.35836
ENV	S91	7.03321	0.14700	0.24993	0.80484
ENV	S92	6.28142	-0.39629	-0.30402	0.23481
ENV	S93	7.46617	0.91858	-0.58360	-0.19013
ENV	S94	7.69529	-0.58882	-0.25741	-0.30403
ENV	S95	6.27650	-0.34140	0.01041	-0.18713

The genotypic and environmental scores are the same as those obtained in Example 1 (using the adjusted means). However, this might not always be the case although the proportionality between scores will always remain the same. It is possible to obtain scores with the opposite sign (different polarity). Nevertheless, this does not change the general trend and associations of the elements within the biplot. This may happen for the same data set when different software for computing the singular value decomposition of a matrix are used.

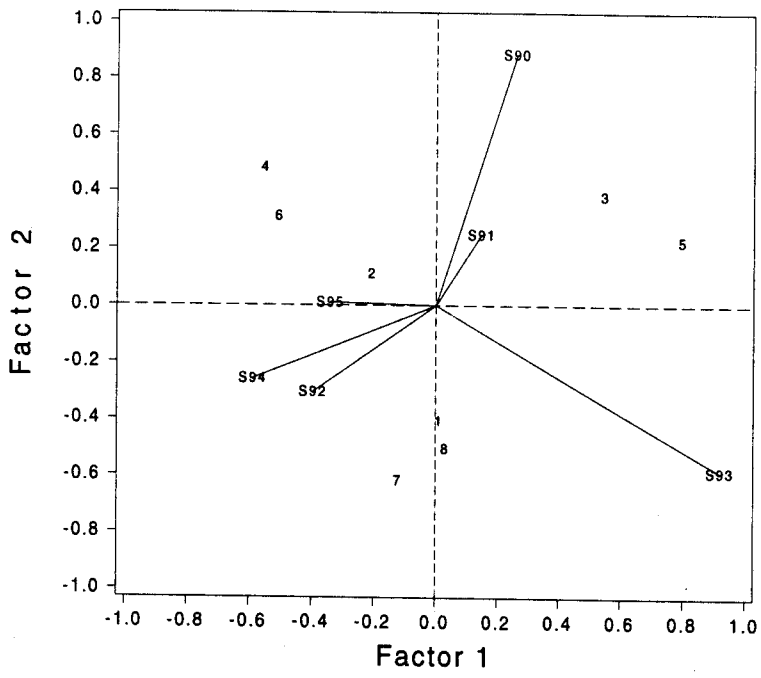

Figure 3.1. Biplot of 8 genotypes and 6 years using all information

This biplot is the same as that presented in Fig 1.1 when the adjusted means were used.

Example 4

This example is the same as Example 2 (24 agronomic practices evaluated during 10 years), but here the complete data set is used. The data file ***example4.dat***, the SAS program ***example4.sas***, the results ***example4.lst***, and biplot ***example4.cgm*** are in the public folder.

Results of the AMMI analysis and the biplot are given in Tables 4.1 - 4.3 and in Fig. 4.1, respectively

Table 4.1. Analysis of variance for the 24 treatments and 10 years using all data

Analysis of variance using all data 1

General Linear Models Procedure

Dependent Variable: YLD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	241	1428317217	5926627	23.52	0.0001
Error	478	120428814	251943		
Corrected Total	719	1548746031			

R-Square	C.V.	Root MSE	YLD Mean
0.922241	7.108937	501.9394	7060.681

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
REP	2	1562209.0	781104.5	3.10	0.0459
ENV	9	373264538.1	41473837.6	164.62	0.0001
GEN	23	773973854.5	33651037.2	133.57	0.0001
ENV*GEN	207	279516615.3	1350321.8	5.36	0.0001

Table 4.2. Results of the Gollob F-test

Sum of squares of the AMMI terms. 3

SSAMMI	PORCENT	PORCENAC	DFAMMI	MSAMMI	F_AMMI	PROBF
151129827.57	54.0683	54.068	31	4875155.73	19.3502	0.00000
39112355.84	13.9929	68.061	29	1348701.93	5.3532	0.00000
36781345.82	13.1589	81.220	27	1362272.07	5.4071	0.00000
20820767.85	7.4488	88.669	25	832830.71	3.3056	0.00000
11994953.86	4.2913	92.960	23	521519.73	2.0700	0.00269
7683755.76	2.7489	95.709	21	365893.13	1.4523	0.08908
6029543.99	2.1571	97.866	19	317344.42	1.2596	0.20535
3559002.45	1.2733	99.140	17	209353.09	0.8310	0.65726
2405062.12	0.8604	100.000	15	160337.47	0.6364	0.84530
0.00	0.0000	100.000	13	0.00	0.0000	1.00000

Table 4.3 Genotypic and environmental scores for the biplot.

Genotypic and environmental scores for the biplot. 4

TYPE	NAME	YLD	DIM1	DIM2	DIM3
GEN	1	7181.17	-8.0746	-4.0617	-27.1772
GEN	10	7750.30	6.3515	-6.8883	-4.9073
GEN	11	7824.63	17.7153	7.8722	2.4006
GEN	12	7767.73	9.4554	-7.1922	-5.2176
GEN	13	7583.77	-1.7508	-6.7256	13.3028
GEN	14	7392.37	-0.4359	-15.1003	14.9586
GEN	15	7367.20	4.9272	-2.2906	7.1998
GEN	16	7139.83	-11.6736	-12.2337	21.2613
GEN	17	7845.30	29.6161	16.1220	-7.5019
GEN	18	7672.00	16.7630	-2.5757	2.6391
GEN	19	7963.03	25.5486	9.9489	-7.7753
GEN	2	6803.10	-13.0718	-24.5781	-15.8283
GEN	20	7550.43	17.6673	3.3123	1.6706
GEN	21	7894.73	10.6621	3.1322	4.3110
GEN	22	7496.83	8.5683	-2.1674	0.0746
GEN	23	7734.63	12.7495	-1.8088	3.5395
GEN	24	7587.40	0.2509	-18.3360	10.0036
GEN	3	6761.87	-22.4812	5.0950	-21.7229
GEN	4	5897.70	-34.1410	6.6126	-16.8658
GEN	5	5472.77	-17.3244	7.9191	5.3736
GEN	6	5421.17	-19.5412	-13.5552	1.5781
GEN	7	4744.47	-22.4810	17.8029	16.3606
GEN	8	4451.70	-25.6245	31.0142	11.9993
GEN	9	8152.20	16.3248	8.6825	-9.6768
ENV	Y88	7563.19	32.5809	18.5787	-23.0636
ENV	Y89	6833.94	-26.3110	-10.0549	-13.2910
ENV	Y90	6658.21	29.3690	18.8136	39.0620
ENV	Y91	7805.71	-4.7118	23.9634	-3.1393
ENV	Y92	6487.46	-4.3659	-29.6159	26.0187
ENV	Y93	7159.08	1.7859	-7.6728	-8.5071
ENV	Y94	8035.19	5.6147	-18.5474	-19.6099
ENV	Y95	5433.01	-35.5846	-7.7201	9.4414
ENV	Y96	7096.25	-38.7977	26.3959	-1.1859
ENV	Y97	7534.75	40.4206	-14.1405	-5.7253

AMMI biplot for Example 4, using all data from Example 2

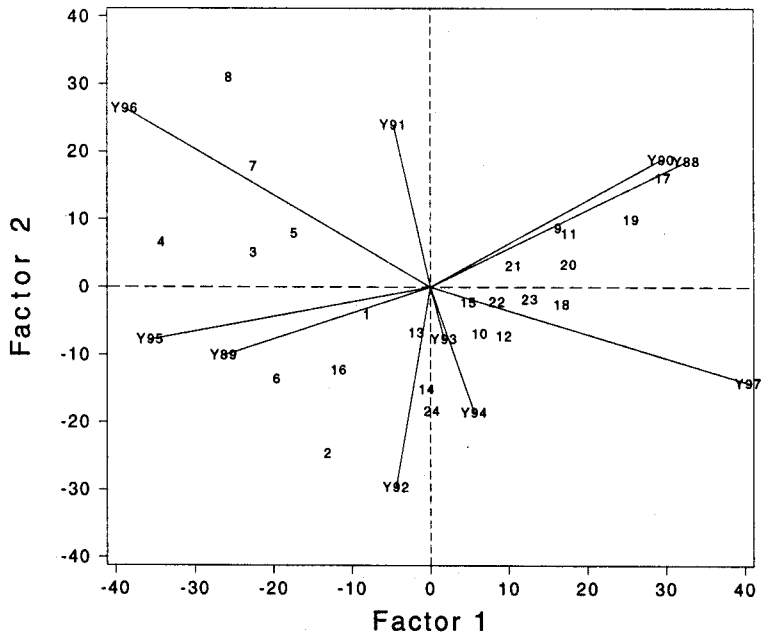

Figure 4.1. Biplot of the 24 treatments and 10 years using all data

Case 3: Computing AMMI analysis with all the information and using SAS PROC MIXED

In this case, the experimental design used for each site (environment) is an incomplete block (for example it could be alpha-lattice or a simple or a rectangular lattice). A mixed model (SAS PROC MIXED) is used because incomplete blocks are considered as random effects.

```
options ps=5000 ls=78;
filename biplot 'example5.cgm';
goptions device=cgmmwwc gsfname=biplot gsfmode=replace;
/*************************************************************************
 * Program for AMMI analysis using all data and obtaining the adjusted means
considering incomplete blocks as random effects
*************************************************************************/
data a;
 infile'example5.dat';
 input env $ rep block gen $ yld;
 cards;
proc sort;
 by env gen;
```

Individual site analyses are performed in order to obtain the adjusted means. Output files are generated using the option ***Make*** and ***Noprint*** for controlling the output of the MIXED procedure.

```
proc mixed data=a method=reml;
 by env;
 id rep block gen;
 classes gen rep block;
 model yld= rep gen;
```

```
random block(rep);
lsmeans gen ;
make 'tests'      out=test1 noprint;
make 'lsmeans'    out=ls1  noprint;
make 'covparms'   out=cov1 noprint;
make 'classlevels' out=levels noprint;
make 'fitting'    out=ajuste noprint;
make 'reml'       out=itera noprint;
```

The next section computes the combined error variance that is stored in file ***mse2***, the adjusted means are assigned to file ***medias***, the number of replicates are stored in ***nrepm***, and the df are in file ***degreem***. These data are stored in one file called ***stats***. These are all used for the Gollob F-test.

```
data errorpon;
 set cov1;
 if covparm ^= 'RESIDUAL' then delete;
proc means noprint;
 var est ;
 output out=covmeans
    mean=mse
    n=nenv;
data mse2;
 set covmeans;
 keep mse nenv;

data medias;
 set ls1;
 yld=_lsmean_;
 keep env gen yld ;
```

```
data rep;
 set test1;
 if source= 'rep';
 nrep=ndf+1;
 keep nrep;
proc means data=rep noprint;
 var nrep;
 output out = nrepm mean=nrep;

data degree;
 set test1;
 if source = 'gen' ;
 dfe=ddf;
 keep dfe;
proc means data=degree noprint;
 var dfe;
 output out = degreem sum=dfe;

data stats;
 merge mse2 nrepm degreem;
 drop _type_ _freq_;
```

Because the PROC MIXED does not present the results of the AOV in the traditional format, such as that given by PROC GLM, we will present the combined analysis using PROC GLM considering the combined intra-block AOV. This step is not necessary and can be omitted, but it is shown for the purpose of illustrating the significance level of the model's terms: env, gen, and gen×env.

```
title1  'Anayisis of variance using all data';
title2  'complete blocks are considered in PROC GLM';
proc glm data=a ;
```

```
class env gen;
model yld = env rep block(rep) gen env*gen/ss4;
```

To obtain the two-way table of residuals (matrix of GEI), a PROC GLM is used with the adjusted means previously calculated (file *medias*). These residuals are those used in the AMMI analysis and the biplot.

```
proc glm data=medias noprint;
 class env gen;
 model yld = env gen / ss4 ;
 output out=outres r=resid;
proc sort data=outres;
 by gen env;
proc transpose data=outres out=outres2;
 by gen;
 id env;
 var resid;

* Singular Value Decomposition (SVD) in SAS IML;

proc iml;
use outres2;
read all into resid;
ngen=nrow(resid);
nenv=ncol(resid);
use stats;
read var {mse} into msem;
read var {dfe} into dfem;
read var {nrep} into nrep;

call SVD (U,L,V,resid);      * SVD for the residual matrix;
```

(Note: From this part on there is no changes in the program found in the two previous cases)

The program ends with the instruction *title1* that indicates the AMMI analysis using the SAS PROC MIXED for analyzing the design in each environment.

title1 f=hwcgm001 h=1.0 'AMMI biplot for example 5, using MIXED models';
run;

Example 5

This data set includes 18 maize genotypes evaluated in 10 environment. Each experiment was arranged in a lattice design (3 incomplete blocks of size 6) with 3 replicates. The response variable is grain yield (kg/ha). The data set ***example5.dat***, the SAS program ***example5.sas***, the results ***example5.lst***, and the biplot ***example5.cgm*** are in the public folder.

Table 5.1. Analysis of variance considering the individual experiments as randomized complete block designs and using PROC GLM

Analysis of variance considering all the data and considering Randomized complete blocks using PROC GLM

Dependent Variable: YLD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	179	1447.274066	8.085330	23.62	0.0001
Error	360	123.253533	0.342371		
Corrected Total	539	1570.527599			

R-Square	C.V.	Root MSE	YLD Mean
0.921521	24.37982	0.585125	2.400037

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
ENV	9	845.1919585	93.9102176	274.29	0.0001
GEN	17	261.3598326	15.3741078	44.90	0.0001
ENV*GEN	153	340.7222748	2.2269430	6.50	0.0001

Results show that all components are highly significant; *env*, *gen*, and *envxgen* at the 1% level. However, these results are only approximated because the correct design (incomplete block) was not considered.

Table 5.2. Gollob F-test for each of the AMMI terms

Sum of squares for the AMMI terms 3

SSAMMI	PORCENT	PORCENAC	DFAMMI	DFE	MSE	MSAMMI	F_AMMI	PROBF
121.767	34.4680	34.468	25	280	0.25648	4.87070	18.9907	0.00000
108.374	30.6769	65.145	23	280	0.25648	4.71192	18.3716	0.00000
43.321	12.2626	77.407	21	280	0.25648	2.06290	8.0432	0.00000
27.052	7.6574	85.065	19	280	0.25648	1.42379	5.5513	0.00000
22.882	6.4771	91.542	17	280	0.25648	1.34600	5.2480	0.00000
12.911	3.6548	95.197	15	280	0.25648	0.86076	3.3561	0.00003
9.747	2.7590	97.956	13	280	0.25648	0.74976	2.9233	0.00053
5.206	1.4735	99.429	11	280	0.25648	0.47325	1.8452	0.04663
2.016	0.5708	100.000	9	280	0.25648	0.22404	0.8735	0.54935
0.000	0.0000	100.000	7	280	0.25648	0.00000	0.0000	1.00000

These results show that the first 7 AMMI terms are significant at the 1% level and that the next term is significant at the 5% level. It can be substantiated that the first three AMMI terms are the most important and explained 77% of the total GEI sum of squares.

Table 5.3. Genotypic and environmental scores for the biplot

Genotypic and environmental scores for the biplot. 4

TYPE	NAME	YLD	DIM1	DIM2	DIM3
GEN	1	4.10233	0.07253	1.24823	0.30563
GEN	10	1.92967	-0.10749	-0.82681	0.49048
GEN	11	2.82833	-0.43483	0.11811	-0.44020
GEN	12	3.04733	0.76254	-0.58011	-0.60342
GEN	13	1.79700	-0.50997	-0.53534	0.00195
GEN	14	1.52600	-0.69362	-0.52859	-0.00380
GEN	15	3.21833	-0.65424	0.62317	-0.34525
GEN	16	2.40867	0.08727	-0.13624	-0.89096
GEN	17	1.87800	-0.56868	0.08191	-0.07677
GEN	18	2.09667	0.82223	-0.39970	-0.05861
GEN	2	2.83367	0.01805	1.08635	-0.29390
GEN	3	2.23900	0.68856	0.42253	1.04426
GEN	4	3.07033	0.72243	0.31783	-0.30111
GEN	5	3.10500	1.13460	0.08246	-0.09708
GEN	6	1.55000	-0.56594	0.30618	0.66133
GEN	7	1.67800	0.00197	-0.53194	0.31244
GEN	8	2.02533	-0.88908	-0.09486	-0.10782
GEN	9	1.86700	0.11367	-0.65317	0.40284
ENV	S1	0.14593	-0.42213	-0.63618	0.47426
ENV	S10	3.26667	1.76935	-0.16631	-0.91405
ENV	S2	1.37870	-0.31456	0.69239	0.61882
ENV	S3	2.88704	-0.57023	1.83965	-0.32776
ENV	S4	3.95130	0.01578	-0.18036	-0.49443
ENV	S5	2.87611	-1.52696	-0.93142	-0.97047
ENV	S6	0.56389	0.31296	-0.62175	0.69062
ENV	S7	3.66889	0.34814	0.52484	-0.05254
ENV	S8	3.27056	0.27105	-0.35135	0.25602
ENV	S9	1.99130	0.11661	-0.16950	0.71953

Note that all genotypes and environments are arranged in an alphanumeric order.

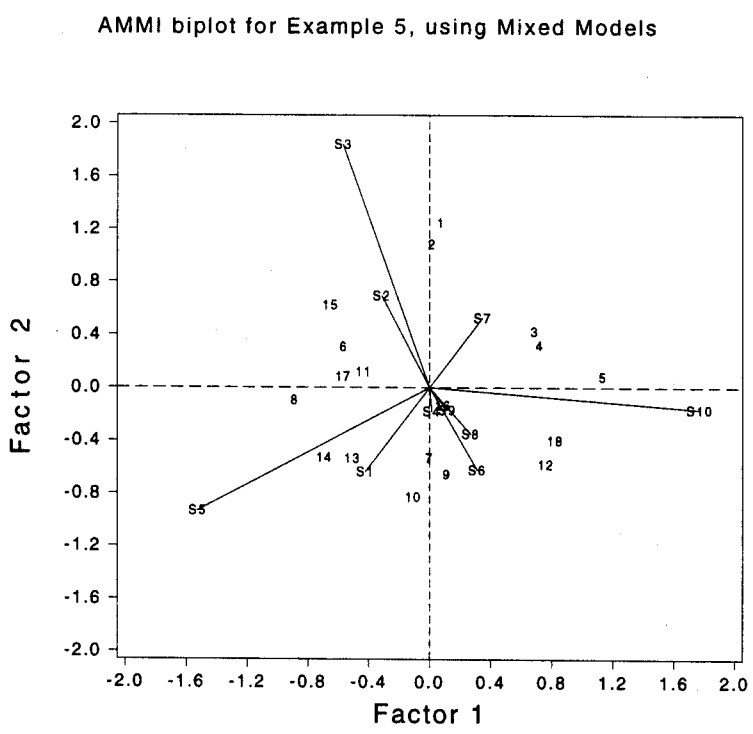

Figure 5.1. Biplot of 18 genotypes evaluated in 10 environments and analyzed as incomplete blocks using PROC MIXED.