

EL ANÁLISIS AMMI Y LA GRÁFICA DEL BIPLLOT EN SAS

Mateo Vargas Hernández

Universidad Autónoma de Chapingo

Unidad de Biometría y Estadística, CIMMYT

José Crossa

Unidad de Biometría y Estadística, CIMMYT

Marzo 2000

EL ANÁLISIS AMMI Y LA GRÁFICA DEL BIPLLOT EN SAS

Mateo Vargas Hernández

Universidad Autónoma Chapingo
Unidad de Biometría y Estadística, CIMMYT

José Crossa

Unidad de Biometría y Estadística, CIMMYT

CIMMYT, INT.

Apdo. Postal 6-641

06600 México DF

México

Marzo 2000

Contenido

Introducción	1
Caso 1. Cálculo de los modelos AMMI utilizando sólo medias ajustadas	3
Caso 2. Análisis utilizando información completa y un modelo de efectos fijos (PROC GLM)	25
Caso 3. Análisis utilizando información completa y un modelo mixto (PROC MIXED)	34
Ejemplos	
Ejemplo 1	15
Ejemplo 2	20
Ejemplo 3	28
Ejemplo 4	31
Ejemplo 5	38

EL ANÁLISIS AMMI Y LA GRÁFICA DEL BIPLLOT EN SAS

Mateo Vargas Hernández y José Crossa

Introducción

En este documento se presentan tres ejemplos de casos en los que se emplearon programas creados en SAS para obtener Modelos AMMI, así como la prueba de Gollob (Gollob, 1967) para determinar la significancia de cada término AMMI. Con estos programas, se obtiene también de manera automática el biplot derivado de los modelos AMMI. Las diferentes modalidades que aquí se presentan son:

1) Cuando se cuenta únicamente con las medias ajustadas. En este caso, las medias ajustadas son suficientes para obtener los modelos AMMI y para graficar el biplot. Sin embargo, para la prueba de Gollob se necesita tener un estimador del error del análisis combinado, los grados de libertad del error y el número de repeticiones, los cuales se deben proporcionar al programa.

2) Cuando se cuenta con toda la información, por ejemplo los valores de la variable de interés medida en cada uno de los ambientes, los genotipos y las repeticiones, pero se utiliza un análisis combinado que se considera que proviene de experimentos individuales en los que se utilizó un diseño de bloques completos. Se asume que los modelos son de efectos fijos y se analizan utilizando PROC GLM. En este caso, se obtiene de forma automática el estimador del error, los grados de libertad del error y el número de repeticiones.

3) Similar al caso anterior, pero ahora se consideran experimentos individuales en los que se utilizó un diseño de látices (bloques incompletos), y se asume que los modelos son de efectos fijos y aleatorios simultáneamente (modelos mixtos). Se utiliza PROC MIXED para la obtención de las medias ajustadas y un estimador del error ponderado de cada uno de los análisis que se hicieron por separado. Asimismo, se emplean grados de libertad y número de repeticiones también ponderados.

En los tres casos anteriores se pretende lograr los mismos objetivos:

- i) Uniformizar los conceptos y resultados de los modelos AMMI, determinar la significancia de cada término AMMI y obtener los biplots,
- ii) Que los investigadores puedan realizar lo anterior de forma simplificada y automatizada.

En términos generales, los tres casos presentan el mismo programa base, con ligeras adecuaciones en algunas secciones, dependiendo del caso que se esté estudiando. Se describirá brevemente cada una de las secciones que conforman los programas.

Todos los archivos que contienen los datos, programas, salidas y gráficos que se mencionan en este manual se pueden obtener en la siguiente dirección de Internet:

www.cimmyt.cgiar.org/biometrics. Aquí seleccione el archivo AMMI.EXE.

Caso 1

Cálculo de los modelos AMMI utilizando sólo medias ajustadas

Para este caso es necesario proporcionar al programa la información correspondiente al valor del estimador del error combinado, los grados de libertad de dicho error y el número de repeticiones. Como notación se utilizarán letras itálicas, que representarán las líneas de programación o instrucciones, y letra normal para el texto explicativo.

En la primera línea se determinan las opciones para la longitud de página (5000 líneas), longitud de línea (78 caracteres), y se indica que no incluya la fecha de los análisis.

```
options ps = 5000 ls=78 nodate;
```

Las siguientes dos líneas se utilizan para generar la gráfica del biplot en formato CGM (Graphics Metafile), que se pueda exportar a Microsoft Word y/o PowerPoint, asignándole el nombre "*ejemplo1.cgm*". Para esto, se emplea la opción *device=cgmmwwc* y el modo de reemplazo, de manera que cada vez que se ejecute el mismo programa el gráfico sea sustituido (*gsfmode=replace*). Otra forma sería que se agreguen los gráficos cada vez que se ejecute el programa utilizando la opción *gsfmode=append*.

```
filename biplot 'ejemplo1.cgm';
```

```
goptions device=cgmmwwc gsfname=biplot gsfmode=replace;
```

Nota: Si se desea visualizar el gráfico en el monitor de la PC, se deben cancelar estas dos líneas utilizando un símbolo de asterisco (*) al inicio de cada una.

En la siguiente sección se describe la forma clásica de leer un archivo de datos externo llamado "*ejemplo1.dat*", en el que se establecen el orden, el nombre y el tipo de las variables. En este caso es importante indicar que la variable que identifica tanto a los

ambientes (*env*) como a los genotipos (*gen*) es de tipo alfanumérico; esto puede lograrse mediante el uso del símbolo de pesos (\$) después del nombre.

```
data raw;  
infile'ejemplo1.dat';  
input env $ gen $ yld;  
cards;
```

Una vez efectuada la lectura de los datos, se procede a realizar un **Análisis de Varianza** con las medias ajustadas. En el análisis de varianza no se puede calcular un estimador del error y, por lo tanto, no se muestran valores de significancia. Asimismo, los valores de las sumas de cuadrados no están corregidos por el número de repeticiones. Para obtener los valores corregidos por número de repeticiones, así como los valores de significancia asociados con éstas, en esta sección deben indicarse algún valor estimado para el error, los grados de libertad del error y el número de repeticiones (*mse*, *dfe*, y *nrep*, respectivamente).

```
title1 'Análisis de Varianza con las medias ajustadas';  
title2 'no corregida por número de repeticiones';  
proc glm data=raw outstat=stats;  
class env gen;  
model yld = env gen env*gen / ss4;
```

Nota: Los resultados del Ejemplo 1 se muestran en el Cuadro 1.1.

```
title1 'análisis de varianza corregida por número de repeticiones';  
data stats2;  
set stats ;  
drop _name_ _type_ ;  
if _source_ = 'ERROR' then delete;  
mse=251943; * aquí se deben establecer los valores del error;  
dfe=478; * grados de libertad del error
```

```

nrep=3;          * y número de repeticiones
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;

```

Nota: Los resultados del ejemplo 1 se muestran en el Cuadro 1.2.

Los valores del error, los grados de libertad del error y el número de repeticiones se guardan y se utilizan en una etapa posterior, para determinar mediante la prueba de Gollob la significancia de cada uno de los términos AMMI.

En la siguiente sección de instrucciones se describe cómo obtener los residuales de la Interacción Genotipo por Ambiente (GEI). Será suficiente modificar el modelo para obtener residuales incluyendo el efecto de genotipos (G+GEI) y realizar las modificaciones respectivas en los procedimientos Sort y Transpose.

```

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;

```

Con la información del cuadro de doble entrada de residuales que se guardó en el archivo *outres2*, se procede a efectuar la **Descomposición del Valor Singular** en IML para calcular el AMMI. La suma de cuadrados de cada término AMMI corresponde simplemente el cuadrado de cada uno de los valores singulares. La proporción de variabilidad explicada por cada término AMMI es el cociente de la suma de cuadrados explicada por cada término AMMI en relación con la suma de cuadrados total de la Interacción Genotipo por Ambiente.

```

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 para la matriz de residuales;
  ss=( L ## 2 ) * nrep;      * sumas de cuadrados para los términos AMMI;
  suma=sum(ss);
  percent=((1/suma) # ss)*100;  * cálculo de proporción de variabilidad;
  mínimo=min(ngen,nenv);

```

El siguiente ciclo sirve para calcular los grados de libertad asociados con cada término AMMI y la proporción de variabilidad acumulada hasta el *i*-ésimo término respectivo.

```

  porcenta=0;
  do i = 1 to mínimo;

```

```

df=(ngenv-1)+(nenv-1)-(2*i-1);
dfa=dfa/df;
porceacu=percent[i,];
porcenta=porcenta+porceacu;
porcenac=porcenac/porcenta;
end;

```

Las siguientes instrucciones se emplean para generar matrices del orden apropiado que contengan los valores de los grados de libertad del error (*dfem*) y del cuadrado medio del error (*msem*), de forma tal que sean compatibles con los resultados calculados para las sumas de cuadrados (*ss*), con la proporción de variabilidad explicada por cada término AMMI (*percent*), con la proporción de variabilidad acumulada (*porcenac*) y con los grados de libertad de cada término AMMI (*dfa*), los cuales se reúnen en una sola matriz denominada *ssdf*.

```

dfe=J(mínimo, 1, dfem);
mse=J(mínimo, 1, msem);
ssdf=ss||percent||porcenac||dfa||dfe||mse;

```

Posteriormente se procede a la obtención de los **scores** para los genotipos y los ambientes, utilizando como ponderación la raíz cuadrada de los valores singulares. Se calculan sólo los scores para los tres primeros términos AMMI, pero se puede extender fácilmente a cualquier número de términos, o restringirlo sólo a los dos primeros haciendo los ajustes pertinentes. Los scores se guardan en la matriz *scores*. Estos valores serán utilizados posteriormente para la graficación del biplot.

```

L12 = L ## 0.5;
scoreg1=U[,1]# L12[1,]; * Scores genotípicos para el primer término;
scoreg2=U[,2]# L12[2,];
scoreg3=U[,3]# L12[3,];
scoree1=V[,1]# L12[1,]; * Scores ambientales para el primer término;

```

```

scoree2=V[,2]# L12[2,];
scoree3=V[,3]# L12[3,];
scoreg=scoreg1||scoreg2||scoreg3;
scoree=scoree1||scoree2||scoree3;
scores=scoreg/scoree;

```

Una vez que se obtienen los resultados del Análisis AMMI en IML de SAS, se generan archivos que contengan la información necesaria para el proceso de Data Step en SAS, a fin de realizar la prueba de Gollob y obtener el biplot correspondiente.

```

create sumas from ssdf;          * creación de archivos para utilizarlos ;
append from ssdf;              * posteriormente en la prueba de Gollob ;
close sumas;                   * y en el biplot ;
create scores from scores;
append from scores ;
close scores;

```

Nota: Aquí termina IML y continúa Data Step de SAS;

En la siguiente sección de instrucciones se realiza la prueba de Gollob para determinar la significancia de cada término AMMI. Los resultados se imprimen en el siguiente orden: Suma de cuadrados de cada término, proporción de variabilidad explicada por cada término, proporción de variabilidad acumulada, grados de libertad, cuadrado medio, valor F y nivel de probabilidad asociado a cada uno de los términos. Se puede modificar la forma de salida de esta información haciendo los cambios pertinentes en la instrucción *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_ammii=msammi/mse;
probf=1-probf(f_ammii,dfammi,dfe);
title1 'Resultados de las Sumas de Cuadrados para los términos AMMI';
proc print data=ssammi noobs;
    var ssammi percent porcenac dfammi msammi f_ammii probf;

```

Nota: Los resultados del ejemplo 1 se muestran en el Cuadro 1.3.

Para **obtener el biplot** en forma automatizada, se requiere generar información relacionada con los nombres que se asignarán a cada uno de los ambientes y/o genotipos, así como identificar el tipo de variable: ambiente (*ENV*) o genotipo (*GEM*). Los valores correspondientes se guardan en los archivos *namegen* y *nameenv*, que posteriormente se combinan en un solo archivo denominado *typename*. También debe obtenerse el rendimiento promedio para cada genotipo/ambiente.

** inicia generación de información para el biplot;*

** para obtener las variables type y name para los genotipos;*

```

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;
```

** para obtener las variables type y name para los ambientes;*

```
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;          * contiene type y name para genotipos y ambientes;  
  set namegen nameenv;
```

Con la información del tipo (type), nombre (name) y rendimiento promedio de las variables ambientes y genotipos, se genera un archivo que además contenga los valores de los scores calculados en una etapa anterior, salvados en el archivo llamado *scores*. Con este nuevo archivo se inicia la etapa de graficación del biplot.

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

```

dim3=col3;
drop col1-col3;

title1 'Resultados de los scores para la graficación del biplot';
proc print data=biplot noobs;
var type name yld dim1 dim2 dim3;

```

Nota:, Los resultados del ejemplo 1 se muestran en el Cuadro 1.4.

Entre las ventajas que ofrece el módulo para la generación de gráficas en SAS (SAS/GRAPH) se encuentra la de poder controlar las características principales de las gráficas, como la asignación de variables a cada uno de los ejes coordenados, los *fonts* (tipo, tamaño, color) para las etiquetas, los títulos de los ejes, los títulos principales, las leyendas, etc., que se guardan en un archivo tipo “*Annotate*” denominado *labels*. en este ejemplo. De esta forma, se puede usar la función *LABEL* para asignar etiquetas a cada una de las coordenadas de ambientes y genotipos, y las funciones *MOVE* y *DRAW* para dibujar los vectores desde el origen hasta la coordenada de los ambientes (o genotipos si se prefiere).

```

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;

```

Nota: Las instrucciones *if dim1 >=0 then position='5'; else position='5';* deben modificarse si, además de la etiqueta para cada coordenada, se desea graficar también un símbolo diferente asociado con la variable ambiente o genotipo. En este caso, el símbolo se dibuja en el centro de la coordenada; la etiqueta se mueve hacia la derecha si los valores de las coordenadas son mayores que 0.0, y hacia la izquierda si son menores

que 0.0. Se recomienda modificar las instrucciones anteriores sustituyendo los nuevos valores: *if dim1 >=0 then position='6'; else position='4';*. Para mayor información sobre otras posibles posiciones de las etiquetas, se recomienda consultar el Manual de Usuarios para SAS GRAPH (SAS / GRAPH, *User's Guide*).

Ahora que se cuenta con toda la información necesaria para obtener el biplot, se empleará el procedimiento GPLOT de SAS. En este caso se empleará como variable el segundo término AMMI (*dim2*) en el eje de las ordenadas (eje Y), y el primer término AMMI (*dim1*) en el eje de las abscisas (eje X). Se pueden realizar las modificaciones que sean necesarias para incluir biplots del tercer término AMMI con respecto al primer y segundo términos. Como archivo tipo "*annotate*" se empleará el archivo generado en la etapa anterior (*labels*), al cual se agrega una instrucción para que dibuje un marco en el gráfico (*frame*); que dibuje líneas de referencia vertical (*vref=0.0*) y horizontal (*href=0.0*) que pasen por el origen. También se determina el color (*cvref, chref*) y tipo (*lvref, lhref*) de las líneas de referencia y se indica cuáles son los ejes verticales (*vaxis*) y horizontales (*haxis*), para manejarlos por separado y tener un mejor control de las características asociadas con los ejes. Finalmente, se indica que no incluya leyendas (*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, datos ejemplo 1, medias ajustadas';
```

```
run;
```

Nótese que en este caso la palabra *symbol* restringe el uso de dicha opción mediante la instrucción *v=none* (*value=none*). Si se desea incluir símbolos, puede modificarse la instrucción, por ejemplo *v=square*, *v=circle*, *v=triangle*, *v=diamond*, *v='*'*, *v='+'*, *v='A'*, *v='B'*, etc. Para mayor detalle, consúltese el Manual de Usuarios de SAS GRAPH. Igualmente se pueden modificar los colores, tamaños, etc. para los símbolos.

Las instrucciones *Axis1* y *Axis2* se presentan con las opciones más comunes, como utilizar: *length = 6.0 in*, para indicar la longitud de los ejes en pulgadas;

order = (-40 to 40 by 10), para controlar los valores mínimo y máximo, así como aumentos en los ejes. Durante una primera ejecución del programa puede cancelarse esta instrucción, de manera que SAS genere los valores por *default*. Si es necesario ajustarlos, podrá efectuarse modificando la opción en una segunda ejecución del programa;

label = (f = hwcgm001 h = 1.2 a=90 r=0 'Factor 2'), para especificar el *font*, tamaño, ángulo, rotación y título del eje respectivo;

value=(h=0.8), para controlar el tamaño de las marcas en los ejes; y

minor=none; para indicar que no se desean divisiones menores en las marcas de los ejes.

Para mayor detalle sobre estas opciones, se recomienda consultar el Manual de Usuarios de SAS GRAPH.

Por último, la instrucción *title1* controla el *font*, el tamaño, el color y la etiqueta del título principal de la gráfica. Se pueden emplear dos o más títulos al mismo tiempo.

Un ejemplo de este tipo de biplots se presenta en la Figura 1.1.

Notas adicionales:

Si se desea obtener una gráfica AMMI diferente, en la que utilice el rendimiento promedio de cada genotipo y ambiente en el eje de las abscisas (eje X) *versus* los valores de los scores para el primer término AMMI en el eje de las ordenadas (eje Y), se puede reemplazar la instrucción *plot dim2*dim1* por *plot dim1*yld* y adecuar o cancelar las opciones asociadas con las líneas verticales y horizontales de referencia (*vref*, *href*, *cvref*, *chref*, *lvref*, *lhref*, etc). Un ejemplo de este tipo de biplot se presenta en la Figura 1.2.

EJEMPLOS

Ejemplo 1

El siguiente ejemplo trata de un experimento con 8 genotipos de trigos harineros evaluados durante 6 años (1990-1995) en Ciudad Obregón, México. En cada uno de los años, los experimentos consistieron en un diseño de bloques completos al azar con tres repeticiones. Los genotipos que se evaluaron provenían de un conjunto histórico: los primeros habían sido liberados en la década de los 60; y los últimos, en la de los 80. El orden de los números del 1 al 8 corresponde al orden de liberación de los genotipos con base en el tiempo (Sayre et. al., 1997).

Los valores promedio de las tres repeticiones se presentan en el archivo llamado ***ejemplo1.dat***.

Para analizar esta información se ejecuta el programa *ejemplo1.sas* de SAS, en el que es necesario introducir los siguientes valores para el estimador del cuadrado medio del error (0.1580245), los grados de libertad del error (94) y el número de repeticiones (3). Los valores del cuadrado medio del error se obtienen al realizar un análisis de varianza con los datos completos, considerándolo como bloques completos al azar y utilizando los valores de la variable respuesta transformada a toneladas por hectárea. Estos datos se obtendrán en el Ejemplo 3, en el que se expone el caso en que se emplea toda la información.

Los resultados se presentan en un archivo denominado *ejemplo1.lst*. Entre estos resultados se incluyen las siguientes secciones.

Cuadro 1.1. Análisis de varianza con las medias ajustadas no corregidas por número de repeticiones.

Análisis de varianza con las medias ajustadas
no corregida por número de repeticiones

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	.	.

Nótese que en este análisis de varianza los grados de libertad del total son sólo 47 (48 observaciones = 6 años × 8 genotipos, no repeticiones). No se puede estimar el cuadrado medio del error y por lo tanto no se presentan valores de significancia para ninguno de los términos del modelo: env, gen y envxgen. Tampoco las sumas de cuadrados están corregidas por el número de repeticiones.

En el Cuadro 1.2 se presentan las sumas de cuadrados ya corregidas por el número de repeticiones. En este punto se lleva a cabo el análisis de varianza completo, con los niveles de significancia asociados a cada término del modelo, en función del estimador del cuadrado medio del error, así como de sus grados de libertad, los cuales se proporcionaron al programa.

Cuadro 1.2. Análisis de varianza corregido por número de repeticiones.

Análisis de varianza corregida por número de repeticiones					3
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

En el Cuadro 1.3 se presentan los resultados de la prueba de Gollob (Gollob, 1967) mediante la cual se determina la significancia de cada uno de los términos AMMI. En la primera columna se encuentra la suma de cuadrados corregida por el número de repeticiones (SSAMMI). La segunda columna muestra el porcentaje de la suma de cuadrados total de la interacción genxenv explicada por cada uno de los términos AMMI (PORCENT); y en la tercera columna se dan estos mismos valores, pero acumulados hasta el término respectivo (PORCENAC). En las siguientes columnas se proporcionan los valores de los grados de libertad (DFAMMI), el cuadrado medio (MSAMMI), el valor de F (F_AMMI), así como el valor de probabilidad asociado con la prueba de F para cada uno de los términos AMMI, respectivamente. En este caso se obtiene que los tres primeros términos AMMI son significativos en un nivel del 1%.

Se ha demostrado, por medio de simulación, que la prueba de Gollob es muy flexible y puede dar como resultado que muchos términos AMMI sean considerados

significativos (Cornellius et. al., 1996). Esta prueba no controla el error tipo I que puede llegar a valores tan altos como el 60%.

Cuadro 1.3. Resultados de la prueba de Gollob para los términos AMMI.

Resultados de las sumas de cuadrados para los términos AMMI							4
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	

Por último, en el Cuadro 1.4 se muestran los resultados del rendimiento promedio (YLD) y los *scores* genotípicos y ambientales para los tres primeros términos AMMI (DIM1, DIM2, DIM3), así como los valores de las variables que sirven para generar el gráfico del biplot (TYPE y NAME). En la variable NAME se encuentran las etiquetas que se utilizarán en el gráfico de las coordenadas correspondientes a cada uno de los genotipos y ambientes. El biplot de este ejemplo se muestra en la Figura 1.1 y también en el archivo *ejemplo1.cgm*, que se genera de forma automática y puede exportarse sin ninguna dificultad a documentos de Microsoft *Word* o *PowerPoint*.

Cuadro 1.4. Información para la gráfica del biplot.

Resultados de los <i>scores</i> para la graficación del biplot						5
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	

AMMI biplot, datos Ejemplo 1, medias ajustadas

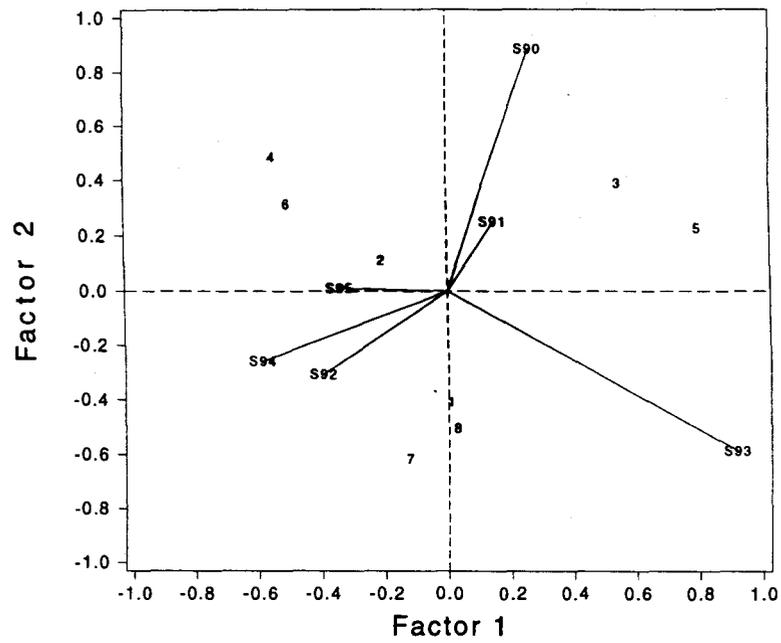


Figura 1.1. Gráfica biplot del ejemplo 1, en el que se emplearon la medias ajustadas.

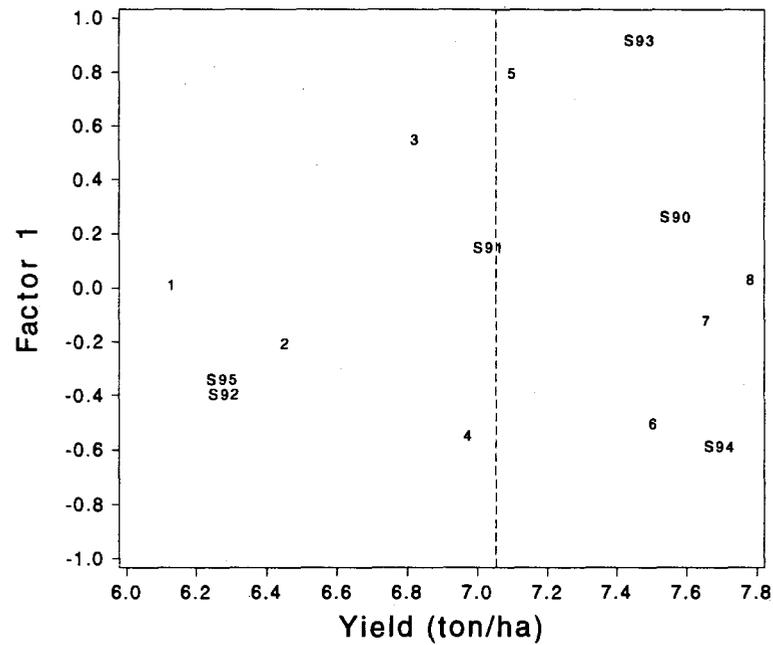


Figura 1.2. Gráfica AMMI alternativa para los datos del ejemplo 1.

En la gráfica alternativa se muestran los genotipos y/o años cuyo rendimiento promedio fue superior al valor de la media general (línea de referencia vertical). Se puede ver claramente cómo los genotipos de liberación temprana (1, 2, 3, y 4) presentaron valores menores que el promedio general, mientras que los genotipos de liberación posterior (5, 6, 7, y 8) presentaron valores promedio mayores que el promedio general. Asimismo, puede visualizarse directamente cuáles fueron los años en que se obtuvieron mejores rendimientos. El programa se encuentra en el archivo denominado ***alterna1.sas***, y la gráfica, en el archivo ***alterna1.cgm***.

Ejemplo 2

En este ejemplo se emplea un conjunto de datos obtenido de un experimento de trigo en el que se incluyeron varios tratamientos para las prácticas de cultivo. Dicho experimento se realizó durante un periodo de 10 años (1988-1997) en Ciudad Obregón, México. Cada año, los experimentos fueron dispuestos en un diseño de bloques completos al azar con tres repeticiones. Los tratamientos se obtuvieron de la combinación de cuatro factores en los siguientes niveles: labranza a 2 niveles (1=sin subsoleo profundo, 2=con subsoleo profundo); cultivo de verano a dos niveles (1=sesbania (*Sesbania spp*), 2=soya (*Glycine max L.*)); estiércol a 2 niveles (1=con estiércol de gallina, 2=sin estiércol de gallina); y dosis de fertilización nitrogenada a 3 niveles (1=0 kg N ha⁻¹, 2=100 kg N ha⁻¹ y 3=200 kg N ha⁻¹). En total se obtuvieron 2x2x2x3=24 tratamientos. Por tanto, utilizando la notación normal para experimentos factoriales, el tratamiento 1 se denota como [1-1-1-1], el tratamiento 2 como [2-1-1-1], el tratamiento 3 como [1-2-1-1], y así sucesivamente hasta llegar al tratamiento 24 [2-2-2-3].

Aun cuando este conjunto de datos no está estrictamente relacionado con la interacción genotipo x ambiente, se consideró conveniente incluirlo, ya que muestra la forma en que las metodologías pueden adaptarse y emplearse en diversas situaciones. En este caso se tendría, entonces, una interacción tratamiento x ambiente.

Los datos se encuentran en el archivo ***ejemplo2.dat*** del sitio de Internet que se mencionó en el ejemplo anterior y consisten en los valores de las medias ajustadas. El

archivo *ejemplo2.sas* contiene el programa; los resultados del análisis AMMI, así como el biplot se encuentran en los archivos *ejemplo2.lst* y *ejemplo2.cgm*, respectivamente.

A continuación se muestran los resultados:

Cuadro 2.1. Análisis de varianza con las medias ajustadas no corregida por número de repeticiones.

Análisis de varianza con las medias ajustadas no corregida por número de repeticiones						1
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	.	.	

Una vez más resulta obvio que al no existir repeticiones tampoco se tiene un estimador del error y, por consiguiente, no se puede determinar el nivel de significancia de los términos del modelo. De manera similar, las sumas de cuadrados calculadas están subestimando los verdaderos valores ya que no se encuentran corregidas por el número de repeticiones. Dado lo anterior, fue necesario proporcionar los valores del estimador del error, los grados de libertad y el número de repeticiones. El Cuadro 2.2 muestra los resultados del análisis de varianza, ya corregido por el número de repeticiones, y con un nivel de significancia asignado a cada uno de los términos del modelo.

Cuadro 2.2. Análisis de varianza corregido por el número de repeticiones.

Análisis de varianza corregida por número de repeticiones						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	

En el Cuadro 2.3 se muestran los resultados de la sumas de cuadrados de cada uno de los términos AMMI, ya corregidas por el número de repeticiones. En este caso puede observarse que los primeros 5 componentes AMMI son significativos en un nivel del 1%, mientras que, por otro lado, se observa que los tres primeros términos explican una proporción del 81% de la suma de cuadrados total de la interacción tratamiento x ambiente. En el Cuadro 2.4 se muestran los valores de los scores para tratamientos y ambientes (años). Nótese que en este caso se cambió el nombre de los ambientes en SAS: la letra S (de sitio) por la letra Y (de *year*). Esta es una tarea muy sencilla que el usuario puede realizar según su criterio.

Cuadro 2.3. Resultados de la prueba de Gollob.

Resultados de las sumas de cuadrados para los términos AMMI							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	

Nota: En este ejemplo, las sumas de cuadrados y cuadrados medios en los Cuadros 2.1, 2.2 y 2.3 tienen muchos dígitos dado que los valores originales para la variable rendimiento en grano se expresaron en kg ha^{-1} . Si se desea reducir el número de dígitos, se deben transformar los valores de rendimiento en ton ha^{-1} . Los cambios que se

realicen se reflejarán también en los valores para el rendimiento promedio, los scores genotípicos y ambientales (del Cuadro 2.4) y los ejes del biplot.

Cuadro 2.4. Resultados de los scores para obtener el biplot.

Resultados de los scores para la graficación del 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	

En la Figura 2.1 se muestra el biplot de este ejemplo utilizando las medias ajustadas.

AMMI biplot, datos Ejemplo 2, medias ajustadas

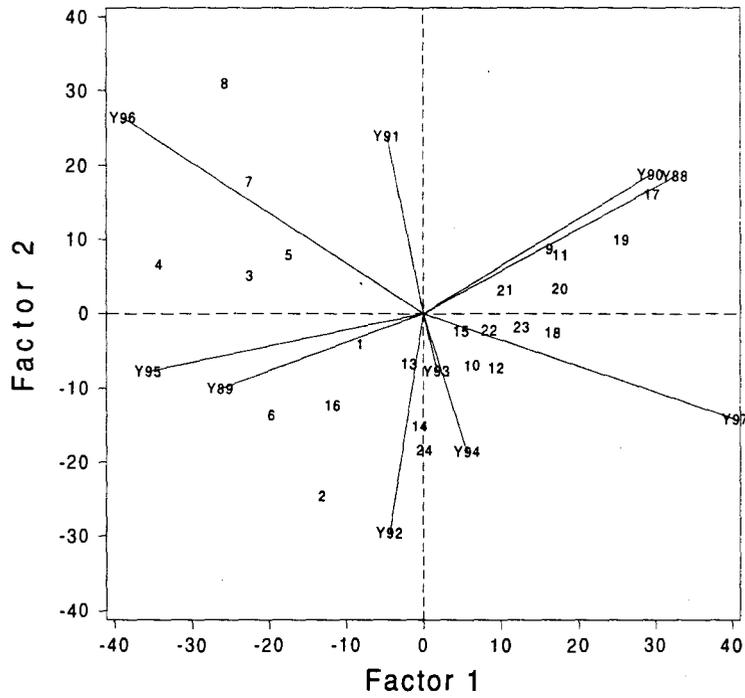


Figura 2.1. Biplot AMMI para 24 tratamientos en un experimento realizado durante 10 años utilizando las medias ajustadas.

Caso 2

Análisis utilizando información completa y un modelo de efectos fijos (PROC GLM)

Como se mencionó en la introducción de este documento, la mayoría de las rutinas y ciclos de programación para los tres casos que se tratan aquí son similares, y sólo en algunas secciones breves se encuentran diferencias específicas de cada caso. En lo sucesivo, se hará mayor énfasis en aquellas instrucciones que difieren de casos que ya se examinaron detalladamente en secciones anteriores. Para las nuevas instrucciones se emplean negritas e itálicas. Donde no se mencione nada al respecto, se debe entender que el tema ya fue explicado.

Obviamente, se cambiará el nombre del archivo de entrada de datos, ya que ahora se cuenta con toda la información para cada ambiente, genotipo y repetición. No es necesario proporcionar los estimadores del error, grados de libertad o número de repeticiones, puesto que el programa los calcula directamente de los datos y los salva para utilizarlos después, cuando se requieran.

```
options ps = 5000 ls=78 nodate;
filename biplot 'ejemplo3.cgm';
goptions device=cgmmwwc gsfname=biplot gsfname=replace;
/*****
* Programa para calcular AMMI, utiliza la prueba de Gollob para determinar la *
* significancia de cada término AMMI, se utiliza toda la información *
* También se obtiene el biplot de AMMI de forma automática *
*****/
data raw;
infile 'ejemplo3.dat';
input env $ rep gen $ yld;
yld=yld/1000;
cards;
```

```

title1 'Análisis de Varianza utilizando todos los datos';
proc glm data=raw outstat=stats;
  class rep env gen;
  model yld = rep env gen env*gen / ss4;

```

**Para obtener el mse y dfe con toda la información para la prueba de Gollob;*

```

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;

```

** Para obtener los residuales en términos de medias de genotipos y ambientes;*

```

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;

```

* **Descomposición del Valor Singular en IML para calcular el AMMI;**

```

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

call SVD (U,L,V,resid);      * SVD para la matriz de residuales;

```

Nota: Como a partir de esta instrucción el resto del programa es básicamente similar al del caso anterior, ya no se ampliará la explicación. No obstante, en los ejemplos que se encuentran en el sitio de internet sí se incluyen los programas completos. El programa termina con la instrucción *title1* que indica que en este proceso se utiliza toda la información.

```

Title1 f=hwcm001 h=1.0 'AMMI biplot, incluye todos los datos del ejemplo 1';
run;

```

Ejemplo 3

Este conjunto de datos corresponde al experimento completo de 8 genotipos con 3 repeticiones evaluados durante 6 años ($8 \times 6 \times 3 = 144$ observaciones) y se encuentra en el archivo *ejemplo3.dat*.

Los resultados se irán presentando en las siguientes secciones y se demostrará que existe una completa correspondencia con los resultados obtenidos utilizando sólo las medias ajustadas (Ejemplo 1). En particular nótese que el valor del estimador del cuadrado medio del error y los grados de libertad corresponden exactamente a los valores que se proporcionaron al programa para el caso en que se usan únicamente los valores de las medias ajustadas.

Cuadro 3.1. Análisis de varianza utilizando todos los datos.

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

Asimismo, nótese que las sumas de cuadrados y los niveles de significancia para cada uno de los términos del modelo (env, gen, envxgen) corresponden a los mismos

valores reportados en el caso en que se utilizan sólo las medias ajustadas. De manera similar, para el caso de la prueba de Gollob se muestra ahora que los tres primeros términos AMMI son significativos en el nivel del 1%.

Cuadro 3.2. Resultados de la prueba de Gollob.

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

Cuadro 3.3. Resultados de los scores para la graficación del 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

Para el caso de los *scores* genotípicos y ambientales se puede observar que son exactamente los mismos valores que se obtuvieron en el ejemplo 1 (utilizando sólo los valores de las medias). Sin embargo, debe aclararse que esto no sucede siempre, puesto que habrá casos en que los valores de los *scores* no serán exactamente iguales, pero sí proporcionales, es decir, no se afectará la relación geométrica entre los elementos, lo cual se verifica de manera sencilla comparando los biplots respectivos. Es posible que para otros conjuntos de datos diferentes de este ejemplo, se obtengan resultados de los *scores* con signo contrario para el caso en que o se utilizan todos los datos o solo las

medias. Existe la posibilidad de que esto suceda incluso en un mismo conjunto de datos cuando se utiliza un *software* diferente para efectuar la descomposición del valor singular de una matriz; esto se debe a que en términos matemáticos dicha descomposición es única, excepto por los signos.

AMMI Biplot, incluye todos los datos del Ejemplo 1

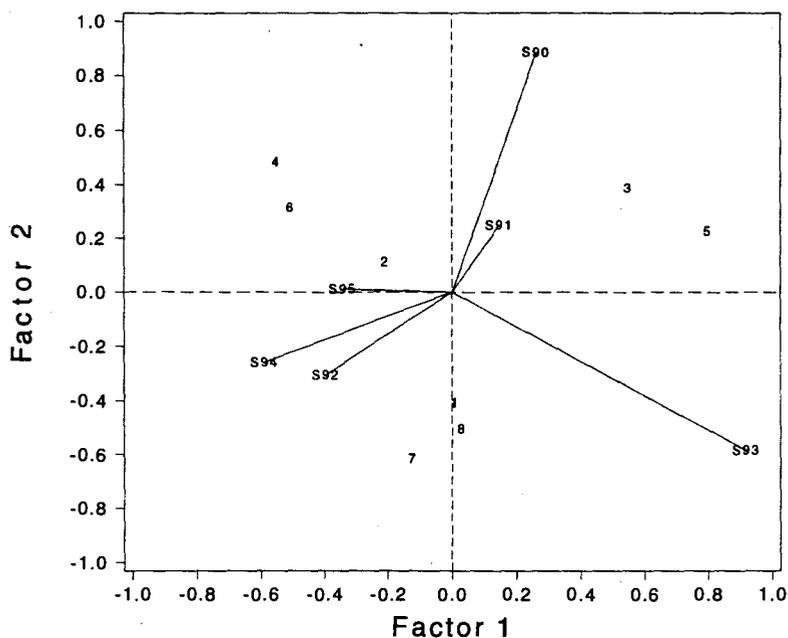


Figura 3.1. Biplot de 8 genotipos y seis años utilizando toda la información.

Se puede verificar de manera sencilla que este biplot es igual al de la Figura 1.1, en el que se utilizaron sólo las medias para realizar el análisis AMMI.

Ejemplo 4

En este ejemplo se trabaja con un conjunto de datos completos correspondiente a los experimentos de prácticas de cultivo que se efectuaron durante 10 años en Ciudad Obregón, México. Esta es la versión completa del Ejemplo 2. Los usuarios interesados encontrarán los datos en el archivo *ejemplo4.dat* en el sitio de Internet.

El programa que realiza los análisis se encuentra en el archivo *ejemplo4.sas*; los resultados correspondientes se muestran en el archivo *ejemplo4.lst*, y el biplot respectivo, en el archivo *ejemplo4.cgm*.

A continuación se muestran tanto los resultados como el biplot, para demostrar que efectivamente se obtienen los mismos resultados igual que cuando sólo se emplean los valores de las medias ajustadas.

Cuadro 4.1. Análisis de varianza para los 24 tratamientos en 10 años utilizando todos los datos.

Análisis de varianza utilizando todos los datos						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	

Cuadro 4.2. Resultados de la prueba de Gollob.

Resultados de las sumas de cuadrados para los términos AMMI							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	

Cuadro 4.3. Resultados de los scores para obtener el biplot.

Resultados de los scores para la graficación del 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, incluye todos los datos del Ejemplo 2

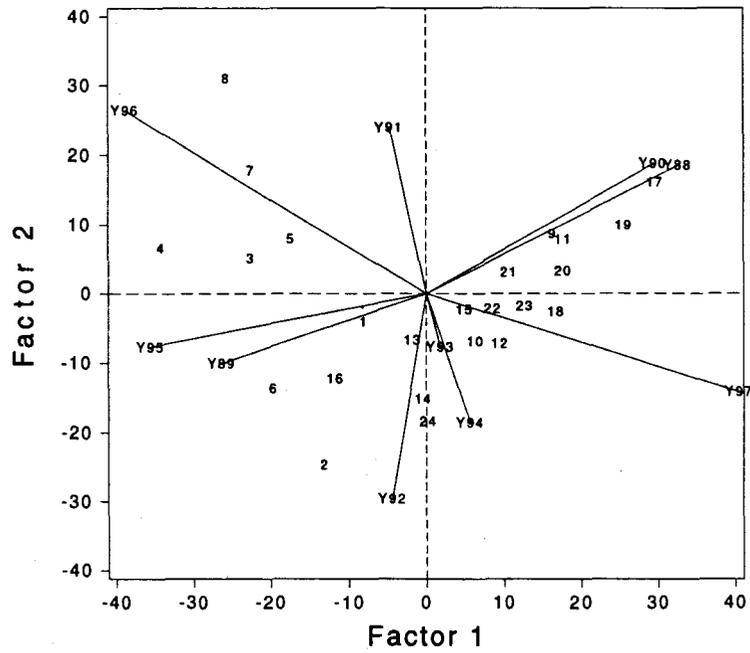


Figura 4.1. Biplot de 24 tratamientos en 10 años utilizando todos los datos.

Caso 3

Análisis utilizando información completa y un modelo mixto (PROC MIXED)

El procedimiento que se describe para este caso se utiliza cuando se cuenta con información completa, incluyendo repeticiones y bloques incompletos. Se asume que los experimentos individuales provienen de diseños de bloques incompletos (látices). Para calcular las medias ajustadas en experimentos individuales, se utilizan un modelo mixto y un estimador del error ponderado, al igual que grados de libertad del error y número de repeticiones también ponderados. El análisis de varianza se obtiene mediante el procedimiento MIXED de SAS.

Nuevamente, se hará hincapié sólo en aquellas secciones que difieren de los casos anteriores. El programa completo se muestra en los ejemplos que se incluyen en el sitio de Internet.

```
options ps=5000 ls=78;
filename biplot 'ejemplo5.cgm';
goptions device=cgmmwwc gsfname=biplot gsfmode=replace;
/*****
* Programa para realizar AMMI utilizando los datos completos      *
* y obteniendo la medias ajustadas como latice en Proc MIXED    *
*****/

data a;
  infile'ejemplo5.dat';
  input env $ rep block gen $ yld;
  cards;
proc sort;
  by env gen;
```

Para obtener las medias ajustadas, se realizan análisis individuales para cada experimento por ambiente. Los archivos de salida se generan con la opción **Make**, y con **Noprint** se controla la impresión de resultados del procedimiento **MIXED**.

```
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;
```

En la siguiente sección se indican los pasos para obtener el estimador del error ponderado, que se salva en el archivo **mse2**, las medias ajustadas que se asignan al archivo **medias**, el número de repeticiones ponderado en el archivo **nrepm** y, por último, los grados de libertad ponderados en el archivo **degreem**. Todos estos datos se combinan posteriormente en un solo archivo denominado **stats** para efectuar la prueba de Gollob.

```
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_;

```

Dado que en **MIXED** no se obtiene propiamente un análisis de varianza clásico como el que obtiene utilizando **GLM**, se realiza a continuación un análisis de varianza combinado, considerando los experimentos individuales como bloques completos y recuperando la información intrabloque. Este análisis es secundario y en realidad no interviene en el análisis AMMI ni en la creación del biplot. Se podría eliminar del programa si se desea, ya que es útil sólo para ilustrar la significancia de los términos del modelo: env, gen y envxgen.

```
title1  'Análisis de varianza utilizando todos los datos';  
title2  'se considera como bloques completos en GLM';  
proc glm data=a ;  
  class env gen;  
  model yld = env rep block(rep) gen env*gen / ss4;
```

Para obtener el cuadro de doble entrada de los residuales de la interacción genotipo por ambiente (GEI), se realiza un análisis de varianza con **GLM** utilizando las medias ajustadas en los pasos anteriores (archivo *medias*). Estos residuales son los que se utilizan para el AMMI y el 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;
```

* Descomposición del Valor Singular en IML para calcular el AMMI;

```

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 para la matriz de residuales;

```

Nota: Una vez más, a partir de este punto casi no habrá cambios en los programas con respecto a los dos casos anteriores. Se finaliza el programa con la instrucción *title1*, lo que indica que se utiliza el procedimiento **MIXED**.

```

title1 f=hwcgm001 h=1.0 'AMMI biplot using MIXED, datos ejemplo 5';

```

```

run;

```

Ejemplo 5

Este conjunto de datos comprende 18 genotipos de maíz evaluados en 10 ambientes (combinaciones de condiciones de sequía y escasez de nitrógeno en diferentes localidades como Tlaltizapán, Poza Rica y Cd. Obregón, México). Cada experimento corresponde a un diseño de bloques incompletos (3 bloques de tamaño 6) con 3 repeticiones. La variable respuesta es rendimiento en grano (kg/ha). No se muestra el conjunto de datos dado su tamaño, pues contiene $18 \times 10 \times 3 = 540$ observaciones; éste

corresponde al conjunto de datos denominado *ejemplo5.dat*, incluido en el sitio de Internet.

El programa para el análisis se encuentra en el archivo *ejemplo5.sas*, los resultados se dan en el archivo *ejemplo5.lst*, y el biplot correspondiente se encuentra en el archivo *ejemplo5.cgm*. A continuación se muestran los resultados más importantes.

Cuadro 5.1. Análisis de varianza considerando el experimento como bloques completos y utilizando PROC GLM.

Análisis de varianza utilizando todos los datos
se considera como bloques completos en 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

En estos resultados se observa que todos los componentes del modelo (*env*, *gen*, así como la interacción *envxgen*) son altamente significativos al 1%. Debe tomarse en consideración que estas cifras corresponden únicamente a una aproximación, puesto que se utilizó un diseño original de bloques incompletos. No obstante, aun con el nivel de aproximación que se muestra, dada la alta significancia que se obtuvo, este punto no debe causar preocupación.

Cuadro 5.2. resultados de la prueba de Gollob para cada uno de los términos AMMI.

Resultados de las sumas de cuadrados para los términos AMMI								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

En estos resultados se observa que prácticamente los primeros 7 términos AMMI son significativos al 1%, mientras que el octavo término es significativo al 5%. Sin embargo, se puede considerar, sin pérdida de generalidad, que los tres primeros términos son los más importantes y explican conjuntamente un 77% de la suma de cuadrados total de la interacción genotipo por ambiente.

Cuadro 5.3. Resultados de los scores para la graficación del biplot.

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

Nótese que tanto los genotipos como los ambientes están en orden alfanumérico; sin embargo, los resultados son correctos, lo cual se puede verificar muy fácilmente realizando un análisis de medias con los genotipos y los ambientes como variables numéricas, por medio de PROC MEANS, y comparando los resultados.

AMMI biplot by MIXED, datos Ejemplo 5

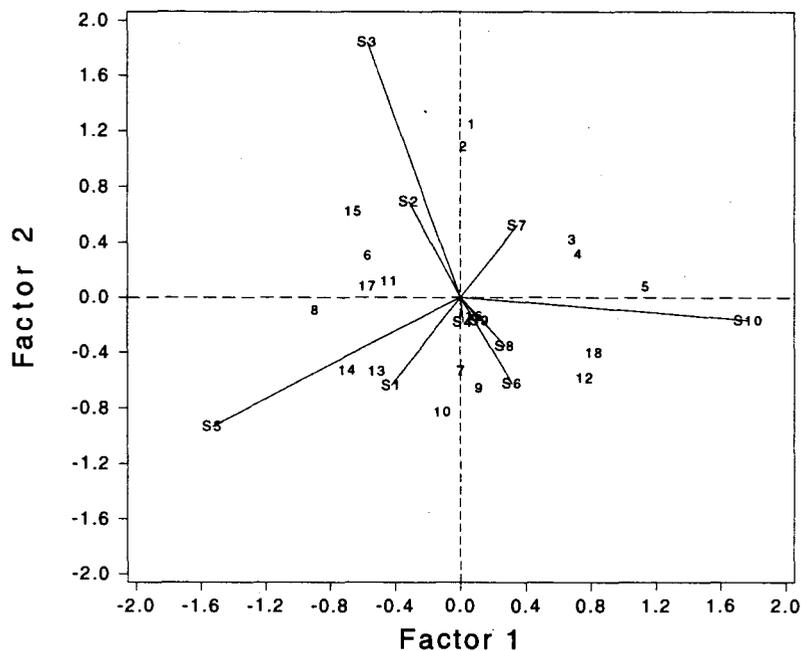


Figura 5.1. Biplot de 18 genotipos evaluados en 10 ambientes y analizados en un diseño de bloques incompletos empleando PROC MIXED.

