

CONSEJERÍA DE AGRICULTURA Y PESCA

Métodos Estadísticos

Para el Estudio de la Estabilidad
Varietal en Ensayos Agrícolas



JOSÉ IGNACIO CUBERO
FERNANDO FLORES

II Edición





MÉTODOS ESTADÍSTICOS

*PARA EL ESTUDIO DE LA ESTABILIDAD
VARIETAL EN ENSAYOS AGRÍCOLAS*

MÉTODOS ESTADÍSTICOS PARA EL ESTUDIO DE LA
ESTABILIDAD VARIETAL EN ENSAYOS AGRÍCOLAS. II Edición

Edita: JUNTA DE ANDALUCÍA. Consejería de Agricultura y Pesca

Publica: Viceconsejería.
Servicio de Publicaciones y Divulgación

Depósito Legal: SE-136-03

Maquetación e Impresión: A.G. Novograf, S.A. (Sevilla)

° Se prohíbe la reproducción parcial o íntegra de esta publicación,
sin la autorización expresa de autor/es, o editor

Σ α **ÍNDICES** \pm ϑ

ÍNDICE

ÍNDICES	5
PREÁMBULO	11



I PARTE: INTRODUCCIÓN

1. INTRODUCCIÓN	21
1.1. Estabilidad y adaptabilidad	21
1.2. Métodos biométricos para estimar la estabilidad: paramétricos univariantes	24
1.2.1. Partición de la Suma de Cuadrados de la Interacción GxE	24
1.2.2. Análisis de regresión	25
1.3. Métodos para medir estabilidad: univariantes no-paramétricos	29
1.4. Métodos para medir estabilidad: Multivariantes.....	29
1.4.1. Componentes Principales	30
1.4.2. Variables ambientales. Regresión Múltiple.....	30
1.4.3. Análisis Factorial	31
1.4.4. Análisis A.M.M.I.....	31
1.4.5. Análisis de Grupos.....	32
1.4.6. Coordenadas Principales	33
1.4.7. Análisis de Correspondencias	33
1.4.8. Análisis Canónico.....	34
1.4.9. Otros métodos.....	34
1.5. Resumen general sobre los métodos expuestos	36



II PARTE: ANALISIS ESTADISTICOS

2. ANALISIS DE LA VARIANZA Y REGRESIÓN CONJUNTA	45
2.1. Datos	45
2.2. Análisis de la Varianza y Regresión Conjunta.....	45
2.2.1. Un paso previo: la homogeneidad de las varianzas.....	45
2.2.2. El Análisis de la Varianza	45
2.2.3. La Interacción GxE	46
2.2.4. Resolución práctica	47
3. PARAMETROS UNIVARIANTES DE ESTABILIDAD MAS UTILIZADOS	55
3.1. Varianza de Estabilidad	55
3.2. Coeficiente de Variación.....	55
3.3. Ecovalencia y Cuadrados Medios de la Ecovalencia	56
3.4. Varianza Ambiental	57
3.5. Medida única de Lin y Binns	57
3.6. Resultados con los datos experimentales	58
4. METODO DE TAI (1971)	63
4.1. Caso práctico.....	64
5. METODOS NO-PARAMETRICOS	69
5.1. Estadísticos no-paramétricos de Hühn	69
5.2. Método de Consistencia de Comportamiento	70
5.3. Método de Kang	70
5.4. Método de Ordenamientos Estratificados	70
5.5. Aplicación práctica	71
5.5.1. Método no-paramétrico de Hühn.....	71





5.5.2. Método de “Consistencia de comportamiento”	72
5.5.3. Método de Kang	73
5.5.4. Método de Ordenaciones Estratificadas	74
6. METODO DE ESKRIDGE	79
6.1. Caso práctico	80
7. METODO ESTRELLA.....	85
8. ANALISIS DE COMPONENTES PRINCIPALES.....	89
8.1. Caso práctico.....	90
8.1.1. ACP de Genotipos. Datos originales de rendimiento.....	90
8.1.2. Datos transformados.....	90
8.1.3. ACP de Ambientes. Datos originales de rendimiento.....	91
8.1.4. Datos transformados.....	92
9. “ADDITIVE MAIN EFFECTS AND MULTIPLICATIVE. INTERACTION”: A.M.M.I.	95
9.1. La validez predictiva.....	96
9.2. Aplicación práctica	96
9.2.1. Procedimientos.....	96
9.2.2. Resultados	97
9.2.3. Estimación de las repeticiones necesarias	98
9.2.4. El “biplot”.....	99
9.2.5. Agrupaciones en el biplot.....	100
9.2.6. Comentarios finales.....	101
9.3. El “ruido” en el Análisis AMMI	101
10. ANALISIS DE GRUPOS	107
10.1. Análisis de grupos según Lin	110
10.2. Análisis de grupos según Fox y Rosielle.....	114
11. ANALISIS DE COORDENADAS PRINCIPALES	119
11.1. Caso práctico.....	122
12. ANALISIS FACTORIAL DE CORRESPONDENCIAS	131
12.1. Caso práctico	131
13. ANALISIS CANONICO	137
13.1. Caso práctico.....	138
14. MODELO SREG. GGE BILOT	143
14.1. Caso práctico	144
15. COMENTARIOS Y CONSEJOS FINALES	149
15.1. Métodos Univariantes Paramétricos	149
15.2. Métodos Univariantes No-Paramétricos	150
15.3. Otros métodos Univariantes.....	150
15.4. Métodos Multivariantes	150
15.5. Comparación de métodos univariantes y multivariantes.....	151
15.6. Plan de Trabajo	152
16. APENDICE	157
16.1. Matrices de entrada para los distintos paquetes estadísticos utilizados	157
16.2. ANOVA y Regresión Conjunta	165
16.3. Estadísticos Paramétricos Univariantes.....	168
16.4. Método de Tai	170

16.5. Estadísticos No-paramétricos	171
16.6. Análisis AMMI	175
16.7. Análisis de grupos según Fox y Rosielle	183
16.8. Análisis de Coordenadas Principales	186
16.9. Análisis Canónico	187
16.10. Modelo GGE biplot	191



INDICE DE FIGURAS

1. Tipos de interacción para dos genotipos hipotéticos	23
2. Líneas de regresión para los 11 genotipos	26
3. Categorías de estabilidad	28
4. Coeficiente de Variación	56
5. Comparación con respuesta máxima	59
6. Gráfica de Tai	65
7. Método de Ketata y col. 1989	70
8. Consistencia del comportamiento (ordenes)	72
9. Consistencia del comportamiento (rendimiento)	73
10. Genotipos 2-6-7 (ordenes bajos)	75
11. Genotipos 10-11-1-5 (ordenes medios)	75
12. Genotipos 8-9-3-4 (ordenes altos)	76
13. Método "Estrella"	85
14. ACP datos sin transformar (ambientes)	90
15. ACP datos transformados (ambientes)	91
16. ACP datos sin transformar (genotipos)	91
17. ACP datos transformados (genotipos)	92
18. Análisis AMMI	100
19. Análisis de grupos Genotipos (UPGMA-DIST. EUCL.)	108
20. Análisis de grupos Ambientes (UPGMA-DIST. EUCL.)	108
21. Análisis de grupos Genotipos (WPGMA-DIST. EUCL.)	109
22. Análisis de grupos Ambientes (WPGMA-DIST. EUCL.)	109
23. Análisis de grupos Genotipos (SINGL-DIST. EUCL.)	110
24. Análisis de grupos Ambientes (SINGL-DIST. EUCL.)	110
25. Análisis de grupos Genotipos (COMPL-DIST. EUCL.)	111
26. Análisis de grupos Ambientes (COMPL-DIST. EUCL.)	111
27. Análisis de grupos Genotipos (METODO LIN 1982)	112
28. Análisis de grupos Ambientes (METODO LIN 1982)	113
29. Análisis de grupos Genotipos (METODO FOX-ROSIELLE)	113
30. Análisis de grupos Ambientes (METODO FOX-ROSIELLE)	114
31. Coordenadas Principales (CICLO L1) (dos dimensiones)	119
32. Coordenadas Principales (CICLO H1) (dos dimensiones)	120
33. Coordenadas Principales (CICLO L1)	120
34. Coordenadas Principales (CICLO L2)	121
35. Coordenadas Principales (CICLO L3)	122
36. Coordenadas Principales (CICLO L4)	123
37. Coordenadas Principales (CICLO H1)	124
38. Coordenadas Principales (CICLO H2)	124
39. Coordenadas Principales (CICLO H3)	125
40. Coordenadas Principales (CICLO H4)	125
41. Coordenadas Principales (CICLO H5)	126





42. Coordenadas Principales (CICLO H6)	126
43. Análisis de Correspondencias	132
44. A.F. de Correspondencias (Genotipos)	132
45. A.F. de Correspondencias (Ambientes)	132
46. Análisis Canónico	137
47. Análisis GGE biplot	145
48. Componentes Principales de todos los métodos utilizados	151

INDICE DE TABLAS



1. ANOVA de la interacción GxE	46
2. ANOVA y Regresión Conjunta de la interacción GxE	47
3. ANOVA y Análisis de la Regresión (caso práctico)	49
4. Rendimiento medio, coeficiente de regresión y S_{-qj} para cada genotipo	50



5. Parámetros de estabilidad \hat{U}_{-i} , s_{-i} , CV_i , W_i , W_i -CM y S_{-xi}	58
6. Parámetros de estabilidad para los cuatro grupos definidos en la Fig. 4	58
7. Índice de Lin y Binns: P_i	59
8. Resultados del Método de Tai	64
9. Cálculo de los estadísticos $S^{(m)}_i$ ($m=1,2,3,6$). Datos transformados	71



10. Cálculo de los estadísticos $S^{(m)}_i$ ($m=1,2,3,6$). Datos originales	72
11. Resultados según el método de la consistencia de comportamiento	73
12. Método de Kang	74
13. Estratificación de rangos	75
14. Valores para los cuatro índices de "primera-seguridad" y las ordenaciones respectivas	80
15. Efectos aditivos e interacción del ANOVA incluyendo los 3 primeros ejes del ACP de la interacción	97
16. Coordenadas genotípicas y ambientales del primer componente del AMMI	98



17. RMS PD ($kg\ ha^{-1}$) para siete modelos basados en datos de rendimiento	98
18. Estimación de parámetros del AMMI para su evaluación predictiva para rendimiento medio de 11 genotipos en 10 ambientes	99
19. Análisis de grupos de los genotipos según el método de Lin	112
20. Análisis de grupos de los ambientes según el método de Lin	112
21. Análisis de grupos de Fox y Rosielle para genotipos	114
22. Análisis de grupos de Fox y Rosielle para ambientes	115
23. Ciclos H y L para el análisis de Coordenadas Principales	121
24. Arbol de mínima dispersión para los primeros genotipos de cada ciclo	123



25. Longitudes medias del AMD para los genotipos más disimilares	124
26. Las tres primeras coordenadas del Análisis Factorial de Correspondencias	132
27. Coordenadas canónicas para cada genotipo	139
28. Coordenadas canónicas para el individuo medio de cada grupo	139



Σ α Π Ω **P R E Á M B U L O** ϑ

PREÁMBULO

Se tiene tendencia a pensar que la labor del mejorador termina con la obtención de las primeras semillas de una nueva variedad (de un nuevo cultivar, para ser más precisos en la terminología), y que el resto –multiplicación, comercialización, conservación, etc.– corresponde a las empresas privadas o públicas. Eso no lo pensará nunca un mejorador profesional, ni ningún alumno de Mejora de Plantas, ni nadie que tenga contacto con este mundo. En realidad, un cultivar no termina nunca de salir de las manos del mejorador: aún cuando éste no se encargue directamente de los ensayos a nivel comercial, siempre estará atento a los datos de estos ensayos, a la respuesta de su material en diferentes localidades, a los factores ambientales que hacen aumentar o disminuir el rendimiento, a la estabilidad de sus variedades en relación a otras obtenciones... en suma, estará interesado en lo que se viene llamando **interacción genotipo-ambiente**, y esa será una información esencial para proseguir el trabajo de mejora.

Huelga decir que ese estudio no sólo interesa al mejorador. A éste se le hace necesario para proseguir su trabajo, pero al extensionista, al agrónomo en general, le resulta esencial por cuanto debe conocer por fuerza el comportamiento de las diferentes variedades existentes en el mercado en los ambientes en que trabaja con objeto de suministrar una información de la mayor importancia: ¿qué cultivares van bien en qué lugares? ¿cuáles podrán recomendarse para todos los ambientes?. Si ninguno, ¿cuáles serán más productivos en los buenos suelos, cuáles “aguantarán el golpe” en los malos?. En general, ¿existe alguna relación entre los diferentes tipos de variedades y los diferentes ambientes en que se pueden sembrar, una relación que permita predecir un cierto comportamiento?

Puede verse que, aunque a primera vista la expresión **interacción genotipo-ambiente** parece excesivamente técnica, su alcance es de enorme importancia práctica. Su estudio es, pues, absolutamente necesario. Hasta hace unos años, los métodos utilizados para dicho estudio eran escasos e insatisfactorios. Ocurría que los que empleaban mejoradores y agrónomos no eran del agrado de los profesionales de la Estadística por simples e imprecisos, por su más que discutible base teórica. A su vez, los análisis estadísticos complejos no eran accesibles a los que trabajaban con material vegetal por su complica-

da elaboración matemática, su dificultad de cálculo y lo problemático de la interpretación de los resultados en términos agronómicos. El acercamiento entre la Estadística y la Agronomía ha propiciado la aparición de numerosos y complejos métodos de estudio de la interacción genotipo-ambiente, perfectamente resolubles gracias a los “paquetes” estadísticos comerciales actualmente existentes y cuya interpretación en términos biológicos no ofrece más dudas que la que resulta de las características del propio material vivo.

No pocos de dichos métodos son de base teórica muy simple. Otros, por el contrario, la tienen muy compleja, muy “sofisticada” según la terminología al uso. Este Manual no entra en las bases teóricas, sea cual sea la complicación de las mismas, a menos que sea estrictamente necesario. Para saciar la sed de rigor matemático se dan abundantes referencias en el texto. Este es un Manual que trata de enseñarle al lector a elegir el método de estudio más apropiado (no siempre lo más complicado es lo mejor), a operar y a interpretar los datos de un caso real. Sólo precisa de un conocimiento elemental de lo que es el diseño de experimentos; podríamos simplificarlo diciendo “un conocimiento de lo que es el análisis de la varianza y de las pruebas estadísticas más comunes”. Eso en cuanto al bagaje teórico. En cuanto al de cálculo, se precisa tener acceso a los paquetes estadísticos usuales, pues los métodos multivariantes no pueden resolverse por medio de operaciones sencillas y estos paquetes resuelven los cálculos más complicados.

Dado que los datos con los que se opera en este manual proceden de un caso real, las interpretaciones de los resultados son relativas a este caso. Eso quitaría generalidad al presente texto como Manual. Para obviar tal limitación, se analizan distintas alternativas (p. ej. “de haberse obtenido un valor, la conclusión hubiera sido”.) con objeto de que el lector pueda encontrar en estas páginas su propio caso. Hemos de decir, sin embargo, que es imposible incluir todos los casos posibles: el material biológico siempre superará en riqueza a todos los cálculos que el hombre pueda realizar. Tarde o temprano, el que practica estos métodos termina encontrando una situación aún no descrita en la literatura científica. Tarde o temprano, pues, el lector terminará analizando sus propios datos y luchando por obtener sus propias conclusiones.

A ayudarle en ese menester es a lo que va dirigido el presente Manual.



NOTA PREVIA

Los datos y corrientes de control necesarios para los distintos paquetes estadísticos utilizados se dan en los apartados del Apéndice. Esos son los datos de partida. Con ellos, siguiendo las instrucciones que se dan para poner en práctica cada método, se obtienen los resultados que figuran en el texto. La práctica usual del análisis de la varianza y de las pruebas estadísticas elementales (pruebas t, F, χ , etc.) se suponen conocidas. A pesar de ello, en el texto se facilitan, cuando se considera oportuno, incluso las operaciones más rutinarias, con objeto de que el lector compruebe que sus cálculos van por buen camino.

El material y las condiciones reales en que se obtuvieron los datos no hacen al caso, por lo que se habla en el Manual de 11 variedades y de 10 ambientes. No obstante, y aunque no se vuelva a mencionar más en el texto, los datos proceden de un ensayo de 15 variedades de habas (*Vicia faba* L.) en 8 localidades andaluzas durante tres años consecutivos, dentro de la Red Andaluza de Experimentación Agrária. Los autores dan las gracias a la Consejería de Agricultura, Pesca y Alimentación de la Junta de Andalucía por haberles permitido disponer de estos datos. Se prefirieron las habas por ser un material sensible a los cambios de ambiente y, por ello, en el que era esperable una alta interacción genotipo-ambiente. El carácter analizado es el del rendimiento, que será en general el carácter problema para el lector.

Una observación sobre terminología. Se utilizan, siempre que es posible, términos castellanos (p. ej. “prueba” en lugar de “test”, etc). Como en este campo, al igual que en tantos otros, los términos ingleses son de uso común, con frecuencia se alternan en el texto los de una y otra lengua; esta doble terminología que el lector encontrará en el texto no obedece, pues, a falta de cuidado sino a todo lo contrario: a familiarizar al lector con la terminología inglesa, de uso universal, mostrándole de cuando en cuando que también existen términos en castellano. Finalmente, se habrá observado que hasta aquí se han manejado indistintamente las palabras “cultivar” y “variedad”. Así se seguirá haciendo en el texto, con clara predilección, al no haber posibilidad de confusión con la terminología que se utiliza en Botánica, por el de uso común en la práctica cotidiana agrícola (esto es, “variedad”) en relación con el más preciso y técnicamente

irreprochable, pero absolutamente pedante en el lenguaje coloquial, de “cultivar”.

En la literatura se habla siempre de “interacción genotipo-ambiente”. Lo que se estudia, en realidad, son “variedades” (o cultivares, como ya se ha dicho), y la identificación de “variedad” con “genotipo” puede hacer temblar a los puristas, y con razón, pues hablar de un genotipo implica un conocimiento profundo del material y una homogeneidad inalcanzable en la práctica, sobre todo en la agrícola. Por bien obtenida que esté una variedad-línea pura de una especie absolutamente autógena, o una variedad-clon de una especie leñosa, es más que dudoso que se tenga la seguridad de que en una y otra sólo existe un genotipo. Sin embargo, dada la práctica unanimitad de la literatura en este caso, se hablará de interacción **genotipo-ambiente** aunque el material de estudio sean **variedades** de todo tipo: se hace la abstracción de que las variedades en ensayo son tan diferentes entre sí que se pueden suponer constituidas por conjuntos de genotipos tales que muestran diferencias mínimas dentro de una variedad y máximas entre variedades.

Por último, un comentario sobre las referencias bibliográficas contenidas en el texto. La finalidad del presente libro es entrenar al lector en diversas técnicas estadísticas de aplicación en agricultura. No es un trabajo de investigación. Por lo tanto, las referencias que se incluyen en el texto son las mínimas necesarias, y se dan cuando hay que otorgar el crédito debido a algún autor concreto (por ejemplo, si es el creador de una nueva técnica, etc). Una proliferación de citas no tendría más efecto que hacer penosa la lectura. Sin embargo, las referencias necesarias se dan al final de cada capítulo: las básicas en las que se encuentran descritos los métodos y algunos trabajos de investigación que los aplican. Sólo encontrará el lector más nombres de los deseados en el primer capítulo, y ello es debido a que en él se hace una descripción somera de los métodos de análisis cuya práctica se detallará en capítulos sucesivos.

MODO DE OPERAR

En cada capítulo (excepto en el 1, que contiene una somera descripción de los diferentes métodos) se comienza por una breve exposición teórica, incluyendo las formulaciones necesarias para el cálculo. A continuación se aplica todo

ello a los datos reales, que se dan en el Apéndice. Deben ser grabados en disco duro o en disquete y *comprobar su correcta grabación* antes de proceder a cualquier tipo de análisis.

En cada caso se mencionan los programas y subprogramas que hay que utilizar. Se reco-

miendan los paquetes estadísticos SAS y NTSYS.

Las corrientes de control necesarias se incluyen en el Apéndice, donde también figuran algunos programas en BASIC elaborados para este Manual.



I. PARTE

INTRODUCCIÓN

Σ α Π Ω \times \pm ϑ

\pm \times \pm ϑ Ω \times α \pm

ϑ Ω \times \pm ϑ Σ α

Π Ω \times \pm ϑ Σ

Ω \times \pm ϑ α Π

α \pm ϑ Σ α \times

Ω \times \pm ϑ Σ α \pm

ϑ Σ α Π Ω \times Σ

1

Σ α Π Ω **INTRODUCCIÓN** ϑ

1. INTRODUCCIÓN

Para una efectiva utilización de los recursos vegetales en programas de Mejora existe la necesidad de evaluar y describir sus variaciones. El problema se complica porque numerosos caracteres se heredan cuantitativamente y son grandemente influenciados por el ambiente, por lo que resulta necesario conseguir la máxima información sobre las diferencias genéticas entre variedades o poblaciones y sus respuestas en diversos ambientes, esto es, lo que se llama **Interacción Genotipo-Ambiente (GxE)**.

La **interacción GxE** surge cuando una variación ambiental tiene distinto efecto sobre genotipos diferentes o, a la inversa, cuando un mismo genotipo responde de distinta manera en diversos ambientes. En otras palabras, cuando no se puede asociar una desviación producida por el ambiente con un ambiente específico (lugar, año, combinación localidad-año, tipo de suelo, etc.) sin tener en cuenta el genotipo sobre el cual aquella actúa. Ello se traduce en una interacción en sentido estadístico entre genotipo y ambiente.

Desde el punto de vista de la Mejora de plantas, la interacción **GxE** representa, sobre todo, un continuo reto: puede ser una eficaz ayuda, pero también puede reducir los progresos de la selección y los rendimientos esperados por el destinatario de todos los programas de Mejora: el agricultor.

1.1. Estabilidad y adaptabilidad

Los investigadores coinciden en general en la importancia de obtener **variedades estables** (se sobrentiende: estables en ambientes muy diversos) de alto rendimiento, pero no se ponen de acuerdo en la definición del concepto de **estabilidad** ni en los métodos estadísticos para su estimación.

La interacción **GxE** puede ser interpretada en parte como resultado de una respuesta diferencial a diversos estreses ambientales (sequía, salinidad, enfermedades, ...); consecuentemente, la mejora para resistencia se traduciría en un aumento de la estabilidad. Pero generalmente, sólo una pequeña parte de la interacción **GxE** puede ser atribuida a determinantes ambientales conocidos; la mayor parte es justamente una cantidad inexplicable en los análisis estadísticos de los ensayos de rendimiento, a la que se conoce con el término de "error experimental". Cuan-

do se estudian estas variaciones en rendimiento se utiliza el término **estabilidad fenotípica** para referirse a las fluctuaciones de la expresión fenotípica del rendimiento mientras el genotipo permanece constante

Hay que decir que los términos de *estabilidad fenotípica*, *rendimiento estable*, y *adaptación* se interpretan de muy distintas formas, lo que motivó la humorística frase de Dorst: **"la palabra adaptación tiene gran adaptabilidad ..."** (Dorst, 1957). Es un sentir común, en efecto, que cada científico tiene su propio punto de vista sobre estos conceptos.

Hay mejoradores que prefieren el término **ADAPTACIÓN** dentro del contexto de una variación en el espacio y utilizan el término **ESTABILIDAD** cuando se refieren al estudio de un sitio durante varios años. Otros muchos utilizan *estabilidad* y *adaptación* para referirse tanto a la dimensión espacial como a la temporal.

La capacidad de algunos genotipos para comportarse óptimamente en un amplio rango de ambientes es lo que Finlay y Wilkinson (1963) definieron como **adaptabilidad general**, y la capacidad de comportarse óptimamente en algunos ambientes específicos, como **estabilidad específica**.

Pueden distinguirse dos conceptos básicos de **estabilidad fenotípica** tal y como estableció Becker (1981):

1) **concepto biológico**: tendrán estabilidad fenotípica aquellos genotipos que muestran variación mínima, medible por la varianza estadística, frente los diferentes ambientes;

2) **concepto agronómico**: un genotipo estable será el que muestre una mínima interacción medida por la **ecovalencia** (véase más abajo su definición).

Tradicionalmente, los mejoradores han utilizado el término de **estabilidad** para caracterizar aquellas variedades (recuérdese: idealmente *genotipos*.) que muestran un rendimiento constante en diferentes ambientes, es decir, los que muestran una varianza mínima frente a éstos. Esta idea de estabilidad está en la misma línea que el concepto de **homeostasis genético** usado en Genética Cuantitativa, e introducido por Lerner (1954) quien lo definió como una propiedad de los organismos a adaptarse a condiciones variables





permitiendo una estabilización frente a las fluctuaciones ambientales; constituye el llamado **concepto biológico de estabilidad**.



Pero en la agricultura moderna, la idea homeostática de estabilidad no resulta interesante debido a que unas nuevas variedades no responderían con incrementos de rendimiento a unas mejores condiciones de crecimiento. Este hecho nos conduce al **concepto de estabilidad agronómica**. Así, un genotipo se considera estable cuando su rendimiento está en el nivel de productividad determinado por un ambiente concreto. Una medida de esta estabilidad es la **ecovalencia (W_i)** propuesta por



Wricke (1962,1964) definida como la contribución de un genotipo concreto a la suma de cuadrados total de la interacción **GxE** en un análisis de la varianza.



La interacción **GxE** se considera de tipo **cuantitativo** si las ordenaciones de los genotipos no cambian al pasar de un ambiente a otro, esto es, si la respuesta diferencial de un genotipo comparado con otro es la misma independientemente del ambiente en cuestión (Fig. 1, b). Para los mejoradores este tipo de interacción cuantitativa es menos importante que la denominada de tipo **cualitativo**, en la cual los genotipos cambian sus ordenaciones relativas en los distintos ambientes (Fig. 1, c). Este tipo de interacción complica el proceso de selección y recomendación de buenos genotipos.



Lin y col. (1986) y Lin y Binns, (1988, 1989), por su parte definen cuatro conceptos estadísticos de estabilidad. Un genotipo se considera estable:

- Tipo 1: si su varianza entre ambientes es pequeña.
- Tipo 2: si su respuesta a los ambientes es paralela a la respuesta media de todos los genotipos del ensayo.
- Tipo 3: si la suma de cuadrados residual de la regresión sobre el índice ambiental es pequeña.
- Tipo 4: si su cuadrado medio dentro de ambientes, esto es, si su variabilidad para cualquier ambiente en diferentes años, es pequeña.



a) No existe interacción, b) Interacción de tipo cuantitativo, c) Interacción de tipo cualitativo.



Se observará claramente el aumento de complejidad en las definiciones. La estabilidad de

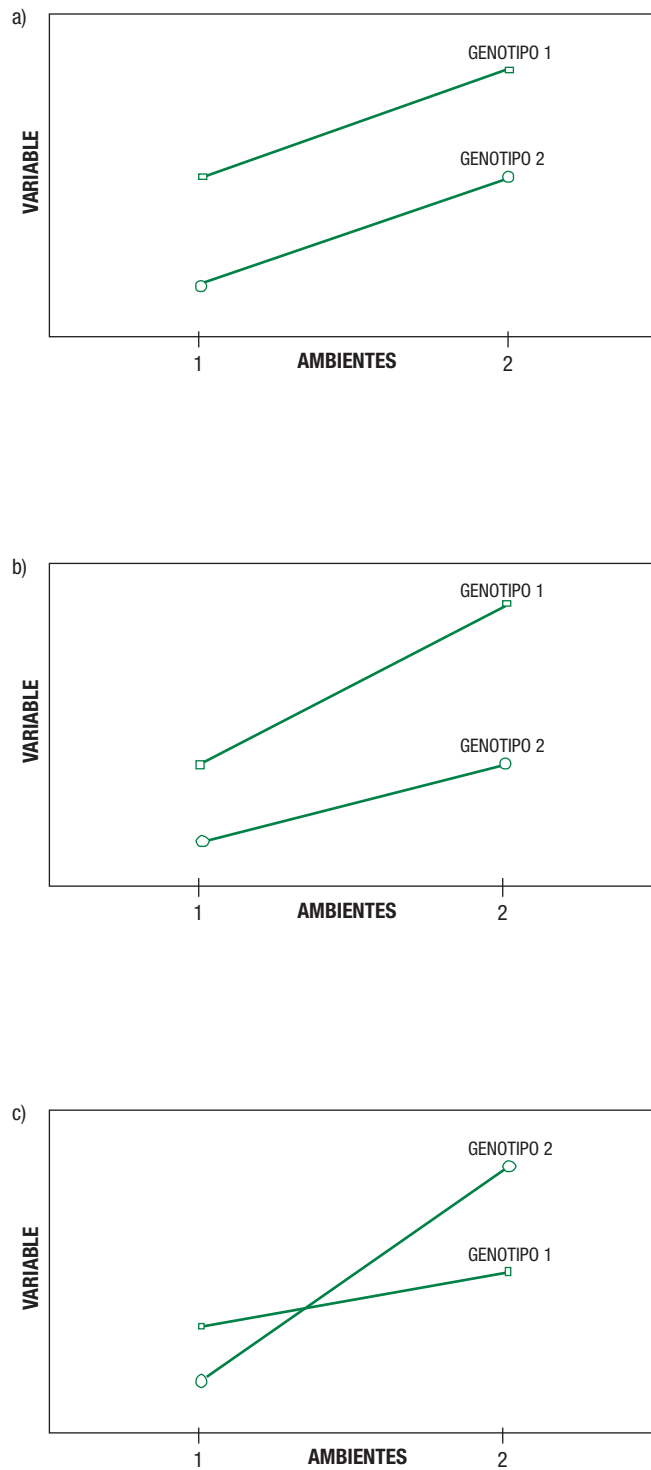
Tipo 1 es análoga al concepto de homeostasis y corresponde a la definición de estabilidad bajo un concepto biológico, diferenciándola de la definida bajo un concepto agronómico, la cual es equivalente al Tipo 2 (ver más arriba).

La estabilidad de Tipo 3 está basada en el concepto de estabilidad de Eberhart y Russell (1966), y se fundamenta en que la variabilidad de un genotipo con respecto al ambiente (localidades, años, localidades x años) puede ser dividida en dos partes: una parte *predecible* correspondiente a una posible regresión entre genotipos y ambientes, y otra *impredecible* correspondiente a los cuadrados medios de la desviación estadística respecto a la regresión lineal. Un genotipo se consideraría estable si la suma residual de cuadrados es pequeña. Todos estos parámetros los obtenemos por medio de un análisis de la varianza (ANOVA) convencional.

La estabilidad de Tipo 4 puede medir la capacidad de un genotipo, en sentido homeostático, para resistir variaciones ambientales impredecibles o súbitas. En contraste con los otros tres tipos de estabilidad, el Tipo 4 tiene la ventaja de proporcionar información acerca de la capacidad de un genotipo para superar la variación debida al componente impredecible, propiedad que no poseen los anteriores tipos de estabilidad. Sin embargo, una desventaja del Tipo 4 es que se requiere incluir en el ensayo un factor "tiempo" añadido al término variedad x localidad. La razón es que sólo entonces podremos separar la variación predecible, proveniente del factor "suelo", de la variación impredecible derivada del factor "tiempo". El factor ideal de tiempo es el "año", pues parece el más natural, pero no tiene por qué serlo; pueden serlo, en efecto, las estaciones, las épocas de siembra, etc.

Todos estos conceptos están en clara contradicción entre sí, habiéndose usado de forma alternativa en muchos trabajos, puesto que un genotipo estable en un sentido puede ser inestable en otros sentidos. Por ejemplo, un genotipo puede ser estable de Tipo 1 pero ser inestable de Tipo 2, o puede ser considerado como inestable desde el sentido de estabilidad de Tipo 3 pero simultáneamente estable de Tipo 2, etc. Es difícil reconciliar todos estos conceptos dentro de un criterio unificado. La razón básica de ésta dificultad es que **la respuesta de un genotipo a distintos ambientes es multivariante**. Y, sin embargo, las definiciones

Figura 1. Tipos de interacción para dos genotipos hipotéticos.



a) No existe interacción, b) Interacción de tipo cuantitativo, c) Interacción de tipo cualitativo.

Σ

α

Π

Ω

\times

\pm

ϑ



anteriores la transforman en un problema univariante, ya que todo se basa en la obtención de un sólo índice de estabilidad.

Se han propuesto numerosos métodos para estimar la estabilidad fenotípica o analizar la interacción **GxE**; la mayoría están fuertemente relacionados con el concepto agronómico de estabilidad, como por ejemplo los *métodos paramétricos univariantes*. Posteriormente, se propusieron *métodos no paramétricos*, los que, por estar basados en las ordenaciones de los genotipos, no necesitan cumplir ninguna condición sobre la distribución de los datos observados (en términos estadísticos, sobre la homogeneidad de varianzas). También se han introducido técnicas de *análisis multivariante* para el análisis de la estabilidad. Estos métodos suministran una información amplia y nueva sobre la verdadera respuesta de los genotipos a los ambientes.

1.2. Métodos biométricos para estimar la estabilidad: paramétricos univariantes

El primer paso en el análisis del comportamiento de una serie de genotipos en diferentes ambientes es la comprobación de la significación de la interacción genotipo-ambiente. Esto requiere llevar a cabo un análisis de la varianza. Cuando la interacción **GxE** es significativa, hemos de estudiar la naturaleza de tal interacción como paso previo a efectuar cualquier inferencia sobre los factores principales, es decir, genotipos y ambientes.

Léon (1985), definió dos conceptos básicos de estabilidad, dependiendo del fin propuesto y del carácter en consideración: **concepto estático** y **concepto dinámico**. Ambos coinciden respectivamente con los ya mencionados conceptos biológico y agronómico de Becker (1981).

La estabilidad fenotípica de acuerdo con el *concepto estático* fue medida ya por Roemer en 1917, y más tarde por numerosos autores, utilizando la varianza de cada genotipo respecto de todos los ambientes del ensayo, que denominó **varianza ambiental** (S_{xi}) la cual detecta todas las desviaciones respecto de la media genotípica (ver fórmula [14]). Este concepto de estabilidad es útil para caracteres como calidad, resistencia a enfermedades y a estreses abióticos (frío, sequía, ... etc.).

Pero un mejorador trata, casi siempre, de buscar genotipos estables y, al mismo tiempo, de alto rendimiento. Ahora bien, la estabilidad medida en base a este concepto estático va normalmente asociada a rendimientos relativamente bajos. Por ello, se recomienda el *concepto dinámico* para estudios de estabilidad de rendimiento.

Frecuentemente, en relación con el rendimiento y otros caracteres cuantitativos, muchos genotipos reaccionan de forma similar a condiciones ambientales favorables o desfavorables. Esta respuesta diferencial al ambiente produce variaciones de la media para el carácter considerado. De acuerdo con el *concepto dinámico*, sólo las desviaciones de un genotipo, respecto a todos los ambientes, se consideran como contribuyentes de inestabilidad en relación con esta reacción general, porque la respuesta general (es decir, la media de todos los genotipos en un determinado ambiente) puede ser interpretada como un efecto (=índice) ambiental.

1.2.1. Partición de la Suma de Cuadrados de la Interacción GxE

La suma de cuadrados de la interacción **GxE** puede ser descompuesta en componentes que ayuden a describir el comportamiento genotípico. Se han propuesto numerosos métodos para caracterizar la contribución genotípica a la suma de cuadrados de la interacción **GxE**, métodos que se verán en detalle en capítulos posteriores. Aquí sólo se dará una breve idea de los mismos.

Salmon (1951) y Plaisted y Peterson (1959) realizaron un análisis de varianza para todos los genotipos tomados de dos en dos, con objeto de estimar la varianza de la interacción para todas las combinaciones posibles de genotipos. La media de las varianzas obtenidas para cada genotipo se utilizaba como un índice de la contribución de un genotipo a la interacción **GxE** total.

Plaisted (1960) propuso estimar la contribución de cada genotipo a la suma de cuadrados de la interacción **GxE** de la siguiente forma: 1º) se realiza el análisis de la varianza con todos los genotipos del ensayo omitiendo en el análisis el genotipo del cual queremos calcular su contribución a la **GxE**; 2º) una medida de la estabilidad del genotipo omitido vendría dada, según Plaisted, por la varianza de la interacción **GxE** obtenida en el análisis anterior.





Wricke (1962) define el término **ecovalencia** (W_i) como la contribución de cada genotipo, en todos los ambientes, a la suma de cuadrados de la interacción (ver fórmula [12]). Si la ecovalencia es pequeña, la estabilidad agronómica (o de Tipo 2) es alta.

Shukla (1972) propuso una estima insesgada de la varianza de la interacción para cada genotipo. A este estadístico lo denominó **varianza de la estabilidad** (ver fórmula [8]).

Se demuestra que la varianza de la estabilidad, para un determinado genotipo, es una combinación lineal de su ecovalencia y ambos resultan en ordenaciones similares (ver [13]). Puede comprobarse también que todos estos estadísticos producen la misma ordenación para un conjunto de genotipos determinado; sin embargo, mientras los tres primeros son meramente índices numéricos, el de Shukla (1972) permite comprobar la homogeneidad de las estimas.

Dado que el establecimiento de testigos en un ensayo de cultivares no resulta siempre fácil, debido a que el mejor testigo en una localidad no tiene por que serlo en otra, Lin y Binns (1988) propusieron una medida única de la superioridad del comportamiento de un genotipo, definiéndola como el cuadrado medio de la distancia entre la respuesta de un genotipo y el genotipo de máxima respuesta, en un ambiente dado. Con este método, la máxima respuesta en un ambiente se convierte en el testigo a considerar. Según dichos autores las ventajas de este método son las siguientes:

- No son necesarios los “genotipos control”, es decir, testigos comunes para todos los ambientes del ensayo: basta con incluir los genotipos mejor adaptados localmente para cada ambiente.
- Sólo se necesita un parámetro para caracterizar la estabilidad de la producción potencial de cada genotipo.
- La adaptabilidad específica de un genotipo se puede identificar dibujando tanto la respuesta real del genotipo de máxima respuesta como la del genotipo en cuestión en función de la media de cada ambiente.
- La diferencia entre la media de la respuesta máxima en todos los ambientes y la media del cultivar más productivo suministra una importante información a la hora de recomendar cultivares; así, cuando esta diferencia no es mucho

mayor que el error estándar de la media de la respuesta máxima, el cultivar más productivo puede ser recomendado en toda la región.

La principal desventaja encontrada por algunos investigadores es que dicho estadístico muestra una alta correlación con el rendimiento, que puede llegar a ser total. De este modo, seleccionar por este estadístico es seleccionar fundamentalmente por rendimiento y no por el binomio rendimiento+estabilidad.

1.2.2. Análisis de regresión

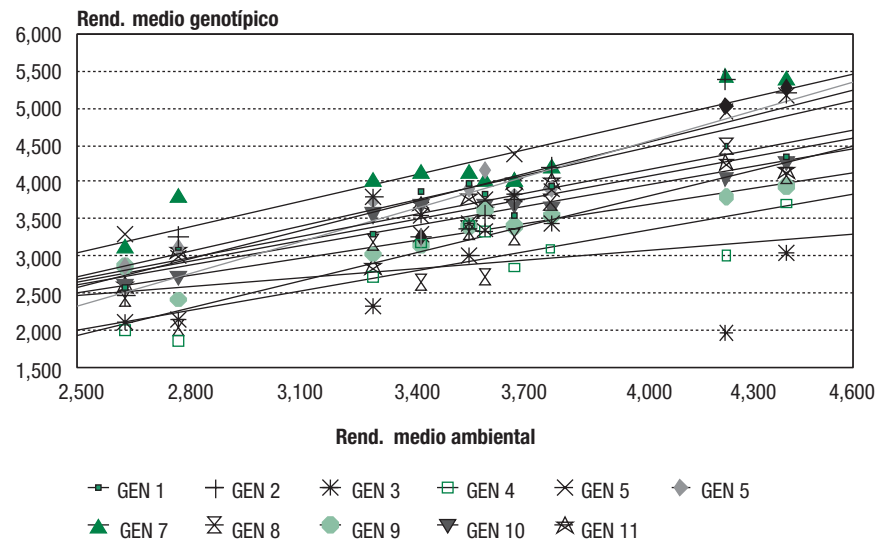
La forma más frecuente de caracterizar la estabilidad fenotípica a través de los ambientes es mediante el coeficiente de regresión de cada genotipo sobre la media ambiental. Por su popularidad entre los mejoradores, se lo considerará aquí con cierto detalle.

El primero en introducir ésta técnica fue Mooers (1921); posteriormente fue descrita y elaborada por Yates y Cochran (1938), quienes utilizaron la media total de todos los genotipos en un ambiente determinado como *índice ambiental*, esto es, como medida del ambiente. El comportamiento de un cierto genotipo o grupo de genotipos se estudiaba por medio de la regresión lineal de las medias del genotipo en cada ambiente en función de los dichos índices o valores ambientales (Fig. 2).

El uso de ésta técnica como base para el análisis de varianza y asociado a pruebas de hipótesis fue discutido por Mandel (1961), quien demostró ser una extensión de la prueba de Tukey (1949) con un grado de libertad para no-aditividad. En efecto, si t es el número de genotipos, Mandel definió un componente con $t-1$ grados de libertad de la suma de cuadrados de la interacción; dicho componente es independiente de los efectos ambientales y representa la suma de cuadrados de la heterogeneidad de las t pendientes de regresión. Mandel demostró que si las pendientes son idénticas éstas se distribuyen como una Chi-cuadrado; basta pues, realizar una prueba **F** para comprobar la existencia de pendientes diferentes, siendo la hipótesis nula el que todas las pendientes sean iguales.

Finlay y Wilkinson (1963) consideraron que una simple comparación de pendientes no es suficiente para determinar estabilidad; la media

Figura 2. Líneas de regresión para los 11 genotipos.



total de cada genotipo también debería entrar en el estudio. Así, los genotipos con pendiente cercana a 1 y alto rendimiento se considerarían como bien adaptados a todos los ambientes.

Como se puede observar, todo ello no es nada más que un redescubrimiento de las ideas de Yates y Cochran ya expuestas en 1938.

En el método del análisis de regresión los efectos de la interacción **GxE** pueden ser expresados de la siguiente forma:

$$(ge)_{ij} = \beta_i e_j + d_{ij}$$

donde β_i es el coeficiente de regresión lineal para el genotipo i y d_{ij} un residuo.

Se han propuesto dos técnicas de regresión aparentemente distintas para explicar la interacción **GxE**. Una calcula la regresión de los efectos de la interacción sobre los efectos ambientales (β_i ; Perkins y Jinks, 1968a,b) y la otra realiza la regresión de valores medios sobre medias de ambientes (b_i ; Finlay y Wilkinson, 1963) (Fig. 2): Ahora bien, ambos estadísticos son equivalentes ya que $\beta_i = b_i - 1$.

Como complemento al coeficiente de regresión, el cuadrado medio residual de las desviaciones de la regresión (S^2_{di}), tal como se lo obtiene en un análisis rutinario de regresión, describe la contribución del genotipo i a la interacción **GxE** (Eberhart y Russell, 1966). Cuando el ajuste al

modelo lineal es bueno, S^2_{di} y ecovalencia están altamente correlacionadas, según Becker (1981).

Estos dos estadísticos del método de regresión, b_i y S^2_{di} , han sido utilizados de diferentes formas. Mientras S^2_{di} está fuertemente relacionado con la parte impredecible de la variabilidad de un genotipo (es decir, puede ser observado como un parámetro de estabilidad), el coeficiente de regresión b_i caracteriza la respuesta específica de los genotipos al efecto ambiental, pudiéndose considerar como un parámetro de respuesta. Los genotipos que no reaccionan frente a los factores ambientales mostrarán una pendiente cero y podrían ser estables de acuerdo con el concepto estático. Por otro lado, los genotipos que muestran una respuesta media a los cambios ambientales (es decir, responden pobremente en los ambientes de bajo rendimiento y responden muy bien en los ambientes de alto rendimiento) tendrán pendientes igual a 1. Para la ordenación de los genotipos será más estable aquél que muestre el valor de S^2_{di} más próximo a cero.

Pinthus (1973) propuso utilizar el coeficiente de determinación (r^2_i), asimismo obtenido en el análisis de regresión, como estimador de la estabilidad de un genotipo en lugar de S^2_{di} . La aplicación de r^2_i tiene la ventaja de que es independiente de las unidades de medida.

Freeman (1973) denomina **análisis de varianza y regresión conjunta** al hecho de

descomponer los grados de libertad $(t-1)(s-1)$, siendo t el número de genotipos y s el de ambientes, y la suma de cuadrados de la interacción $G \times E$, en dos componentes, uno con $t-1$ grados de libertad, para comprobar la heterogeneidad de las regresiones, y otro que recoge las desviaciones de la regresión con $(t-1)(s-2)$ g.l. Cada uno de estos componentes se compara con el cuadrado medio del error experimental; si la heterogeneidad de las regresiones es significativa, podemos rechazar la hipótesis nula de que todos los genotipos tienen los mismos coeficientes de regresión y, por consiguiente, la misma estabilidad. Además, el componente de heterogeneidad puede compararse con el cuadrado medio de las desviaciones para comprobar si las regresiones explican una parte importante de la interacción observada.

Eagles y col., (1977) sugieren que de la fracción denominada "heterogeneidad de las regresiones", en el análisis de varianza y regresión conjunta, se puede extraer un componente con un grado de libertad que nos indicaría la tendencia de las líneas de regresión a *converger* en un punto. Cuando las líneas de regresión tienden significativamente a converger en un punto nos indica la existencia de una alta correlación entre los coeficientes de regresión y las medias genotípicas en todos los ambientes. Esta conclusión la podemos deducir sin más que ver la ecuación [9], ecuación que nos permite calcular la suma de cuadrados del factor convergencia.

A pesar de la gran utilización que ha tenido la técnica de regresión, los problemas inherentes a la misma son numerosos e importantes, habiendo sido objeto de discusión en muchos trabajos. Aunque es un método práctico y funcional, es de débil base teórica ya que:

1) La media genotípica no es independiente de las medias marginales de los ambientes. Realizar la regresión de una variable sobre la otra no siendo independientes viola uno de los requisitos fundamentales del análisis de regresión. Esta interdependencia es un problema grave cuando el número de genotipos es pequeño, pero cuando éste es superior a 15 el problema es menor. Como solución, Freeman y Perkins (1971) sugirieron que la regresión podría estar basada sobre medidas independientes de los ambientes; así, por ejemplo, se podrían realizar los cálculos omitiendo sucesivamente el genotipo en cuestión en el análisis de la regresión, o bien añadiendo genotipos tes-

tigo al ensayo. Sin embargo, si el número de genotipos es grande los resultados difieren muy poco cuando empleamos estas medidas independientes de las que se obtienen utilizando la media ambiental.

2) La segunda limitación estadística es que el error asociado con las pendientes de los genotipos no es estadísticamente independiente de éstos, porque la suma de cuadrados para la desviación, con $(t-1)(s-2)$ g.l., no puede ser subdividida ortogonalmente en t componentes correspondientes a genotipos.

3) Implícito en el método está el admitir que existe una relación lineal entre la interacción y las medias ambientales. Cuando este requisito no se da, la efectividad del análisis se reduce y los resultados pueden ser engañosos ya que el análisis requiere que una alta proporción de la interacción pueda ser explicada mediante regresión lineal. Otros autores han propuesto análisis de regresión no lineales; por ejemplo, Hill y Baylor (1983) utilizaron un contraste ortogonal del análisis de la varianza subdividiendo la variación frente a ambientes (años y localidades) para cada genotipo en efectos lineales y cuadráticos de los ambientes.

4) Los coeficientes de regresión son estimaciones sesgadas debido al hecho de suponer que en el análisis de regresión, la variable independiente, en este caso la media ambiental, ha sido medida sin error. Este sesgo depende del número de genotipos del ensayo y del cociente varianza ambiental/varianza del error. En un buen diseño experimental (por ejemplo, un gran número de repeticiones y/o contrastando condiciones ambientales) este cociente sería grande y el sesgo pequeño, pero en otras muchas condiciones (por ejemplo, un número pequeño de repeticiones, etc), el sesgo puede ser importante.

5) Existe una interdependencia biológica y algebraica entre las pendientes y las sumas de cuadrados de las desviaciones de la regresión. Esto puede ser consecuencia del propio procedimiento de ajustar los datos a una línea recta.

6) Otro problema es que cuando en nuestro ensayo existen pocos ambientes que sean o de alto o de bajo rendimiento, el ajuste de un genotipo concreto puede estar grandemente influenciado por su comportamiento en éstos pocos ambientes, con posibilidad de obtener resultados muy dudosos. Es decir, el coeficiente y las desviaciones de la regresión pueden





variar sustancialmente al no disponer de ambientes extremos.

7) La relativa estabilidad de dos genotipos cualquiera depende no sólo del grupo de localidades analizadas en el ensayo sino también del resto de genotipos ensayados. Se ha demostrado que la estabilidad de un genotipo depende del comportamiento medio del grupo en el cual está siendo comparado.



8) Cuando el conjunto de genotipos es grande, parece existir una correlación positiva entre producciones medias y los coeficientes de regresión.

9) Estas técnicas son más descriptivas que predictivas, y por ello, no mejoran el conocimiento sobre la interacción **GxE**.



10) Con bastante frecuencia, no hay suficiente variabilidad genética dentro del conjunto de genotipos estudiados.

11) Para determinar en un ensayo cual de los genotipos se comporta como superior a los demás, los mejoradores tienen la dificultad de definirse entre el rendimiento medio, pendiente y desviación de la regresión, porque la respuesta genotípica a los ambientes es intrínsecamente multivariante y el análisis de regresión lo transforma en un problema univariante.



Revisando estas consideraciones estadísticas, Skråppa (1984) concluyó que en aquellos casos en que las desviaciones de la regresión, así como la convergencia de las líneas de regresión, no son significativas, y cuando la heterogeneidad de las pendientes es significativa, si se

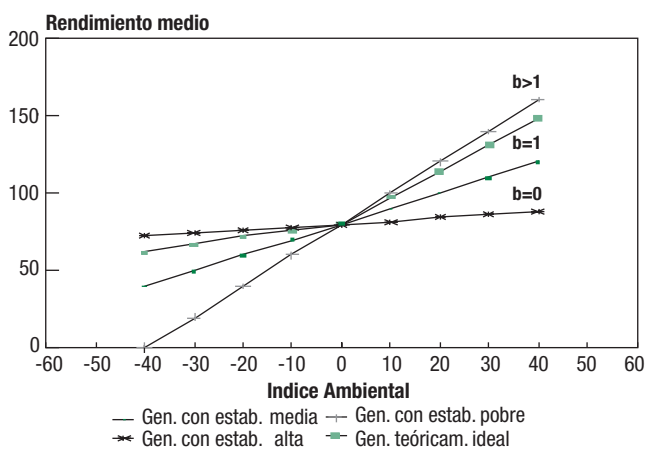


dispone de un número suficientemente grande de genotipos, ambientes y repeticiones, y con homogeneidad de varianzas (necesaria para el análisis de varianza), el método de regresión puede ser recomendable.

Regresiones segmentadas. El análisis de regresión convencional presupone una relación lineal entre la interacción **GxE** y los valores aditivos ambientales (e_j). Cuando, sin embargo, esta relación es significativamente no lineal, el análisis de regresión simple no puede ser utilizado plenamente como herramienta predictiva. Verma y col. (1978) proponen una modificación a la regresión lineal convencional; sugieren dividir los ambientes en dos grupos según el signo del índice ambiental: los ambientes de bajo rendimiento (o signo negativo) y los de alto rendimiento (signo positivo). Se efectuaría la regresión lineal para cada grupo de ambientes (con el índice ambiental, como siempre, en abscisas), con un punto de truncadura en el valor cero (Fig. 3). Esta modificación es muy simple y es un método más conveniente para detectar diferencias entre genotipos que complicadas regresiones cuadráticas.

Según Virk y col. (1988) al utilizarse como índice ambiental, en los modelos de regresión lineal, la media de todos los genotipos, la consecuencia de medir el valor ambiental de esta forma es que la respuesta media de todos los genotipos frente a todo el conjunto de ambientes podría tener pendiente uno. Por ello, dichos autores modifican el modelo anterior de la Fig. 3, que tiene un punto fijo de truncamiento (pues sólo considera dos grupos de ambientes), por un modelo en el cual existen varios puntos de truncamiento específicos para cada genotipo y determinados por los cambios

Figura 3. Categorías de estabilidad.



de respuesta frente a la serie de ambientes establecidos en el ensayo. El análisis se aproxima así a un estudio de regresión no lineal, aunque de más fácil ejecución.

En resumen, siendo el análisis de regresión el método más utilizado por los mejoradores en el estudio de la interacción **GxE**, no hay que olvidar que los resultados dependen, en gran medida, del grado de linealidad así como de la independencia existente entre el índice ambiental y el valor de los genotipos ensayados.

1.3. Métodos para medir estabilidad: univariantes no-paramétricos

Cuando utilizamos las pruebas de significación para estadísticos paramétricos que nos miden la estabilidad, realizamos implícitamente suposiciones sobre sus distribuciones y sobre la homogeneidad de sus varianzas. En muchos casos, dichas hipótesis previas ni siquiera se prueban. Sin embargo, con medidas no-paramétricas de estabilidad no tenemos que tener en cuenta suposiciones de este tipo. Las pruebas no-paramétricas se basan en el orden de los genotipos en los distintos ambientes; por tanto, un genotipo será estable si su posición en el ordenamiento general de todos los genotipos es similar en los distintos ambientes. Otra ventaja de los estadísticos no-paramétricos frente a los paramétricos es que son menos sensibles a errores de medida y que la adición u omisión de una o varias observaciones no causa grandes variaciones en las estimaciones. Según Becker y León (1988), el primero que utilizó esta técnica fue Haldane (1946).

Normalmente los números de orden de los genotipos en cada ambiente se determinan por sus valores observados X_{ij} (=valor medio respecto de las repeticiones); sin embargo, la estabilidad fenotípica puede ser medida independientemente de los efectos genotípicos sin más que utilizar para la ordenación los valores X_{ij} corregidos [$X^*_{ij}=X_{ij}-(X_{i.}-X_{..})$], donde $X_{i.}$ es la media marginal del genotipo **i** y $X_{..}$ la media total del ensayo. De esta forma, el número de orden de cada genotipo sólo dependerá de la interacción y del error experimental.

Los estadísticos no-paramétricos calculados en base a las ordenaciones de los datos corregidos, que permiten una mejor interpretación de los resultados que el de los datos sin corregir,

fueron propuestos por Hühn (1979) y Nassar y Hühn (1987).

Ketata y col., (1989) propusieron el uso simultáneo de la posición media y su desviación estándar para cada genotipo. Representando gráficamente estas dos variables para todos los genotipos podremos seleccionar aquellos que ocupen una situación de posición media alta y de una desviación estándar pequeña. Dichos autores afirman que para un alto número de genotipos ambas variables no están correlacionadas y entonces es posible seleccionar independientemente para alto rendimiento y baja variabilidad ambiental, esto es, para **consistencia de comportamiento**.

Kang (1988) ideó un nuevo método no-paramétrico para seleccionar genotipos estables de alto rendimiento: se establecen las ordenaciones de los rendimientos medios de los genotipos, dándole el valor 1 al genotipo de más alto rendimiento. De igual forma se establecen las ordenaciones de las varianzas de estabilidad definida por Shukla (1972) (vease 1.2.1), con el valor más bajo igual a 1. Los dos números de orden se suman: la suma de menor valor podría corresponder al genotipo más deseable.

El método estratificado de ordenaciones, propuesto por Fox y col. (1990) para analizar la adaptación general, consiste en contar el número de ambientes para los cuales un genotipo ocupa los primeros órdenes, los medios o los bajos del ensayo. Un genotipo con la mayoría de los números de orden en el primer tercio superior se considera como ampliamente adaptado. Es por tanto, un método no-paramétrico muy simple e independiente de toda escala.

1.4. Métodos para medir estabilidad: Multivariantes

El análisis multivariante es apropiado para analizar matrices de dos dimensiones, como son las que contienen datos de **G** genotipos y **E** ambientes. La respuesta de cualquier genotipo en los **E** ambientes puede representarse en un espacio **E-dimensional**, siendo las coordenadas para un eje los rendimientos (o cualquier otro carácter) genotípicos en un ambiente.

Se han utilizado dos grupos de técnicas multivariantes para estudiar la estructura interna de la interacción **GxE**:





1) **Técnicas de ordenación**, tales como Análisis de Componentes Principales, Análisis de Coordenadas Principales y Análisis Factorial. Estas técnicas representan la relación genotipo-ambiente tan fielmente como es posible hacerlo en un espacio de pocas dimensiones. En las gráficas resultantes de estos análisis observamos cómo los genotipos o los ambientes similares están más cercanos entre sí que aquellos cuyo comportamiento es distinto al de los anteriores. Estas técnicas son efectivas para mostrar relaciones y reducir el “ruido” de nuestros datos.

2) **Técnicas de clasificación**, como el Análisis de Grupos y el Análisis Discriminante. Estos métodos agrupan entradas similares en forma de nubes de puntos (“clusters”).

1.4.1. Componentes Principales

El Análisis de Componentes Principales (ACP) es uno de los métodos multivariantes más utilizados y desde luego el más antiguo, cuyo fundamento se le deba a Pearson, nada menos que en 1901. Evidentemente el uso de ordenadores ha facilitado la ejecución de las operaciones matemáticas necesarias, de enorme complejidad en la mayor parte de los casos, y de realización poco menos que imposible por procedimientos tradicionales. Su atractivo radica en transformar los datos de un grupo de coordenadas en otro de más fácil interpretación, el cual preserva, tanto como sea posible, la configuración original del grupo de puntos y concentra un gran porcentaje de la variación o estructura de los datos inicial en las primeras nuevas variables, llamadas *componentes principales*, obtenidas en la transformación. En este proceso de reducción de los datos se pierde inevitablemente alguna información original, pérdida que se compensa por la posibilidad de obtener una mejor interpretación de los datos reales. Este análisis puede reducir efectivamente la estructura de una matriz de interacción **GxE** de **G** puntos (genotipos) en **E** dimensiones (ambientes) a un subespacio de pocas dimensiones.

Hay autores que, bajo ciertas condiciones, lo consideran un modelo equivalente al de la regresión lineal. Williams (1952) demuestra que, efectivamente, la estimación de los coeficientes de regresión por mínimos cuadrados es equivalente a la extracción del primer componente principal en el estudio del comportamiento genotípico. La validez del modelo puede ser entonces evalua-

da extrayendo nuevos componentes principales o, más fácilmente, examinando la matriz de correlación residual, después de extraer el primer componente. Por ejemplo, Perkins y Jinks (1968a,b), estudiando la altura de planta en líneas de *Nicotiana rustica*, encontraron que los genotipos del ensayo podían ser divididos en dos grupos siguiendo el sistema anterior: los genotipos que mostraban una correlación residual positiva formaban un grupo y los que mostraban una correlación negativa otro.

Ahora bien, como ya se ha comentado, el método de la regresión lineal utiliza un sólo estadístico, el coeficiente de regresión, para describir el patrón de respuesta de los genotipos frente a los ambientes, y mucha de la información perdida se recoge en la desviación. Por otra parte, el Análisis de Componentes Principales es una generalización de la regresión lineal que supera esta dificultad generando varios estadísticos, las coordenadas en los ejes de los componentes principales, para describir el patrón de respuesta de un genotipo. No es de extrañar que, pues, que el Análisis de Componentes Principales sea más eficaz que el método de regresión lineal en la descripción del comportamiento genotípico.

La principal dificultad del estudio de la interacción GxE mediante el ACP, estriba en la interpretación de los componentes principales extraídos, ya que puede ocurrir que éstos no muestren ninguna relación obvia con las condiciones ambientales.

1.4.2. Variables ambientales. Regresión Múltiple

Otra posibilidad es caracterizar un ambiente por las medidas de los parámetros que pueden describirlo, tales como temperatura, humedad, precipitación, radiación, etc., ya que ellos pueden explicarnos, al menos parcialmente, la interacción **GxE**. Varios autores han intentado relacionar el comportamiento de caracteres genotípicos, como rendimiento, en los diversos ambientes con variables ambientales particulares mediante simple regresión lineal. Cuando la variable independiente domina el carácter estudiado, entonces, por supuesto, la regresión lineal lo pondrá de manifiesto. Sin embargo, cuando actúan simultáneamente varias variables independientes (incluso cuando actúan aditivamente), o cuando varias de ellas están correlacionadas entre



sí, el análisis de regresión simple no es apropiado. Así pues, cuando están implicados muchos factores, la utilización de la regresión múltiple parece una necesidad.

Esto que parece tan evidente desde un punto de vista teórico no resulta tan sencillo de llevar a la práctica, pues ha de investigarse qué variables y de qué tipo (edáficas, climáticas, estados fisiológicos, etc.) han de ser incluidas en el modelo y, además, en la mayor parte de los casos no se dispone de los datos necesarios. Otra dificultad es que si se consideran simultáneamente muchas variables ambientales, tienen que obtenerse regresiones polinómicas de orden elevado para obtener un buen ajuste.

Varios autores han utilizado el análisis de regresión múltiple incluyendo variables climáticas. Para predecir los rendimientos de diversos genotipos, utilizaron la técnica de regresión “paso-a-paso”, esta consiste en realizar el análisis de regresión de manera exploratoria, es decir, tomando diversas combinaciones de las variables independientes, a fin de obtener un mínimo de varianza residual no explicada en términos del mismo número de variables independientes, apartando cualesquiera variables independientes potenciales que no eliminen una porción significativa de la variación. Se necesita aún un estudio muy profundo de investigación para determinar la utilidad real de este método.

1.4.3. *Análisis Factorial*

El Análisis Factorial es un procedimiento de ordenación relacionado con el Análisis de Componentes Principales. Los factores son similares a los componentes principales. De la misma forma que en éste, un gran número de variables correlacionadas se reducen a un pequeño número de factores. La variabilidad se explica, por el contrario, en términos de **factores generales** comunes a todas las variables y de **factores únicos** a cada variable. Los ejes de los factores generales pueden ser rotados a posiciones oblicuas conforme a las hipótesis previas que se establezcan con objeto de maximizar los factores principales y minimizar los secundarios.

El Análisis Factorial se ha utilizado mucho para estudiar la relación existente entre componentes de rendimiento y características morfológicas de

los cultivos. Pocos investigadores lo han utilizado, sin embargo, para el estudio del comportamiento de los genotipos en un ensayo multilocal.

1.4.4. *Análisis A.M.M.I.*

Mandel (1971) introdujo un nuevo tipo de análisis para el estudio de una matriz bidimensional **GxE** utilizando para ello un método multiplicativo que denominó AMMI “Additive Main effect and Multiplicative Interaction”. El primer paso es realizar el análisis de la varianza y posteriormente particionar la suma de cuadrados de la interacción en una serie de componentes multiplicativos (coordenadas genotípicas x coordenadas ambientales). Prácticamente en este último decenio la utilización del Análisis de Componentes Principales se ha realizado como componente del Análisis AMMI.

El AMMI estudia primeramente los efectos aditivos principales de genotipos y ambientes por medio de un análisis de la varianza y describe la parte no-aditiva de la variación, esto es, la interacción **GxE**, por medio de un Análisis de Componentes Principales.

El método AMMI tiene tres objetivos principales:

1) **Es un método de diagnosis.** Resulta apropiado para el comienzo de los análisis estadísticos del ensayo ya que nos diagnostica la posibilidad de poder estudiar submodelos representativos de los datos totales del ensayo.

2) **Clarificar la interacción GxE.** El AMMI nos da el patrón y comportamiento de los genotipos y ambientes conjuntamente.

3) **Mejora la precisión de estimas del rendimiento.** Se ha demostrado que se obtiene la misma precisión en la estima de rendimiento con el análisis AMMI que con otros tipos de análisis con un número de repeticiones de 2 a 5 veces mayor. Esta mejora en la estima puede utilizarse para rebajar costes, ya que podemos reducir el número de repeticiones, o bien, incluir más tratamientos en el experimento.

Gauch (1985, 1988) indicó la posibilidad de utilizar el modelo AMMI con un enfoque **predictivo**, además del **postdictivo**. Desde el punto de vista **postdictivo** los datos se utilizan para construir un modelo que explique los datos y comparar los valores esperados con los datos



de partida. Bajo un enfoque **predictivo** los datos se dividen en dos partes: los que van a servir para elaborar el modelo (es decir, los datos a los cuales les aplicamos el Análisis AMMI) y los que se reservan para la prueba de validez. A partir de los datos del modelo obtenemos los valores esperados que compararemos con los datos de validación.



Es importante determinar qué modelo AMMI es el más ajustado a nuestros datos. Para ello, existe un estadístico denominado RMS PD (como veremos en el capítulo 9) cuyo valor nos indica cual es el modelo más preciso. Así, podemos construir varios modelos: el AMMI0, que sólo estima los efectos aditivos de genotipos y ambientes sin incluir en el modelo ningún componente principal, esto es, es un análisis de la varianza simple; el AMMI1 que combina los efectos aditivos del AMMI0 con el efecto interacción **GxE** estimado por el primer componente principal (ACP 1); y así sucesivamente hasta incluir todos los componentes principales en el modelo.



Cuando el modelo más preciso es el AMMI1 se pueden representar las coordenadas ACP 1 de genotipos y ambientes (en ordenadas) frente a las medias marginales de los mismos (en abscisas): es el denominado **biplot** de Gabriel (1971).

1.4.5. Análisis de Grupos



El Análisis de Grupos es un método de clasificación que ha sido ampliamente utilizado para aumentar la eficacia de los ensayos o bien como un método alternativo cuando no pueden usarse los métodos univariantes tales como la regresión lineal.

El Análisis de Grupos puede realizarse bien sobre caracteres medidos, o bien sobre cualquier otra variable independiente. El objetivo es identificar grupos de ambientes con un nivel similar de interacción **GxE** a fin de identificar los mejores genotipos para determinadas condiciones. Lo que hacemos es agrupar los genotipos de acuerdo con su respuesta, es decir, agrupar genotipos en subgrupos cualitativamente homogéneos en cuanto a su estabilidad, de tal forma que *dentro* de dichas agrupaciones no haya interacción **GxE**, pero que sí exista *entre* ellas.

Para dilucidar la interacción **GxE** mediante el Análisis de Grupos se han propuesto, al menos, 10 medidas diferentes de disimilaridad y 8 estra-

tegias diferentes de clasificación. La elección que se haga de ellas resultará en grupos distintos. No existe un criterio objetivo para aceptar o rechazar una elección determinada. La comparación de los métodos de análisis de grupos disponibles se basa en dos aspectos: uno es la definición de la *medida de similitud* (o disimilitud) y el otro es el *algoritmo* usado para efectuar la agrupación.

Básicamente hay dos grandes clases de *medidas de similitud*, según el criterio sea *único* o *múltiple*. Para el primero se distinguen cuatro grupos, obteniéndose las medidas de similitud a partir de uno sólo de ellos:

- Distancia euclídea,
- Distancia tipificada,
- Índice de disimilitud, y
- Coeficiente de correlación.

A su vez, dentro de cada grupo existen dos tipos de índices: el que nos define la similitud usando conjuntamente efectos genéticos y la interacción **GxE**, y aquellos en que la similitud está basada solamente en la interacción **GxE**.

En contraste con el criterio único anterior, el criterio múltiple desarrollado por Lefkovitch (1985) utiliza un algoritmo que permite efectuar más de una medida de las relaciones existentes entre cada par de genotipos. Definió la disimilitud de genotipos por medio de tres medidas:

- Media marginal,
- Varianza entre ambientes y,
- Distancia entre ambientes.

Esta alternativa se basa en que, para construir grupos homogéneos, los genotipos pueden ser agrupados usando todas esas medidas secuenciales o simultáneamente en un procedimiento de agrupación condicional (Lefkovitch, 1980, 1982). Al contrario que el criterio único, esta descomposición proporciona un medio de análisis para investigar cada medida por separado.

Frecuentemente, la decisión sobre el valor de un cultivar por parte del mejorador se basa en la comparación de dicho cultivar con uno o más testigos. Lin y Binns (1985) sugieren el estudio de las interacciones entre los testigos y los cultivares ensayados para determinar qué genotipos, si no muestran interacción **GxE**, pueden ser directamente comparados con los





testigos, ignorando la variabilidad entre ambientes. Una vez comprobado que los cultivares no muestran interacción **GxE** significativa con ninguno de los testigos conocidos, mediante la utilización de una prueba **t** se pueden agrupar aquellos cultivares que o son superiores, o inferiores o estadísticamente iguales al testigo.

Esta metodología no está exenta de críticas. Así:

1) Existen gran número de algoritmos de clasificación, y cada uno de ellos puede producir diferentes agrupaciones (dendrogramas).

2) El nivel de corte en el dendrograma para obtener grupos de cultivares, puede ser decidido arbitrariamente.

3) Existen diferentes medidas de disimilitud (Lin y col., 1986, mencionaron nueve), obteniéndose resultados distintos con cada una de ellas, y

4) puede conducir al establecimiento de grupos difíciles de justificar y, por consiguiente, traducirse en resultados erróneos.

Como Cormack (1971) dijo: **“en algunas situaciones es mejor no efectuar el análisis de grupos y dejar al criterio del investigador el establecimiento o no de tales grupos”**.

1.4.6. Coordenadas Principales

El Análisis de Coordenadas Principales fue desarrollado por Schoenberg (1935), Young y Householder (1938) y Torgerson (1952). Fue popularizado por Gower (1966), el cual lo relacionó con otros métodos estadísticos. Es una generalización del Análisis de Componentes Principales en el cual se utiliza una medida de similitud entre los individuos. Sus objetivos y limitaciones son similares. Parte de una matriz de similitud para obtener una representación geométrica de los individuos basada en una distancia (*distancia de Gower*), razonablemente compatible con la similitud (Cuadras, 1981).

Westcott (1987) propuso un método gráfico para apreciar estabilidad genotípica. En él, el Análisis de Coordenadas Principales está

basado en una adecuada medida de similitud entre genotipos. La base del método es utilizar el Análisis de Coordenadas Principales sobre ciclos (L y H) de series ambientales. Para ello, los ambientes se dividen en dos grupos: i) aquellos cuya media marginal no supera a la media total del ensayo y ii) aquellos que la sobrepasan. Los primeros forman progresivamente lo que se denomina ciclos L (por ejemplo, ambientes de bajo rendimiento, si es éste el carácter medido) y los segundos constituyen lo que se denomina ciclos H (ambientes de alto rendimiento). Para cada ciclo, tanto L como H, se realiza una representación gráfica de las primeras dos coordenadas principales. Con las sucesivas representaciones se pone de manifiesto directa e inmediatamente los buenos genotipos, mientras que se desecha el resto.

Este método tiene varias ventajas:

- Es fiable cuando en los datos incluimos ambientes extremos (de alto o bajo rendimiento),
- No depende del conjunto de genotipos incluidos en el análisis,
- Es muy fácil identificar los genotipos más estables en la secuencia de gráficas resultantes.

1.4.7. Análisis de Correspondencia

De una manera general puede considerarse que los objetivos que se persiguen cuando se aplica el Análisis de Correspondencias son similares a aquellos perseguidos con la aplicación del Análisis de Componentes Principales, y pueden resumirse en los dos puntos siguientes:

- Estudio de las relaciones existentes entre los genotipos entre sí y entre los ambientes entre sí.
- Estudio de la relaciones entre los genotipos por un lado y los ambientes por otro.

La elección entre el Análisis de Componentes Principales y el de Correspondencias dependerá del tipo matriz de datos (filas=genotipos, columnas=ambientes) que se desee analizar. Así como el Análisis de Componentes Principales es útil para datos cuantitativos, el Análisis de Correspondencias es apropiado para representar tablas de frecuencias, aunque en la actualidad se ha empezado a utilizar en el estudio de la Interacción GxE.

El Análisis de Correspondencias puede considerarse como un caso particular del Análisis de Componentes Principales aunque con propiedades que permiten una mejor representación de la estructura de ciertas tablas de datos. Como peculiaridad hay que señalar que en el Análisis de Correspondencias la distancia que se define entre puntos no es una distancia euclídea clásica, como en el ACP, etc., sino una *distancia Chi-Cuadrado*.

1.4.8. Análisis Canónico

Las técnicas de representación de datos desarrolladas en los apartados anteriores se han basado en distancias como la euclídea, la distancia de Gower, la distancia chi-cuadrado, etc. En el Análisis Canónico se utiliza la *distancia de Mahalanobis D*.

El Análisis Canónico es un método de *representación de grupos* a lo largo de ejes con máximo poder de discriminación, en relación a la distancia D . El Análisis Canónico resume la *variación entre grupos* de forma parecida a como el Análisis de Componentes Principales resume la variación total.

El Análisis Canónico se utiliza generalmente después de un Análisis de Grupos para cerciorarse de la fiabilidad existente en la formación de dichos grupos. En tanto que el Análisis de Grupos clasificaba los genotipos de acuerdo con su semejanza fenética, el Análisis Canónico empezando con unos grupos ya definidos, procede a investigar las interrelaciones entre estos grupos.

1.4.9. Otros métodos

Dominancia estocástica. La Dominancia estocástica es una técnica poco usada en el análisis de la interacción y en la estabilidad genotípica. Los agricultores pueden considerar nuevas tecnologías en general y a las nuevas variedades en particular como de mayor riesgo que las tradicionales. El riesgo puede entonces actuar como un impedimento para su adopción. Se han propuesto gran cantidad de parámetros univariantes de estabilidad para identificar genotipos de gran adaptabilidad. Muchos de estos parámetros han sido desarrollados con el interés implícito de que los mejoradores puedan utilizar el rendimiento genotípico y la estabilidad al efectuar su selección. Pero en muy pocos se ha indicado explícitamente como desarrollar un índice

basado en ambos (rendimiento genotípico y estabilidad) que pueda ayudar al mejorador en la selección. Consecuentemente, los mejoradores le han dado más importancia, en el desarrollo de estos índices, bien al rendimiento, o bien a la estabilidad según sus propios criterios. Si la estabilidad la podemos expresar como una *medida de variabilidad o de incertidumbre*, entonces las técnicas adecuadas para tomar decisiones bajo *condiciones de riesgo* pueden utilizarse para desarrollar estos índices de estabilidad que ponderen la importancia tanto del rendimiento como de la estabilidad.

El proceso de Dominancia Estocástica puede identificar, bajo ciertas condiciones, variedades que puedan ser preferidas, por su bajo riesgo, por los agricultores. Es una técnica que puede ayudar a otros métodos analíticos para la interpretación de los datos en la mejora de plantas u otros ensayos.

El **Método “primera-seguridad”** de Eskridge, propuesto inicialmente por Kataoka (1963) como un medio de estimar cómo se toman decisiones en inversiones financieras cuando existe el riesgo de que deriven a, por ejemplo, bancarota y aplicado a problemas de mejora por Eskridge (1990), está basado sobre una probabilidad aceptable para el mejorador de que *no ocurra un suceso indeseable*, como, por ejemplo, la elección de cultivares que tengan una alta probabilidad de producir bajos rendimientos. Esta probabilidad se utiliza para calcular un intervalo de confianza que representa los niveles de rendimiento para los cuales los rendimientos muy bajos sólo se tendrían con un $\alpha\%$ de probabilidad.

La importancia de éste método es que los modelos de decisión resultantes pueden ser utilizados para desarrollar índices que cuantifiquen en cuánto ha valorado un mejorador el rendimiento y la estabilidad cuando está desarrollando cultivares para un amplio intervalo de ambientes. El mejorador tendrá primero que especificar una probabilidad razonable (α) de que ocurra un rendimiento bajo (así, 1 de cada 20, es decir, $\alpha=0.05$). Utilizando éste α se calcula un intervalo de confianza para cada genotipo. Este intervalo representa el nivel de rendimiento que se espera si ocurre el desastre de obtener, por alguna causa, rendimientos bajos para un nivel de probabilidad del $\alpha\%$. Los genotipos con mayor valor del extremo superior de este intervalo serán los seleccionados.



Un importante resultado de los ensayos multilocales es la construcción de una tabla de dos vías donde se sitúan los rendimientos estimados de cada genotipo en varios ambientes. Esta tabla es la base para el estudio de la interacción genotipo ambiente. El estimador más natural y más utilizado para la construcción de estas tablas GxE es la media entre repeticiones para cada genotipo en cada ambiente.

Se ha demostrado, sin embargo, que existen otros estimadores que pueden ser predictivamente más precisos que la media; por ejemplo, los estimadores basados en modelos multiplicativos como el AMMI (Gauch, 1992) y los factores de reducción (Cornelius et al 1994, 1996), y el mejor predictor lineal insesgado (BLUP) basado en un análisis de varianza de dos vías (Cornelius et al 1994, 1996; Piepho, 1994).

En el AMMI tanto “genotipos” como “ambientes” son considerados como factores fijos (Piepho, 1994). Sin embargo, según Stroup y Multize (1991) cuando el número de genotipos es lo suficientemente grande debería considerarse como factor aleatorio, pero si el factor “genotipo” es aleatorio, la interacción GxE también lo será. Stroup y Multize (1991) consideran que si, en un ensayo de variedades, estas están entre 20 y 100, el BLUP es más eficaz que el BLUE. Piepho (1994) aclara que considerar el factor “genotipo” como aleatorio no necesariamente elimina nuestro interés de estimar el rendimiento de genotipos concretos.

Piepho (1994) concluyó que las estimaciones obtenidas con el BLUP son predictivamente más precisas que las obtenidas por el modelo AMMI, y en Piepho (1998) completó el análisis señalando que el BLUP era ligeramente superior a los factores de reducción de Cornelius et al (1996), aunque Cornelius y Crossa (1999) concluyeron justamente lo contrario.

Más usual es considerar los efectos ambientales como aleatorios y los efectos genotípicos como fijos, y por esta razón Henderson (1975) recomendó los modelos mixtos que proporcionan el mejor estimador lineal insesgado (BLUE) para genotipo (efecto fijo) y el mejor predictor lineal insesgado (BLUP) para los efectos aleatorios (Kang, 1998). Según Piepho (1999) la mayoría de los modelos estadísticos para el estudio de la interacción GxE pueden expresarse por medio de modelos lineales mixtos, donde “ambientes” se considera como factor aleatorio y

“genotipos” como fijo. Los parámetros de este tipo de modelos pueden estimarse por máxima verosimilitud restringida (REML) (Denis et al. 1997; Piepho, 1998). Una de las mayores ventajas de aplicar la metodología REML es su utilidad para el manejo de datos no equilibrados. Piepho (1999) propone utilizar el procedimiento MIXED del SAS (SAS, 1997) para la resolución de los principales modelos lineales utilizados en el estudio de la interacción GxE, y ofrece las corrientes de control en SAS correspondientes al modelo de la varianza de estabilidad de Shukla (1972), a los modelos de regresión de Finlay-Wilkinson (1963) y de Eberhart-Russel (1966), al de la varianza ambiental (Becker y Leon, 1988) y finalmente, a un modelo multiplicativo como el AMMI (Gauch, 1988).

Cuando se poseen datos adicionales de variables ambientales o genotípicas como datos meteorológicos, análisis de suelo, reacciones a enfermedades, días a floración, marcadores moleculares, etc, la regresión factorial (Denis, 1988; van Eeuwijk et al 1996) puede utilizarse para determinar la influencia de estas covariables sobre la interacción GxE. Los modelos de regresión factorial (RF) son modelos lineales que explican la interacción GxE en función de las respuestas diferenciales de los genotipos a las distintas covariables ambientales o genotípicas incluidas en el modelo; según Vargas et al (1999), tiene la ventaja de poder probar estadísticamente la influencia de estas covariables sobre la interacción GxE del rendimiento. Los modelos RF son también modelos multiplicativos como el AMMI; la principal diferencia es que los primeros se construyen directamente como una función de covariables ambientales o genotípicas. Cuando ambos tipos de covariables se integran simultáneamente en el modelo de RF se realiza una regresión múltiple paso a paso para cada covariable, siendo, según Vargas et al (1999), una herramienta útil para seleccionar las más importantes covariables que explican la interacción GxE.

Cuando realizamos un análisis de QTLs con datos en diferentes localidades, puede existir interacción GxE cuando los alelos favorables sólo se muestran en algunos ambientes. Romagosa et al (1996) estudiaron la utilidad del modelo AMMI para identificar QTLs con un valor suficiente para poderse introducir en un esquema de selección ayudada por marcadores, bajo la hipótesis de que si la adaptación tiene base genética se podrían identificar aquellas regiones del genoma



responsables de la expresión genotípica diferencial en diferentes ambientes por medio del mapeo de QTLs basado en las coordenadas genotípicas del AMMI. Concluyeron que cuando pueda mapearse QTLs en una población, las coordenadas genotípicas resultantes del análisis AMMI constituyen un método adecuado para el estudio de la naturaleza y de la magnitud de las interacciones QTL x ambiente.



Yan et al (2000) desarrollaron un método denominado "GGE biplot" para el análisis gráfico de los datos procedentes de un ensayo multi-ambiental (tómese ambiente como año, localidad, combinación año-localidad, etc.). Las letras "GGE" se refieren a que en el modelo se unen el efecto principal de genotipos (G) más el efecto multiplicativo de la interacción GxE (GE), siendo estos dos efectos las dos fuentes de variación más relevantes en la evaluación del comportamiento de distintos genotipos en diferentes ambientes. El "GGE biplot" se construye dibujando los dos primeros componentes derivados de una descomposición singular (SVD) de la matriz de datos ambientalmente centrada. Este método desde su difusión por Yan et al (2000) ha recibido numerosos comentarios positivos de reconocidos biométricos como Crossa, Gauch, Hunt, etc. De ahí que le dediquemos un nuevo capítulo en esta segunda edición de este Manual.



1.5. Resumen general sobre los métodos expuestos

Obtener datos de ensayos en múltiples localidades ayuda a los investigadores a estimar rendimientos con más precisión y seleccionar los mejores genotipos.

Para genetistas, mejoradores y agronomistas, los estadísticos paramétricos de estabilidad, obtenidos por regresión lineal, son matemática-



mente muy simples y fácilmente interpretables desde un punto de vista biológico. Sin embargo, éste método tiene graves inconvenientes mencionados en un apartado anterior de este capítulo. De ahí que se deba sustituir, en la mayor parte de los casos, por otros métodos, ya sean univariantes paramétricos, univariantes no-paramétricos o multivariantes.

Se ha utilizado un amplio número de métodos multivariantes para analizar los ensayos multilocales y estimar la estabilidad del rendimiento. En algunos de ellos se superan las limitaciones de la regresión lineal, pero los resultados son, a menudo, difíciles de interpretar en relación a la interacción **GxE** (como es el caso del Análisis de Componentes Principales y el Análisis de Grupos).

Ciertas técnicas multivariantes, o combinaciones de las mismas, ofrecen una información biológica relevante y son estadísticamente muy simples (por ejemplo, el AMMI).

La combinación del Análisis de Varianza y Componentes Principales en el modelo AMMI, junto con las estimas predictivas, hacen que éste sea un método muy válido para explicar la estructura de la interacción y estimar con más precisión los rendimientos, pero se necesita una investigación más intensa para cuantificar la probabilidad de acertar en la selección de un genotipo utilizando el valor **predictivo** del AMMI, comparado con los valores **predictivos** de otros modelos.

Evidentemente, se debe prestar la mayor atención a las colecciones de material y a los análisis e interpretación de variables ambientales y fisiológicas. Estas ayudarían a caracterizar más particularmente tanto a los genotipos como a las áreas geográficas, y así poder explicar mejor ciertos aspectos de la interacción.



REFERENCIAS CITADAS

- BECKER, H.C. 1981:Correlations among some statistical measures of phenotypic stability. *Euphytica* 30, 835-840.
- BECKER, H.C. and J. LEON. 1988: Stability Analysis in Plant Breeding. *Plant Breeding* 101, 1-23.
- CORMACK, R.M. 1971: A Review of classification. *Journal of the royal stat. society.* A134, 321-367.
- CORNELIUS P. L. and J. CROSSA. 1999. Prediction Assessment of Shrinkage Estimators of Multiplicative Models for Multi-Environment Cultivar Trials. *Crop Science*, 39:998-1009.
- CORNELIUS, P.L., CROSSA, J. and M.S. SEYEDSADR. 1994. Test and estimators of multiplicative models for variety trials. *Proc 1993 Kansas State Univ. Conf. On Applied Statistics in Agriculture.* Manhattan, Kansas, pp. 156-169.
- CORNELIUS, P.L., CROSSA, J. and M.S. SEYEDSADR. 1996. Statistical tests and estimators of multiplicative models for genotype by environment data. In: Kang M.S., Gauch, H.G. Jr. (eds) *Genotype by environment interaction.* CRC Press, Boca Raton, pp. 199-234.
- CUADRAS, C.M. 1981: Métodos de análisis multivariante. Ed. por Eunibar. Barcelona.
- DENIS, J.B. 1988. Two-way analysis using covariates. *Statistics* 19: 123-132.
- DENIS, J.B., PIEPHO, H.P. and F.A. VAN EEUWIJK. 1997. Modelling expectation and variance for genotype by environment data. *Heredity* 79:162-171.
- DORTS, J.C. 1957: Adaptation. *Euphytica* 6, 247-254.
- EAGLES, H.A., P.N. HINZ, and K.J. FREY. 1977: Selection of superior cultivars of oats by using regression coefficients. *Crop Sci.* 17, 101-105.
- EBERHART, S.A. and W.A. RUSSELL. 1966: Stability Parameters for Comparing Varieties. *Crop Sci.* 6, 36-40.
- ESKRIDGE, K.M. 1990: Selection of stable cultivars using a safety-first rule. *Crop Sci.* 30, 369-374.
- FINLAY, K.W. and G.N. WILKINSON, 1963: The analysis of adaptation in a plant breeding programme. *Aust. J. Agric. Res.* 14, 742-754.
- FOX, P.N., B. SKOVMAND, B.K. THOMPSON, H.J. BRAUN and R. CORMIER. 1990: Yield and adaptation of hexaploid spring triticale. *Euphytica* 47, 57-64.
- FREEMAN, G.H. 1973: Statistical methods for the analysis of genotype-environment interactions. *Heredity* 31 (3), 339-354.
- FREEMAN, G.H. and J.M. PERKINS. 1971: Environmental and genotype-environmental components of variability. VIII. *Heredity* 27, 15-23.
- GABRIEL, K.R. 1971: The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453-467.
- GAUCH, H.G. 1985: Mimeo 85-7. Cornell Univ. Ithaca, New York.
- GAUCH, H.G. 1988: Model selection and validation for yield trials with interactions. *Biometrics* 44, 705-715.
- GAUCH, H.G. Jr. 1992. *Statistical analysis of regional yield trials.* Elsevier, Amsterdam.
- GOWER, J.C. 1966: Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 3, 325-338.
- HALDANE, J.B.S. 1946: The interaction of nature and nurture. *Ann. Eugenics.* 13, 197-205.
- HENDERSON, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31: 423-447.
- HILL, R.R. Jr., and J.E. BAYLOR. 1983: Genotype x environment interaction analysis for yield in alfalfa. *Crop Sci.* 23, 811-815.
- HILL. M. O. 1974: Correspondence Analysis: A Neglected Multivariate Method. *Applied statistics* 23, 340-354.

• HÜHN, VON M. 1979: Beiträge zur erfassung der phänotypischen stabilität. EDV in Medizin und Biologie 10 (4), S. 112-117.

• KANG, M.S. 1988: A Rank-sum method for selectig high -yielding, stable corn genotypes. Cereal Res. Comm.16, 113-115.

• KANG M.S. 1998. Using genotype by environment interaction for crop cultivar development, Adv. Agron. 62:199-252.

• KATAOKA, S. 1963: A stochastic programming model. Econometrica 31, 181-196.

• KETATA, H., S.K. YAN and M. NACHIT. 1989: Relative consistency performance across environments. Int. Sym. on Physiology and Breeding of Winter Cereals for stressed Mediterranean Environments. Montpellier. July 3-6, 1989.

• LEFKOVITCH, L.P. 1985: Multi-criteria clustering in genotype-environment interaction problems. Theor. Appl. Genet. 70, 585-589.

• LEFKOVITH, L.P. 1982: Conditional cluster, muster, and probability. Math. Biosci. 60, 207-234.

• LEFKOVITH, L.P. 1980: Conditional clustering. Biometrics 36, 43-58.

• LEON, J. 1985: Beiträge zur erfassung der phänotypischen stabilität unter besonderer berücksichtigung unterschiedlincher heterogenitäts- und heterozygotiegrade sowie einer zusammenfassenden beurteilung von ertragshöhe und ertragssicherheit. Dissertation, Christian-Albrechts-Universität Kiel.

• LERNER, I.M. 1954: Genetic homeostasis. Edinbourg, London: Oliver and Boyd.

• LIN, C.S. and M.R. BINNS. 1985: Procedural approach for assessing cultivar-location data: pairwise genotype-environment interactions of test cultivars with checks. Can. J. Plant Sci. 65:1065-1071.

• LIN, C.S., M.R. BINNS, and L.P. LEFKOVITCH, 1986: Stability analysis: Where do we stand?.Crop Sci. 26, 894-900.

• LIN, C.S. and M.R. BINNS. 1988: A superiority measure of cultivar performance for cultivar x location data. Can. J. Plant Sci. 68, 193-198.

• LIN, C.S., and M.R. BINNS, 1989: Comparison of unpredictable environmental variation generated by year and by seeding-time factors for measuring type 4 stability. Theor. Appl. Genet. 78, 61-64.

• MANDEL, J. 1961: Non-additivity in two-way analysis of variance. J. Amer. Statist. Assoc. 56, 878-888.

• MANDEL, J. 1971: A new analysis of variance model for non-additive data. 13, 1-18.

• MOOERS, C.A. 1921: The agronomic placement of varieties. J. Amer. Soc. Agronomy 13, 337-352.

• NASSAR, R., and M. HÜHN, 1987: Studies on estimation of phenotypic stability: Test of significance for nonparametric measures of phenotypic stability. Biomet. 43, 45-53.

• PEARSON, K. 1901. Philos. Mag. 2, 559-572.

• PERKINS, J.M. and J.L. JINKS. 1968a: Environmental and genotype-environmental components of variability. III. Heredity 23, 339-356.

• PERKINS, J.M. and J.L. JINKS. 1968b: Environmental and genotype-environmental components of variability. IV. Heredity 23 (4), 525-535.

• PIEPHO, H.P. 1994. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects multiplicative interaction (AMMI) analysis. Theor. Appl. Genet. 89: 647-654.

• PIEPHO, H.P. 1998. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. Theor. Appl. Genet. 97:195-201.

• PIEPHO, H.P. 1999. Stability analysis using the SAS system. Agronomy Journal, 91:154-160.

• PINTHUS, M.J. 1973: Estimate of genotypic value: A proposed method. Euphytica 22, 121-123.

• PLAISTED, R.L. and L.C. PETERSON. 1959: A technique for evaluating the ability of selections to yield consistently in different locations or seasons. Amer. Potato J. 36, 381-385.



• PLAISTED, R.L. 1960: A shorter method for evaluating the ability of selection to yield consistently over locations. *Amer. Potato J.* 37, 166-172.

• ROEMER, T. 1917: Sind die ertragsreichen sorten ertragssicherer? *Mitt. DLG* 32, 87-89.

• ROMAGOSA, I., ULLRICH, S.E., PATRICK, F. HAN and P. M. HAYES. 1996. Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theor. Appl. Genet.* 93:30-37.

• SALMON, S.C. 1951: Analysis of variance and longtime variety tests of wheat. *Agron. J.* 43, 562-570.

• SAS Institute. 1997. SAS/STAT software: changes and enhancements through release 6.12. SAS Inst., Cary, NC.

• SCHOENBERG, I.J. 1935: Remarks to Maurice Frechet's article "Sur la définition axiomatique d'une classe d'espace distances vectoriellement applicable sur l'espace d'Hilbert. *Annals of Mathematics (2nd Series)* 36, 724-732.

• SHUKLA, G.K., 1972: Some Statistical aspects of partitioning genotype-environmental components of variability. *Heredity* 29, 237-245.

• SKROPPIA, T. 1984: A critical evaluation of methods available to estimate the genotypex-environment interaction. *Studia Forestalia Suecuca* 166, 3-14.

• STROUP, W.W. and D.K. MULITZE. 1991. Nearest neighbor adjusted best linear unbiased prediction. *Am. Stat.* 45:194-200.

• TORGERSON, W.S. 1952: Multidimensional scaling. I. Theory and method. *Psychometrika* 2, 267-276.

• TUKEY, J.W. 1949: One degree of freedom for non-additivity. *Biometrics* 5, 232-242.

• VAN EEUWIJK, F.A., DENIS, J.B. and M.S. KANG. 1996. Incorporating additional information on genotypes and environments in models

for two-way genotype by environment tables. p. 15-49. In M.S. Kang and H.G. Gauch (ed.) *Genotype-by-environment interaction*, CRC Press, Boca Raton, FL.

• VARGAS, M., CROSSA, J., VAN EEWIJK, F.A., RAMIREZ, M.E. and K. SAYRE. 1999. Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype x environment interaction. *Crop Sci.* 39:955-967.

• VERMA M.M., G.S. CHAHAL and B.R. MURTY. 1978: Limitations of Conventional Regression Analysis A Proposed Modification. *Theor. Appl. Genet.* 53, 89-91.

• VIRK, D.S., P.S. VIRK, B.K. MANGAT and G. HARINARAYANA. 1988: Linear and non-linear regression analysis of genotype x environment interactions in pearl millet. *Theor Appl. Genet.* 75, 736-740.

• WESTCOTT, B. 1987: A method of assessing the yield stability of crop genotypes. *J. Agric. Sci., Camb.* 108, 267-274.

• WILLIAMS, E.J. 1952: The interpretation of interactions in factorial experiments. *Biometrika* 39, 65-81.

• WRICKE, G. 1962: Uber eine methode zur erfassung der ökologischen streubreite in feldversuchen. *Z. Pflanzenzüchtg* 47, 92-96.

• WRICKE, G. 1964: Zur berechnung der okovalenz be sommerweizen und hafer. *Z. Pflanzenzüchtg* 52, 127-138.

• YAN, W., HUNT, L.A., SHENG, Q. and Z. SZLAVNICS. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* 40:597-605.

• YATES, F. and W.G. COCHRAN. 1938: The analysis of groups of experiments. *J. Agr. Sci.*, 28, 556-580.

• YOUNG, G. and A.S. HOUSEHOLDER. 1938: Discursion of a set points in terms of mutual distances. *Psychometrika* 3, 19-22.

III. PARTE

ANÁLISIS ESTADÍSTICOS

Σ α Π Ω \times \pm ϑ

\pm \times \pm ϑ Ω \times α \pm

ϑ Ω \times \pm ϑ Σ α

Π Ω \times \pm ϑ Σ

Ω \times \pm ϑ α Π

α \pm ϑ Σ α \times

Ω \times \pm ϑ Σ α \pm

ϑ Σ α Π Ω \times Σ

2

ANÁLISIS DE LA VARIANZA Y REGRESIÓN CONJUNTA

2. ANÁLISIS DE LA VARIANZA Y REGRESIÓN CONJUNTA

2.1. Datos

Se parte de un ensayo de 11 genotipos (G1 a G11) en 10 ambientes (A1 a A10) con un diseño de bloques al azar con tres repeticiones (ver los datos reales en el Apéndice).

2.2. Análisis de la varianza y regresión conjunta

El primer paso en el análisis del comportamiento de una serie de genotipos en diferentes ambientes es la identificación estadística de una interacción **GxE** significativa. Esto requiere llevar a cabo un análisis de la varianza de los ensayos en estudio. Este análisis sólo será posible si las varianzas del error (ϵ_{ijk}) (es decir, las varianzas de las celdas que contienen el nivel inferior de datos) **no** son estadísticamente heterogéneas: recordemos que se trata de uno de los supuestos en que se fundamenta el análisis de la varianza.

2.2.1. Un paso previo: la homogeneidad de las varianzas

Cuando existen más de dos grupos, la **prueba de Barlett para homogeneidad de varianzas** es el más indicado. El estadístico final de significación es (χ^2); el valor experimental se compara, como es usual, con un valor crítico tabulado de la distribución chi-cuadrado (χ^2). El estadístico a calcular viene dado por la expresión:

$$\chi^2 = 2.3026 \left[\sum_1^a (n_i - 1) \log s^2 - \sum_1^a (n_i - 1) \log s_i^2 \right] \quad (1)$$

donde:

a = número de varianzas cuya homogeneidad se prueba.

$n_i - 1$ = grados de libertad de la varianza i-ésima.

s_i^2 = varianza i-ésima, y

$$s^2 = \frac{\sum_1^a (n_i - 1) s_i^2}{\sum_1^a (n_i - 1)} \quad (2)$$

El estadístico χ^2 así obtenido se divide por un factor de corrección C, cuyo valor es:

Por tanto, el valor del estadístico χ^2 ajustado será:

$$\chi^2 \text{ ajustado} = \chi^2 / C \quad (3)$$

el cual se distribuye mediante una $\chi^2_{[a-1]}$ cuando la hipótesis nula H_0 es verdadera, es decir, cuando todas las varianzas son homogéneas.

De esta forma si el valor χ^2 ajustado es mayor que $\chi^2_{[a-1]}$ concluiremos con que las **a** varianzas estudiadas son significativamente heterogéneas.

Desgraciadamente, el test Barlett es excesivamente sensible a las desviaciones de la normalidad en los datos y un χ^2 significativo puede, por tanto, indicar no-normalidad en vez de heteroscedasticidad.

Existe también un método rápido que, aunque no tan eficiente, presenta la ventaja de su sencillez. Este es la **prueba de la Fmax** que se basa en la distribución de probabilidad acumulativa del cociente entre las varianzas mayor y menor. Se hallan las varianzas mayor, s^2_{max} , y la menor, s^2_{min} . Se calcula entonces el cociente de varianzas máximo s^2_{max}/s^2_{min} . Existe una Tabla de la distribución acumulativa de $F_{max, \alpha[a, n-1]}$ (ver Tablas en Steel y Torrie, 1985); esta Tabla supone que los grados de libertad de todas las varianzas son iguales, pero podemos hacer una prueba aproximada utilizando el **menor** de los grados de libertad de las dos varianzas utilizadas en el cálculo.

Si el cociente de varianzas es mayor que el estadístico F_{max} hallado en la Tabla entonces las varianzas se considerarían heterogéneas con probabilidad $P < \alpha$.

2.2.2. El Análisis de la Varianza

El modelo matemático del análisis de varianza de los datos observados es, en nuestro caso, el siguiente: sean **g** genotipos en **e** ambientes, y **r** el número de observaciones de cada genotipo en cada ambiente,

$$Y_{ijk} = \mu + g_i + e_j + (ge)_{ij} + r_{jk} + \epsilon_{ijk} \quad (4)$$

donde:

Y_{ijk} = el valor fenotípico del genotipo i-ésimo en la k-ésima repetición del j-ésimo ambiente,

μ = media total de todas las observaciones, genotipos y ambientes,





g_i = efecto aditivo del genotipo i -ésimo ($i=1, \dots, g$),

e_j = efecto aditivo del ambiente j -ésimo ($j=1, \dots, e$),

$(ge)_{ij}$ = interacción **GxE** del genotipo i -ésimo y del ambiente j -ésimo,

r_{jk} = efecto de la observación k -ésima en el ambiente j -ésimo, ($k=1, \dots, r$),

ϵ_{ijk} = variación residual de la observación k -ésima del genotipo i -ésimo y del ambiente j -ésimo.



Se trata, de nuevo en nuestro caso, de un modelo mixto de dos factores con genotipo como factor fijo (pues se está interesado en esas variedades concretas) y ambiente y repetición como aleatorios. El análisis de la varianza para el modelo mixto [4] es como se muestra en Tabla 1. Otras situaciones tendrán lógicamente un modelo matemático [4] y un análisis (Tabla 1) distintos. El análisis numérico se da en la Tabla 3. Para realizarlo se puede utilizar el paquete estadístico SX, o si el diseño es muy desequilibrado (es decir, distinto número de observaciones por factor, o dicho de otra forma, existencia de valores perdidos o no tomados) el procedimiento GLM del paquete estadístico SAS.

2.2.3. La interacción **GxE**

El análisis anterior es el más simple que podemos realizar y con el determinamos algo tremendamente importante: la *existencia o no existencia* de una interacción Genotipo-Ambiente (GxE) signifi-

cativa. A partir de aquí describiremos todos aquellos métodos más utilizados para el estudio de la interacción GxE.

El primero de ellos, y por ende el más utilizado, es el método de la Regresión, conocido como **análisis de la varianza y regresión conjunta**. Este análisis utiliza el análisis de la varianza tal y como hemos visto anteriormente, pero nos descompone la varianza total de la interacción **GxE** en dos partes, una debida a la heterogeneidad de los coeficientes de regresión y otra a las desviaciones de estas regresiones. Tal análisis se expone en la Tabla 2 y los resultados obtenidos para los datos del ejemplo en la Tabla 3.

En el método de la regresión se asume que la interacción está linealmente relacionada con el efecto ambiental, es decir:

$$(ge)_{ij} = b_i e_j + d_{ij} \tag{5}$$

donde b_i es el *coeficiente de regresión lineal* para el genotipo i y d_{ij} un residuo.

De esta forma el modelo lineal [4] queda como sigue:

$$Y_{ijk} = \mu + g_i + b_i e_j + r_{jk} + \epsilon'_{ijk} \tag{6}$$

El análisis de la varianza para nuestro modelo mixto se muestra en la Tabla 2.

Como complemento al coeficiente de regresión (b_i), *el cuadrado medio residual de las desviaciones de la regresión* (S^2_{di}) describe la contribución del genotipo i a la interacción **GxE**, propuesta por Eberhart y Russell (1966). Ambos parámetros se calculan como sigue:

Tabla 1. ANOVA de la interacción **GxE**.

F. de V.	G.L.	S.C.	F
Ambientes(E)	e-1	$\sum_j Y^2_{.j} / rg - FC$	CM(E)/CM(RE)
Rep./Amb.(RE)	e(r-1)	$\sum_{jk} Y^2_{.jk} / g - SC(E) - FC$	
Genotipos(G)	g-1	$\sum_i Y^2_{i.} / re - FC$	CM(G)/CM(GxE)
Genot x Amb(GxE)	(g-1)(e-1)	$\sum_{ij} Y^2_{ij} / r - SC(G) - SC(E) - FC$	CM(GxE)/CM(Er)
Error(Er)	e(g-1)(r-1)	$\sum_{ijk} Y^2_{ijk} - \sum_{ij} Y^2_{ij} / r - SC(RE)$	
Total	egr-1	$\sum_{ijk} Y^2_{ijk} - FC$	

FC = $\sum_{...} Y^2_{...} / egr$

Los **CM** se obtienen dividiendo las **SC** por los **g.l.** correspondientes.



Tabla 2. ANOVA Y REGRESIÓN CONJUNTA de la interacción GxE.

F. de V.	G.L.	S.C.	F
Ambientes(E)	e-1	$\sum_j Y_{.j}^2 / rg - FC$	CM(E)/CM(RE)
Rep./Amb.(RE)	e(r-1)	$\sum_{jk} \sum Y_{.jk}^2 / g - SC(E) - FC$	
Genotipos(G)	g-1	$\sum_i Y_{i..}^2 / re - FC$	CM(G)/CM(GxE)
Genot x Amb(GxE)	(g-1)(e-1)	$\sum_{ij} \sum Y_{ij.}^2 / r - SC(G) - SC(E) - FC$	CM(GxE)/CM(Er)
Heterog. (H)	g-1	$\sum_j (\beta_j - 1)^2 \sum Z_j^2$	CM(H)/CM(D)
Desv.Regres(D)	(g-1)(e-2)	Por sustracción de la suma SC(GxE)	CM(D)/CM(Er)
Error(Er)	e(g-1)(r-1)	$\sum_{ijk} \sum Y_{ijk}^2 - \sum Y_{ij.}^2 / r - SC(RE)$	
Total	egr-1	$\sum_{ijk} \sum Y_{ijk}^2 - FC$	

FC= $\sum Y_{..}^2 / egr$

$Z_j = Y_{.j} / rg - Y_{..} / rge$

$\beta_j = (\sum Y_{ij} / r * Z_j) / \sum Z_j^2$

Los CM se obtienen dividiendo las SC por sus g.l.

- coeficiente de regresión:

$$b_i = 1 + \frac{\sum_j (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}) (\bar{Y}_{.j} + \bar{Y}_{..})}{\sum_j (Y_{.j} - \bar{Y}_{..})} \quad (7)$$

- cuadrado medio residual:

$$s^2_{di} = \frac{1}{E-2} [\sum_j (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2 - (b_i - 1)^2 \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2] \quad (8)$$

El cuadrado medio residual para cada línea se compara por medio de una prueba **F** utilizando el valor aproximado $F^* = \sigma^2_i / \sigma^2_o$, con (e-2) y $e * g * (r-1)$ grados de libertad, siendo σ^2_o la varianza del error dentro de ambientes. Es decir, en la Tabla 1, σ^2_o sería $[SC(RE) + SC(Er)] / [g.l.(RE) + g.l.(Er)]$ y $\sigma^2_i = S^2_{di}$. Si F^* resulta **no significativo** nos indica que S^2_{di} se puede considerar no significativamente distinta de cero.

Cuando la prueba de Bartlett nos ha dado significativa y, por tanto, las varianzas del error de los ambientes son heterogéneas, el cuadrado medio de la desviación de la regresión se puede utilizar como error en aquellos contrastes en que los datos de partida sean las medias genotípicas en cada ambiente. Dado que los parámetros presentados hasta ahora han sido calculados a partir de los datos individuales (repeticiones) y no de medias genotípicas se ha optado aquí por utilizar σ^2_e

(varianza del error) en los contrastes en lugar de los cuadrados medios de la desviación de la regresión.

Se puede extraer un componente más, con un sólo grado de libertad, para probar la tendencia de las líneas de regresión a converger en un punto. En efecto, la suma de cuadrados de la heterogeneidad entre las regresiones lineales se puede dividir en dos componentes, uno debido a la convergencia de las líneas de regresión en un punto, con un grado de libertad, y otro debido a la no-convergencia.

La fórmula utilizada para calcular la suma de cuadrados para convergencia es:

$$SS_{conv} = R^2 H^2 \quad (9)$$

donde R es la correlación entre los coeficientes de regresión y las respectivas medias marginales de los genotipos (es decir, las medias genotípicas en cada ambiente), y H^2 es la suma de cuadrados debida a heterogeneidad de las regresiones.

2.2.4. Resolución práctica

Para la resolución del ejemplo que a continuación se expone se ha utilizado una "corriente de control" o secuencia de órdenes reconocibles por el paquete estadístico que se está utilizando (ver Apéndice) elaborada por Romagosa para el paquete estadístico SAS.



La prueba de Bartlett ($\chi^2=65.5$) y $F_{\max}=6.838$ para homogeneidad de varianzas para los 10 ambientes indicó que los errores experimentales eran heterogéneos. En estos casos lo aconsejable es encontrar una transformación de los datos que baje el valor de χ^2 en la prueba de Bartlett. Para ello se pueden ensayar varios tipos de transformaciones; las más comunes son la raíz cuadrada de Y_{ijk} , esto es, de cada valor experimental (adecuada cuando los datos consisten en números enteros pequeños), el log Y_{ijk} (cuando la media esté positivamente correlacionada con la varianza, es decir, a mayores medias corresponden mayores varianzas) y la angular o arco-seno de la raíz cuadrada de Y_{ijk} (especialmente apropiada para porcentajes y proporciones). Los resultados obtenidos con las dos primeras transformaciones han sido los siguientes: 1) raíz cuadrada: $\chi^2=74.49$, $F_{\max}=6.629$, y 2) logarítmica: $\chi^2=76.34$, $F_{\max}=6.652$. Puede verse que las transformaciones resultan en poca o ninguna mejora de la homogeneidad.

La principal consecuencia de que las varianzas del error sean heterogéneas es que el estadístico F no se distribuye según una F de Snedecor, lo que afecta a la determinación de las significaciones estadísticas en el Análisis de la Varianza, lo que puede ser un serio problema si la significación de la interacción **GxE** fuese marginal (es decir, rondase los valores del 5% o del 1%). En nuestro caso, la gran significación de la interacción **GxE** ha permitido el análisis de la varianza para todos los ambientes utilizando los datos originales, aunque siempre es conveniente resaltar, a efectos de posteriores estudios de otros autores, el hecho de que las varianzas del error son heterogéneas.

De este modo, para completar la Tabla 3 se ha procedido de la siguiente forma:

1º) Si se posee el paquete estadístico SAS existe una "corriente de control" diseñada por Romagosa (ver Apéndice).

2º) Si no se posee el paquete estadístico SAS se pueden utilizar otros paquetes estadísticos menos complicados, por ejemplo, el SX, en el cual todas las aplicaciones vienen dadas por medio de ventanas de menús. Así, con el SX, el cálculo de la suma de cuadrados de la heterogeneidad entre regresiones lineales se realiza obteniendo las respectivas regresiones entre los valores de interacción de cada genotipo y los índices ambientales. De esta forma, con este paquete estadístico, los pasos a seguir serían:

2.1) A partir de la matriz de datos **A** 10x11 (ver Apéndice) se obtiene la matriz de interacción **INT** (ver Apéndice) sin más que restar a cada dato de la matriz **A** las respectivas medias marginales y sumarle la media general, así, por ejemplo, para cualquier dato de **A** (Y_{ij}) obtenemos el correspondiente dato de la matriz **INT** (Y^*_{ij}) de la siguiente forma: $Y^*_{ij} = Y_{ij} - Y_{.j} - Y_{i.} + Y_{..}$

2.2) Se forma la matriz compuesta por las columnas de la matriz de interacción **INT** a la cual se le añade una nueva columna: la de las medias ambientales; por tanto, sería una matriz 10x12: matriz **INTAMB** (ver Apéndice).

2.3) A partir de esta matriz, en el SX se pueden realizar las sucesivas regresiones entre las 11 primeras columnas (correspondientes a los 11 genotipos) y la última (correspondiente a la media total de todos los genotipos en cada ambiente), obteniéndose así, para cada genotipo, la suma de cuadrados debida a la regresión. Se suman todas las sumas de cuadrados debidas a la regresión para los 11 genotipos. Dicha suma se multiplica por el número de repeticiones del ensayo y obtenemos como resultado la suma de cuadrados del término que en la Tabla 3 aparece como Heterog (H).

La suma de cuadrados debida a la convergencia se calcula como hemos comentado anteriormente por medio de la fórmula [9].

Los resultados obtenidos para nuestro ejemplo (Tabla 3) y la explicación de los mismos se realiza a continuación.

Del estudio del ANOVA (Tabla 3), se desprende que el factor de variación Genotipo (**G**) es muy significativo ($F=12.68; P<0.0001$), al igual que el factor Ambiente (**E**). La interacción **GxE** resulta ser muy significativa ($F=3.21; P<0.0001$) por lo que se puede deducir que los factores **G** y **E** no son independientes, existiendo por consiguiente interacción entre ellos.

Por otra parte, la descomposición de la SC total pone de manifiesto que el efecto ambiental (**E**) es mucho mayor (41.37%) que el efecto genotípico (**G**) (24.48%) y el de su interacción **GxE** (17.37%).

Una vez comprobada la existencia de la interacción **GxE**, cabe preguntarse por la naturaleza

Tabla 3. ANOVA y análisis de regresión.

F. de V.	G.L.	S.C.(*)	R ²	F	Pr>F
Total	329	22751.6			
Ambiente	9	9412.6	41.37#	19.48	0.0001
Rep/Amb	20	1073.7	4.70#	3.92	0.0001
Genotipo	10	5571.2	24.48#	12.68	0.0001
GenxAmb	90	3953.5	17.37#	3.21	0.0001
Heterog.	10	708.5	17.92 ^a	1.74	0.0856
Conv.	1	316.4	44.65 ^b	7.26	0.0246
Desv. Conv.	9	392.2	55.35 ^b	3.18	0.0013
Desv. Reg.	80	3244.9	82.08 ^a	2.96	0.0001
Error	200	2740.5	12.04#		

(*) S.C por 10⁴. NOTA: CM=SC/GL ; CM error = 137,025.

tanto por ciento que representa la SC respecto de la SC total.

^a tanto por ciento que representa la SC respecto de la SC(GxE).

^b tanto por ciento que representa la SC respecto de la SC(Heterog.).

de la misma. Para ello suponemos la existencia de una relación lineal entre la interacción **GxE** (ge_{ij}) y el valor aditivo ambiental (e_j) tal como expusimos en [5]. En la Tabla 3 se muestra la descomposición de la suma de cuadrados (SC) de la interacción **GxE** en una **componente lineal**, también llamada **heterogeneidad de las regresiones (Heterog.** en la Tabla 3) y en un componente no lineal o **residual (Desv. Reg.** en la Tabla 3).

Tras comparar ambos componentes con el error experimental [$F_{\text{Heterog}} = \text{CM}(\text{Heterog}) / \text{CM}(\text{Er}) = 5.17^{***}$; $F_{\text{Desv.Reg.}} = \text{CM}(\text{Desv.Reg.}) / \text{CM}(\text{Er}) = 2.96^{***}$] y, además, la componente lineal (Heterog.) con la desviación de la regresión (Desv. Reg.) [$F_{\text{Heterog}} = \text{CM}(\text{Heterog.}) / \text{CM}(\text{Desv.Reg.}) = 1.74$ N.S.], resulta que la heterogeneidad de regresiones es altamente significativa cuando se le compara con los cuadrados medios del error y no significativa respecto de los cuadrados medios de la desviación de la regresión, hecho que no sorprende dado que sólo el 17.92% de la interacción **GxE** (columna R², fila "Heterog." en Tabla 3) puede atribuirse a las regresiones lineales. El componente residual (82.08%), por contra, resulta ser muy significativo ($P < 0.0001$).

Como la heterogeneidad de regresiones resulta no significativa respecto de la desviación de la regresión, la utilidad del método de análisis de regresión para la interpretación de los datos resulta, en este caso, poco útil.

Como se ha dicho antes, se puede extraer aún otro componente con un grado de libertad para probar la tendencia de las líneas de regresión a converger en un punto. Este componente ("Conv." en Tabla 3), calculado como se indica en [9], es significativo al 5%, denotando que gran parte de la

variación en la heterogeneidad de las regresiones se debe a una **convergencia** de las líneas de regresión en un punto común. La convergencia indica una correlación significativa entre las medias de los genotipos y los coeficientes de regresión, que en nuestro caso resulta ser $r = 0.67^{**}$ (este coeficiente de correlación lo obtenemos en el listado de salida de la corriente de control realizada para el SAS, para éste método). Así, para el ejemplo que estamos analizando, un genotipo especialmente adaptado a ambientes de bajo rendimiento caracterizado por un rendimiento medio alto y con bajo coeficiente de regresión, sería raro encontrarlo en el material de trabajo.

Así pues, el modelo de regresión no sería el adecuado para describir la estabilidad de nuestros genotipos. Esta relativa falta de linealidad para estos datos confirman las observaciones de muchos autores en los que la interacción **GxE** no resultó ser función lineal del efecto ambiental (e_j), principio fundamental en el que está basado el método de la regresión como se manifestó en [5]. En estos casos, para la selección de genotipos estables puede utilizarse el parámetro S^2_{di} (el cuadrado medio residual) dado en [8], el cual describe una medida de irregularidades impredecibles en las respuestas a los ambientes, es decir, suministra la verdadera medida de estabilidad (Langer y col., 1979). Los resultados obtenidos para nuestro ejemplo son los que se muestran en la Tabla 4. Dado que, en nuestro caso, el método de Regresión no ha resultado muy fiable, la selección la efectuaríamos por medio del parámetro S^2_{di} . Según éste parámetro los genotipos seleccionados serían aquellos que tuvieran S^2_{di} no significativamente distinta de cero, es decir, los genotipos 1, 6, 7, 9, 10 y 11.



Tabla 4. Rendimiento Medio, Coeficiente de Regresión (b_i) y S^2_{di} , para cada genotipo.

Genotipo	Rend. Medio	Coef. Regres.	S^2_{di}
1	3699.37	.9637	49843.79
2	3825.07	1.4456 #	162724.80 **
3	2877.50	.3858 #	464869.80 ***
4	2920.35	.8743	123878.00 *
5	3896.24	1.1374	127512.10 *
6	3907.15	1.2733 #	78695.36
7	4241.11	1.1412	87471.60
8	3193.40	1.2132	122333.20 *
9	3320.56	.7652 #	32766.83
10	3517.98	.8777 #	28145.89
11	3617.71	.9222	73808.71

: Coeficientes de Regresión significativamente distintos de la unidad.

$\sigma^2_o = 57,791$ con 220 g.l.

* $P < 0.05$ ** $P < 0.01$ *** $P < 0.001$

REFERENCIAS CITADAS

• LANGER, I., K.J. FREY and T. BAILEY. 1979: Associations among productivity production response and stability indexes in oat varieties. Euphytica 28, 17-24.

• STEEL, R.G.D. y J.H. TORRIE. 1985: Bioestadística: Principios y procedimientos. Ed. Mc Graw-Hill.

OTRAS REFERENCIAS ACONSEJADAS

• EAGLES, H.A., P.N. HINZ, and K.J. FREY. 1977: Selection of superior cultivars of oats by using regression coefficients. Crop Sci. 17, 101-105.

• KROONENBERG, P.M. and K.E. BASFORD. 1989: An investigation of multi-attribute genotype response across environments using three-mode principal component analysis. Euphytica 44, 109-123.

• EBERHART, S.A. and W.A. RUSSELL. 1966: Stability Parameters for Comparing Varieties. Crop Sci. 6, 36-40.

• MANDEL, J. 1961: Non-additivity in two-way analysis of variance. J. Amer. Statist. Assoc. 56, 878-888.

• FINLAY, K.W. and G.N. WILKINSON, 1963: The analysis of adaptation in a plant breeding programme. Aust. J. Agric. Res. 14, 742-754.

• McINTOSH, M.S. 1983: Analysis of combined experiments. Agron. Jour. 75, 153-155.

• FREEMAN, G.H. and J.M. PERKINS. 1971: Environmental and genotype-environmental components of variability. VIII. Heredity 27, 15-23.

• MUNGOMERY, V.E. 1978: Genetic analyses of environmental interactions and effects of competition in soybeans. Unpublisher Ph. D. Thesis, Department of Agriculture, University of Queensland.

• FREEMAN, G.H. 1973: Statistical methods for the analysis of genotype-environment interactions. Heredity 31 (3), 339-354.

• PERKINS, J.M. and J.L. JINKS. 1968b: Environmental and genotype-environmental components of variability. IV. Heredity 23 (4), 525-535.

• HINZ, P.N. and H.A. EAGLES. 1976: Estimation of a transformation for the analysis of some agronomic and genetic experiments. Crop Sci. 16, 280-283.

• POONI, H.S. and J.L. JINKS. 1980: Non-linear genotype- environment interactions. II. Statistical models and genetical control. Heredity 45, 389-400.

- SHORTER, R. 1972: Influence of genotype and environment on chemical composition of soy-bean seeds (*Glycine max* L. Merrill) Unpublished M. Agr. Sc. Thesis, University of Queensland.
- SOKAL, R.R. and F.J. ROHLF. 1979: Biometría. Principios y métodos estadísticos en la investigación biológica. H. Blume Ediciones.
- TAI, G.C.C., 1971: Genotypic Stability Analysis and Application to Potato Regional Trials. Crop Sci. 11, 184-190.
- TUKEY, J.W. 1949: One degree of freedom for non-additivity. Biometrics 5, 232-242.
- VERMA M.M., G.S.CHAHAL and B.R.MURTY. 1978: Limitations of Conventional Regression Analysis A Proposed Modification. Theor. Appl. Genet. 53, 89-91.
- VIRK, D.S., N.B. SINGH, M. SRIVASTAVA and G. HARINARAYANA. 1984. Regression analysis for general adaptation in pearl millet using different environmental indices. Theor. Appl. Genet. 68, 509-513.
- VIRK, D.S., P.S. VIRK, B.K. MANGAT and G. HARINARAYANA. 1988: Linear and non-linear regression analysis of genotype x environment interactions in pearl millet. Theor Appl. Genet. 75, 736-740.
- YATES, F. and W.G. COCHRAN. 1938: The analysis of groups of experiments. J. Agr. Sci., 28, 556-580.

 Σ α Π Ω \times \pm ϑ



3

**PARÁMETROS
UNIVARIANTES DE
ESTABILIDAD MÁS UTILIZADOS**

Σ α Π Ω \times \pm ϑ



3. PARÁMETROS UNIVARIANTES DE ESTABILIDAD MÁS UTILIZADOS

Como ya se indicó en la Introducción los parámetros univariantes de estabilidad más utilizados en el estudio de la interacción GxE son los siguientes:

1) Varianza de estabilidad (σ^2_i) y (s^2_i) de Shukla.

2) Coeficiente de Variación de Francis y Kannembert (CV_i).

3) Ecovalencia de Wricke (W_i) y **Cuadros Medios de la ecovalencia**.

4) Varianza ambiental de genotipos (S^2_{xi}).

5) Medida única de Lin y Binns.

3.1. Varianza de estabilidad

Shukla (1972) definió dos varianzas (parámetros) de estabilidad σ^2_i y s^2_i . La primera se calcula por medio de la fórmula:

$$\sigma^2_i = \frac{(g(g-1))\sum_j(Y_{ij}-\bar{Y}_i-\bar{Y}_j+\bar{Y}_{..})^2 - \sum_i \sum_j (Y_{ij}-\bar{Y}_i-\bar{Y}_j+\bar{Y}_{..})^2}{(g-1)(g-2)(e-1)} \quad (10)$$

donde:

Y_{ij} = valor medio del carácter medido (por ejemplo, rendimiento) para el genotipo i en el ambiente j .

- g = número de genotipos,
- e = número de ambientes.

Cuando el número de genotipos g es grande, para probar si $\sigma^2_i=0$, el estadístico calculado de la forma $F^* = \sigma^2_i / \sigma^2_o$, donde σ^2_o es la varianza del error dentro de ambientes, se aproxima a un distribución F de Snedecor con $(e-1)$ y $e^*g^*(r-1)$ grados de libertad para cada σ^2_i . De esta forma, para saber si un determinado valor de σ^2_i es estadísticamente no distinto de cero, efectuamos en primer lugar el cociente σ^2_i / σ^2_o , tal y cómo se ha comentado en el capítulo anterior. En segundo lugar, el valor de F^* así calculado lo contrastamos con el de las tablas de la F de Snedecor con $(e-1)$ grados de libertad para el numerador y $eg(r-1)$ grados de libertad para el denominador. Si el valor de F^* contrastado resulta no

significativo, entonces la σ^2_i no es, estadísticamente hablando, distinta de cero.

Información adicional sobre la estabilidad de cada genotipo se puede obtener por medio de la introducción de un covariante que puede representar fertilizaciones y/o prácticas de campo distintas. El covariante más utilizado es el denominado por Shukla como Z_j , que para el ambiente j -ésimo se define como la diferencia entre el efecto aditivo del ambiente j (esto es, $Y_{.j}$) y la media total del ensayo ($Y_{..}$), esto es, $Z_j=Y_{.j}-Y_{..}$. Por tanto, este covariante es en realidad un índice ambiental. Las ecuaciones 15 y 16 de Shukla (1972) son las que se utilizan para extraer el efecto del covariante en componentes asignables a cada genotipo (s^2_i).

La ecuación 15 de Shukla es la del coeficiente de regresión lineal para el genotipo i -ésimo:

$$b_i = \frac{\sum_j (Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_{..}) * Z_j}{\sum_j Z_j^2} \quad (11)$$

La ecuación 16 de Shukla nos da su parámetro de estabilidad s^2_i :

$$s^2_i = \frac{g}{(g-2)(e-2)} [S_i - \frac{\sum_j S_j}{g(g-1)}] \quad (12)$$

donde:

$$S_i = \sum_j (Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_{..} - b_i * Z_j)^2 \quad (13)$$

Cuando g es grande, se procede con s^2_i de igual forma que para σ^2_i . Así, para probar si $s^2_i=0$, $F^* = s^2_i / \sigma^2_o$ se aproxima a un distribución F de Snedecor con $(e-2)$ y $e^*g^*(r-1)$ grados de libertad para cada s^2_i , procediéndose como antes.

3.2. Coeficiente de Variación

Dado que la respuesta genotípica puede tener una gran varianza, Francis y Kannenberg (1978) sugirieron una medida de mayor consistencia para rendimiento. Esta medida es el **coeficiente de variación** (CV_i), que es el cociente, expresado generalmente bajo la forma de porcentaje, entre la desviación estándar (σ) y la media. La Figura 4 representa la gráfica para el rendimiento medio en función del CV_i .

El rendimiento total del ensayo y el CV medio dividen la gráfica en 4 grupos:



GRUPO I: Alto rendimiento, pequeña variación.
 GRUPO II: Alto rendimiento, gran variación.
 GRUPO III: Bajo rendimiento, pequeña variación.
 GRUPO IV: Bajo rendimiento, gran variación.

Considerando como genotipo estable aquel que posee un rendimiento alto y consistente, sólo los genotipos del Grupo I pueden considerarse como estables.



El Grupo III también es consistente, pero pueden considerarse a los genotipos incluidos en el como inestables debido a que los rendimientos están por debajo del rendimiento medio del ensayo para la mayoría del intervalo ambiental.

Se puede realizar algunas comparaciones utilizando los parámetros ya comentados. Así, obteniendo los coeficientes de regresión por grupos, los grupos III y IV (en los que $b_i \approx 0$; ver Tabla 7) podrían ser considerados por encima de los de media estabilidad ($b_i \approx 1$; método de Finlay y Wilkinson, 1963). Sin embargo, cuando consideramos las S^2_{di} de Eberhart y Russell (1966) y la σ^2_i de Shukla (1972), son los grupos III y I los que hay que considerar como estables.



Veamos las diferencias concretas entre éstos métodos: el genotipo 5, en el Grupo I, es considerado como estable según el método del coeficiente de variación, pero viendo la alta significación para los parámetros σ^2_i y S^2_{di} (ver las Tablas 5 y 4, respectivamente) debería considerarse como inestable. Por el contrario, los genotipos 9 y 10, que no aparecen en el Grupo I, y por tanto, según el método del coeficiente de variación se considerarían como inestables, según los valores de las significaciones para σ^2_i y S^2_{di} deben considerarse como estables.



Siempre nos encontraremos con resultados como los comentados en el párrafo anterior. Es decir, tendremos genotipos que de acuerdo con la mayoría de los métodos utilizados se deben considerar como estables, y que para otros sin embargo, estarán caracterizados por ser los mas inestables. Por último tendremos aquellos genotipos que tras el estudio de la interacción GxE con varios métodos no quedan definidos claramente. Así pues, además de los cálculos estadísticos se precisa un buen conocimiento de nuestro material vegetal para discutir adecuadamente los resultados.

El método del CV-medio se diseñó en principio como ayuda en estudios sobre la base fisiológica de la estabilidad del rendimiento (Francis, 1977). Se trata de un método que agrupa más que caracteriza genotipos individualmente.

3.3. Ecovalencia y Cuadrados Medios de la Ecovalencia

- La ecovalencia de Wricke (1962) viene expresada por:

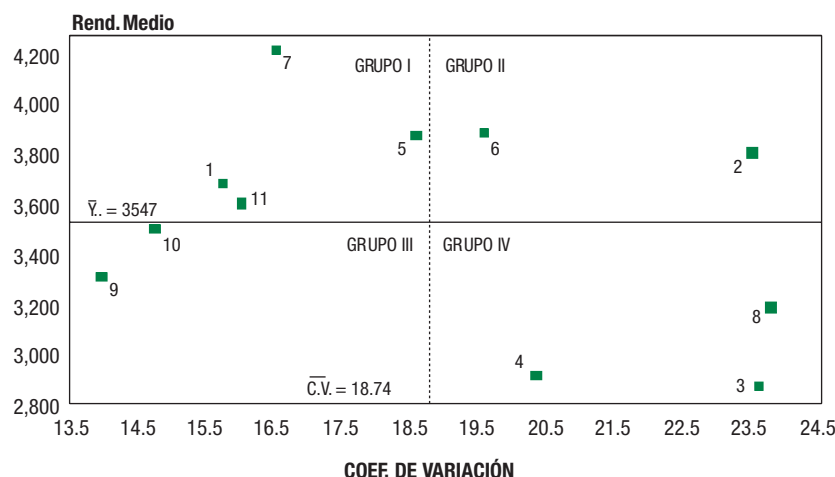
$$W_i = \sum_j (Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_{..})^2 \quad (14)$$

Puesto que la ecovalencia mide la contribución de un genotipo i a la interacción **GxE**, el genotipo con $W_i=0$ será el más estable.

Como puede deducirse de la expresión [10] la varianza de estabilidad es una función lineal de la ecovalencia:

$$\sigma^2 = \frac{G \cdot W_i}{(g-2) \cdot (e-1)} - \frac{C.M. \text{ interacción GxE}}{(g-2)} \quad (15)$$

Figura 4. Coeficiente de Variación





Muchos autores han utilizado la suma de cuadrados W_i (ecovalencia) para una comparación relativa entre cultivares, pero pocos trabajos utilizan los cuadrados medios de la ecovalencia (W_i -CM) propuesto por Kang y Miller (1984). El cuadrado medio de la ecovalencia se compararía de la misma forma que la σ^2_i de Shukla (véase el apartado 3.1). Esto supondría dividir la suma de cuadrados de la interacción **GxE** de cada cultivar por $(e-1)(g-1)/g$ y comparar los cuadrados medios con σ^2_o , suponiendo que W_i -CM se distribuye como una F de Snedecor (hipótesis que necesita ser verificada). Así, para el genotipo 1, de la Tabla 5 vemos que $W_1 = 402491.8$, y el valor de $(e-1)(g-1)/g = 8.1818$. Con estos valores, el cálculo de los cuadrados medios de la ecovalencia sería tal como sigue: $W_1\text{-CM} = 402491.8/8.1818 = 49193.4$; este valor dividido por la $\sigma^2_o = 45675$ nos da el estadístico que debe ser contrastado según una F de Snedecor con 9 ($=e-1$) g.l. en el numerador y 220 g.l. en el denominador. Pero Kang y Miller (1984) sugieren que ésta sería una hipótesis válida si σ^2_i y W_i -CM estuvieran fuertemente correlacionadas y si σ^2_i / σ^2_o se distribuyera aproximadamente como una F de Snedecor. La posibilidad de probar la significación de W_i -CM podría incrementar la utilidad y efectividad de este estadístico como índice de estabilidad.

Kang (1985) y Kang y col., (1987) realizaron un estudio exhaustivo de estos dos métodos que reveló que la ecovalencia (W_i) era un paso intermedio para el cálculo de σ^2_i y que el uso de W_i como una medida separada de estabilidad no es necesario dado que σ^2_i es un parámetro insesgado (es decir, depende para su cálculo de menos parámetros que la ecovalencia) que puede ser probado por significación estadística.

Es significativo que tanto W_i -CM como σ^2_i muestren prácticamente los mismos genotipos como estables (ver Tabla 5). Los valores de σ^2_i y W_i -CM para cada genotipo son muy similares; el orden relativo de los genotipos para ambos parámetros, sería exactamente el mismo, pero los cálculos para la obtención de W_i -CM son menores y mas simples que los utilizados para calcular σ^2_i . El método de Shukla sería el preferido, sin embargo, si se utilizase un covariante y, por tanto, escogiéramos el parámetro (s^2_i).

3.4. Varianza ambiental

La estabilidad fenotípica, según el **concepto estático**, fue medida por Roemer (1917) y

más tarde por numerosos autores usando la varianza obtenida de los resultados de un genotipo en todos los ambientes del ensayo. Se denominó **varianza ambiental**:

$$S^2_{xi} = \frac{\sum_j (Y_{ij} - \bar{Y}_i)^2}{(e-1)} \quad (16)$$

Desde el punto de vista del **concepto dinámico**, los estadísticos más utilizados en la partición de la interacción **GxE** son la **ecovalencia** y la **varianza de estabilidad**. La **varianza ambiental**, como se ha comentado en la Introducción, es útil para caracteres de calidad, para resistencia a enfermedades o caracteres estresantes, pero para rendimiento, un mejorador intenta encontrar simultáneamente genotipos estables y de alto rendimiento. Por contra, este parámetro nos evalúa la estabilidad de los genotipos desde el punto de vista del concepto estático, asociado generalmente con rendimientos bajos.

3.5. Medida única de Lin y Binns

Lin y Binns (1988) propusieron una medida única de la superioridad del comportamiento de un genotipo i (P_i). P_i se define como el cuadrado medio de la distancia entre la respuesta de un genotipo y el genotipo de máxima respuesta, en un ambiente dado, expresándose de la forma:

$$P_i = \frac{\sum_j (Y_{ij} - M_j)^2}{2n} \quad (17)$$

donde Y_{ij} es el rendimiento del genotipo i -ésimo en el ambiente j -ésimo, M_j es la máxima respuesta de entre todos los genotipos en el ambiente j -ésimo, y n es el número de ambientes. Un valor pequeño de éste índice de superioridad implica adaptación general de un genotipo. Este índice integra rendimiento medio y estabilidad relativa en un sólo parámetro.

Como la distribución de éste parámetro no es conocida, Lin y Binns utilizaron como prueba de comparación el cuadrado medio del error (CME) de la ANOVA por el estadístico F de Snedecor para una determinada probabilidad. Así, $CME = 137,025$ con 200 g.l., y el valor de F (Snedecor) con un nivel del 5% y 10 g.l. (número de ambientes) es de 1.83; por tanto, el punto de corte para P_i sería $137,025 \times 1.83 = 250,755.75$.



3.6. Resultados con los datos experimentales

Los resultados de la aplicación de éstos estadísticos a los datos ejemplo se muestran en las Tablas 5, 6 y 7. Su cálculo puede hacerse partiendo de la matriz de datos del Apéndice y siguiendo un programa en lenguaje BASIC que nos calcula todos los parámetros univariantes comentados, excepto el P_i , para el cual existe una corriente de control realizada para el paquete estadístico SAS. Comentaremos brevemente los resultados para cada estadístico.



3.6.1. Respecto de las Varianzas de Shukla σ^2_i [10] y s^2_i [12], así como de los estadísticos relacionados con la ecovalencia W_i y W_i -CM (véase 3.3), se puede decir que los resultados son muy similares y prácticamente los genotipos



considerados como estables son los mismos (Tabla 5). Estos genotipos son: 1, 6, 7, 9, 10 y 11.

3.6.2. Según el método del Coeficiente de Variación los genotipos mas estables serían aquellos incluidos en el Grupo I de la Figura 4 (Tablas 5 y 6). Estos son los genotipos **1, 11, 5 y 7**.

3.6.3. En cuanto a la Varianza Ambiental los genotipos más estables serían el **9, 10, 11, 1, 4 y 3**, de los cuales tan sólo el 1 y el 11 superan ligeramente el rendimiento medio total del ensayo (Tabla 5). Se comprueba así, como hemos comentado anteriormente, que este parámetro nos evalúa la estabilidad de los genotipos desde el punto de vista del concepto estático, asociado generalmente con rendimientos bajos.



Tabla 5. Parámetros de estabilidad σ^2_i , s^2_i , CV_i , W_i , W_i -CM y S^2_{xi} .

GEN.	σ^2_i	s^2_i	CV_i	W_i	W_i -CM	S^2_{xi}
1	38389.8	45897.4	15.73	402491.8	49193.4	338689.7
2	237453.2***	183863.1**	23.48	1868321.0	228350.4***	807001.8
3	634877.2***	553151.4***	23.58	4794808.0	586032.1***	460401.0
4	124434.3*	136383.6*	20.33	1036091.0	126633.4*	352371.9
5	129582.4*	140825.3*	18.57	1074001.0	131266.8*	523394.7
6	98155.3	81160.4	19.55	842582.7	102982.3	583765.1
7	86491.0	91886.9	16.51	756691.3	92484.5	490541.2
8	134251.7*	134495.6*	23.75	1108384.0	135469.1*	575251.3
9	40682.2	25025.6	13.95	419372.0	51256.6	214700.8
10	20100.2	19377.8	14.74	267812.9	32732.7	269178.1
11	66262.7	75187.9	16.00	607737.6	74279.1	335133.4

$\sigma^2_o = 57,791$ con 220 g.l.
 * $P < 0.05$ ** $P < 0.01$ *** $P < 0.001$



Tabla 6. Parámetros de estabilidad para los cuatro grupos definidos en la Fig. 4.

GRUPO	REND. MEDIO	b_i MEDIA	S^2_{di} MEDIA	σ^2_i MEDIA
I	3863.60	1.0411	84658.	80181.
II	3866.11	1.3594	120710.*	167804.**
III	3419.27	0.8214	30455.	30391.
IV	2997.08	0.8244	237026.***	297854.***



Tabla 7. Índice de Lin y Binns: P_i .

GENOTIPO	RENDIMIENTO MEDIO	P_i
7	4241.11	18079.87
6	3907.16	240540.41
5	3896.25	277814.63
2	3825.07	357083.54
1	3699.38	510353.34
11	3617.71	672759.88
10	3517.99	799177.34
9	3320.56	1242578.43
8	3193.40	1512318.61
4	2920.35	2413706.44
3	2877.50	3109467.61

3.6.4. Si $1.83 \times 137,025 = 250,755.75$ se utiliza como punto de corte para P_i (véase 3.5) sólo los genotipos **7 y 6** se seleccionarían como no diferentes de la respuesta máxima.

Como podemos ver en la Tabla 7 el estadístico P_i , para nuestros datos, ordena los genotipos según sus rendimientos medios y no por su estabilidad. Por tanto, seleccionar por este estadístico resultaría en seleccionar por rendimiento, como ya se comentó en la Introducción.

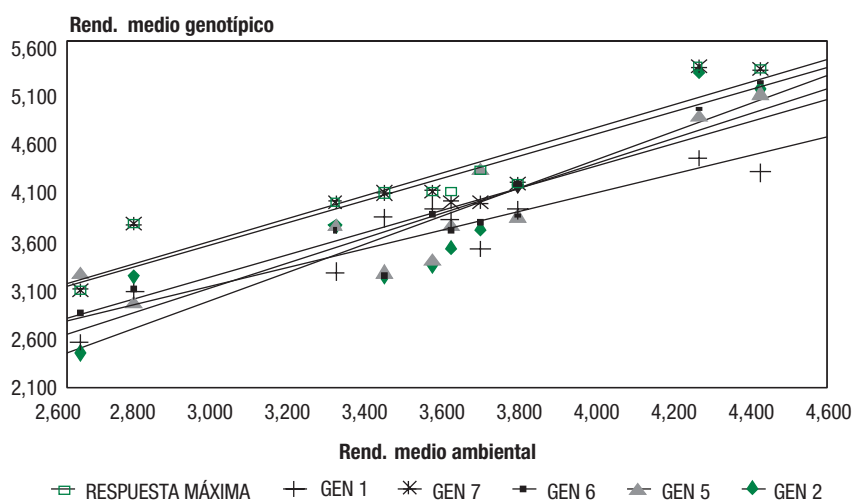
Puede complementarse la selección de los genotipos más estables por medio de la construcción de la gráfica de la Figura 5, donde se representan los rendimientos medios de cada genotipo en cada ambiente en función del rendimiento ambiental para ser comparados con la respuesta máxima.

En la Figura 5 vemos como el genotipo 7 es el de mayor rendimiento medio en la mayoría de los ambientes del ensayo, sobresaliendo en la gráfica del resto de genotipos, siendo casi imposible distinguir un segundo genotipo que destaque de entre los restantes.

Como comentario final cabe repetir aquí lo dicho en la Introducción: distintos métodos dan con frecuencia distintos resultados. Esto no es una paradoja: es consecuencia de la enorme complejidad del material biológico en su interacción con el ambiente.

Por ello, conviene siempre aplicar **más de un método**, y decidir en base a los resultados obtenidos y al conocimiento del material biológico.

Figura 5. Comparación con respuesta máxima



REFERENCIAS CITADAS

• EBERHART, S.A. and W.A. RUSSELL. 1966: Stability Parameters for Comparing Varieties. *Crop Sci.* 6, 36-40.

• FINLAY, K.W. and G.N. WILKINSON, 1963: The analysis of adaptation in a plant breeding programme. *Aust. J. Agric. Res.* 14, 742-754.

• FRANCIS, T.R. 1977: Yield stability studies in short-season maize (*Zea mays* L.) Ph. D. Thesis University of Guelph, Guelph, Ont.

• FRANCIS, T.R. and L.W. KANNENBERG. 1978: Yield stability studies in short-season maize. I. A descriptive method for grouping genotypes. *Can. J. Plant Sci.* 58, 1029-1034.

• KANG, M.S. and J.D. MILLER. 1984: Genotype x Environment interactions for cane

and sugar yield and their implications in sugar-cane breeding. *Crop Sci.* 24, 435-440.

• KANG, M.S., J.D. MILLER AND L. L. DARRAH. 1987: A note on relationship between stability variance and ecovalence. *J. Hered.* 78, 107-107.

• KANG, M.S. 1985: Sas program for calculating stability-variance parameters. *J. Hered.* 76, 142-143.

• LIN, C.S. and M.R. BINNS. 1988: A superiority measure of cultivar performance for cultivar x location data. *Can. J. Plant Sci.* 68, 193-198.

• SHUKLA, G.K., 1972: Some Statistical aspects of partitioning genotype-environmental components of variability. *Heredity* 29, 237-245.

• WRICKE, G. 1962: Über eine methode zur erfassung der ökologischen streubreite in feldversuchen. *Z. Pflanzzüchtg* 47, 92-96.

OTRAS REFERENCIAS ACONSEJADAS

• BECKER, H.C. and J. LEON, 1988: Stability Analisis in Plant Breeding. *Plant Breeding* 101, 1-23.

• GRAY, E. 1982: Genotype x environment interactions and stability analysis for forage yield of orchardgrass clones. *Crop Sci.* 22, 19-23.

• JOWETT, D. 1972: Yield stability parameters for sorghum in East Africa. *Crop Sci.* 12, 314-317.

• NGUYEN, H.T., D.A. SLEPER, and K.L. HUNT. 1980: Genotype x environment interactions and stability analysis for herbage yield. *Crop Sci.* 20, 221-224.

• ROEMER, T. 1917: Sind die ertragsreichen sorten ertragssicherer? *Mitt. DLG* 32, 87-89.

• WRICKE, G. and W.E. WEBER. 1986: Quantitative Genetics and Selection in Plant Breeding. De Gruiter.

4

MÉTODOS DE TAI (1971)

4. MÉTODO DE TAI (1971)

Tai (1971) describe otro método para estimar parámetros de estabilidad. Es similar al método de Eberhart y Russell (ver Capítulo 2) en el sentido de que ambos métodos estiman la respuesta lineal de una variedad respecto de los efectos ambientales. Sin embargo, el método descrito por Tai difiere del propuesto por Eberhart y Russell en la estimación de los parámetros que determinan estabilidad: Tai relaciona el modelo matemático para el análisis de la varianza con la estimación del potencial genotípico de una variedad para estabilizar su comportamiento frente a las variaciones ambientales.

El modelo matemático en el cual se fundamenta el método de Tai está basado en el método de **análisis de relación estructural**. Dado que la demostración del modelo matemático del método de Tai queda fuera de los objetivos de este manual, para más detalle se recomienda examinar el libro de Kendall y Stuart (1961) en lo concerniente al análisis de relación estructural. Aún así, a continuación se describirá brevemente aquellos aspectos más importantes para la comprensión de este método.

A partir del método de análisis de relación estructural, se pueden derivar los siguientes estimadores de máxima verosimilitud bajo el supuesto de que las variables e_j (efecto de j -ésimo ambiente) y $(ge)_{ij}$ (efecto de la interacción **GxE**) están distribuidas normalmente:

$$\sigma_e^2 + \sigma_\lambda^2 = s_e^2 = \sum e_j^2 / (e-1) \quad (18)$$

$$\alpha_i^2 \sigma_e^2 + \sigma_{\delta_i}^2 + \sigma_\epsilon^2 = s^2(ge)_i = \sum (ge)_{ij}^2 / (e-1) \quad (19)$$

$$\alpha_i^2 \sigma_e^2 = s_e(ge)_i = \sum e_j(ge)_{ij} / (e-1) \quad (20)$$

donde σ_e^2 es la varianza de los efectos ambientales, $\sigma_{\delta_i}^2$ es la varianza de las desviaciones de la respuesta lineal, σ_λ^2 es la varianza del error asociado con los efectos ambientales, σ_e^2 es la varianza del error asociado con los efectos de la interacción, s_e^2 es la varianza muestral de los efectos ambientales, $s^2(ge)_i$ es la varianza muestral de los efectos de la interacción del genotipo i -ésimo en los e ambientes y $s_e(ge)_i$ es la covarianza muestral entre los efectos ambientales y la interacción.

Siendo MSL, MSB y MSE los cuadrados medios de los efectos ambientales, de las repeticiones dentro de ambientes y del error, respec-

tivamente, obtenidos en el análisis de la varianza (ver Capítulo 2), se demuestra que:

$$s_e^2 = \text{MSL}/gr, \sigma_\lambda^2 = \text{MSB}/gr \text{ y } \sigma_\epsilon^2 = [(g-1)(\text{MSE}/r)] \quad (21)$$

donde g es el número de genotipos y r el número de repeticiones del ensayo. Haciendo $\lambda_i = (\sigma_{\delta_i}^2 + \sigma_\epsilon^2)/\sigma_e^2$ e introduciendo las relaciones anteriores en [19], resulta:

$$\sigma_e^2 + \sigma_\lambda^2 = \text{MSL}/gr \quad (22)$$

$$\alpha_i^2 \sigma_e^2 + \lambda_i \sigma_e^2 = s^2(ge)_i \quad (23)$$

$$\alpha_i^2 \sigma_e^2 = s_e(ge)_i \quad (24)$$

$$\sigma_\lambda^2 = \text{MSB}/gr \quad (25)$$

$$\sigma_\epsilon^2 = (g-1)\text{MSE}/gr \quad (26)$$

Resolviendo este sistema de ecuaciones obtenemos las estimaciones de los parámetros α_i y λ_i para el genotipo i -ésimo:

$$\alpha_i = \frac{s_e(ge)_i}{(\text{MSL}-\text{MSB})/gr} \quad \text{y} \quad (27)$$

$$\lambda_i = \frac{s^2(ge)_i - \alpha_i s_e(ge)_i}{(g-1)\text{MSE}/gr}$$

α_i mide la respuesta lineal del genotipo i a los efectos ambientales, y λ_i la desviación de la respuesta lineal en unidades relativas de la varianza del error (MSE).

Se puede obtener un intervalo de predicción para $\alpha_i=0$ con un nivel de probabilidad P considerando la correlación entre los pares de efectos e_j y $(ge)_{ij}$ para el genotipo i .

Denominando t_a al estadístico t con un nivel de probabilidad de $a=1-P$ con $(e-2)$ grados de libertad, los dos límites de predicción para α_i correspondientes a $\alpha_i=0$ se demuestra que son:

$$\pm t_a \left[\frac{\lambda(g-1)\text{MSE}\text{MSL}}{(\text{MSL}-\text{MSB})[(e-w)\text{MSL}-(t_a^2+n-2)\text{MSB}]} \right]^{1/2} \quad (28)$$

expresión que es función de λ .

De esta forma, a partir de [28] podemos construir unas regiones de confianza (parábolas que pasan por el origen) a un determinado nivel de probabilidad, dentro de las cuales, el valor de α_i no es significativamente distinto de cero (Fig. 6).





Se puede construir también un intervalo de confianza teórico para un hipotético valor de λ_0 por medio de una distribución F con $n_1=e-2$ y $n_2=e(g-1)(r-1)$ grados de libertad, de esta manera:

Para $\lambda_0 = 1$ el intervalo a un nivel de probabilidad P es

$$F_a(n_2,n_1) \leq \lambda_0 \leq F_a(n_1,n_2)$$

donde $F_a(n_2,n_1)=1/F_a(n_1,n_2)$ y $2a=1-P$.



El intervalo para $\lambda_0 > 1$ se puede construir por medio de una distribución F no centrada (ver Kendall y Stuart, 1961; o Scheffe, 1959). Siendo $F'_a(n_1,n_2)$ el límite superior para un hipotético valor de λ_0 mayor que la unidad, tenemos que, aproximadamente:

$$F'_a(n_1,n_2) = \lambda_0 F_a [n_1 \lambda_0^2 / (2\lambda_0 - 1), n_2] \quad (29)$$

Dado que un genotipo perfectamente estable no cambiaría su comportamiento de un ambiente a otro, esto es equivalente a decir, para el método de Tai, que $\alpha = -1$ y $\lambda = 1$. Eberhart y Russell definieron como genotipos estables los que tenían $b = 1$ y $s^2_{di} = 0$, lo que es equivalente a tener los valores $\alpha = 0$ y $\lambda = 1$ en el modelo de Tai. El valor de $\alpha = 0$ y $\lambda = 1$ lo define Tai como genotipos de mediana estabilidad y los valores de $\alpha = -1$ y $\lambda = 1$ como genotipos de perfecta estabilidad.

4.1. Caso práctico

Pasemos ahora al cálculo y discusión de los parámetros de estabilidad según el método diseñado por Tai:

La distribución de los valores de α_i y λ_i de los 11 genotipos se muestran en la Fig. 6 y sus valores en la Tabla 8. Para ello se han utilizado las fórmulas del apartado [27] y para las curvas de confianza la fórmula [28]; se han representado las curvas correspondientes a los niveles del 20, 10 y 5%.

Tan sólo dos genotipos, el 2 y el 9, han mostrado un valor significativamente distinto de $\alpha = 0$, al nivel del 20% (Tabla 8).

El genotipo 3 ha mostrado un valor significativo al 5% de $\lambda > 1$; por tanto, se sitúa fuera de la zona de comportamiento genotípico estable. Por otro lado los genotipos 1, 9 y 10 han mostrado un valor de λ significativamente menor de uno, de ahí que también se consideren como genotipos inestables.

Los genotipos situados en la región de estabilidad media y, por tanto también considerados como deseables son: **2, 4, 5, 6, 7, 8, y 11**. Según este método entre los genotipos situados en la zona de estabilidad media seleccionaríamos los de mayor rendimiento; éstos serían el **7** y el **6** que no difieren estadísticamente en sus rendimientos medios a un nivel del 5%.

Como podemos ver en la Figura 6, existe más variación en el estadístico λ sugiriendo que el componente relativamente impredecible (así λ en el modelo de Tai y S^2_{di} en el modelo de la regresión, indicándonos ambos la desviación de la respuesta lineal) de la interacción **GxE** puede ser mucho más importante que el componente predecible (α y b_i , coeficientes de la respuesta lineal). Estos resultados corroboran los resultados del Capítulo 2.

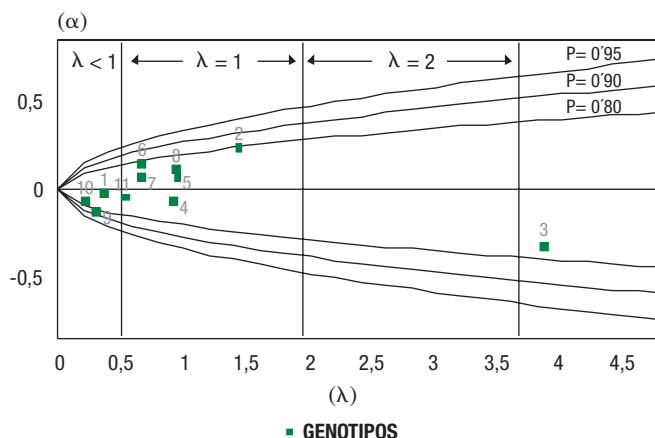
Tabla 8. Resultados del método de Tai.⁽¹⁾

GENOTIPO	α	λ
1	-0.018084	0.362717#
2	0.222528+	1.435433
3	-0.306657	3.854765*
4	-0.062764	0.917683
5	0.068642	0.948011
6	0.136454	0.666558
7	0.070535	0.659362
8	0.106482	0.944867
9	-0.117234+	0.308609#
10	-0.061054	0.223194#
11	-0.038851	0.542428

(1) * = Valor λ mayor que Fa derivado de F Snedecor con: $n_1 = g-2$; $n_2 = g(c-1)(r-1)$; $a = 0.05$
 # = Valor λ menor que Fa derivado de F Snedecor con: $n_1 = g-2$; $n_2 = g(c-1)(r-1)$; $a = 0.05$
 + = Valor α distinto de 0 para el 20%.



Gráfico 6. Gráfica de Tai



Σ

α

Π

Ω

\times

\pm

ϑ

REFERENCIAS CITADAS

• EBERHART, S.A. and W.A. RUSSELL. 1966: Stability Parameters for Comparing Varieties. *Crop Sci.* 6, 36-40.

• KENDALL, M.G. and A. STUART. 1961: The advanced of statistics. Vol. 2, 3 rd ed. Griffin and Co. Ltd. London.

• SCHEFFE, H. 1959: The analysis of variance. Jhon Wiley Sons, Inc., New York.

• TAI, G.C.C., 1971: Genotypic Stability Analysis and Application to Potato Regional Trials. *Crop Sci.* 11, 184-190.

OTRAS REFERENCIAS ACONSEJADAS

• KANG, M.S. and J.D. MILLER. 1984: Genotype x Environment interactions for cane and sugar yield and their implications in sugar-cane breeding. *Crop Sci.* 24, 435-440.

• LIN, C.S., M.R. BINNS, and L.P. LEFKOVITCH, 1986: Stability analysis: Where do we stand?. *Crop Sci.* 26, 894-900.

• TAI, G.C.C. 1975: Analysis of genotype-environment interactions based on the method of

path coefficient analysis. *Can. J. Genet. Cytol.* 17, 141-149.

• TAI, G.C.C. 1979: Analysis of genotype-environment interactions of potato yield. *Crop Sci.* 19, 434-438.

• YUE, G, S. K. PERNG, T. L. WALTER, C. E. WASSOM and G. H. LIANG. 1990: Stability Analysis of Yield in Maize, Wheat and sorghum and its Implications in Breeding Programs. *Plant Breeding.* 104, 72-80.



5

MÉTODOS NO PARAMÉTRICOS

5. MÉTODOS NO-PARAMÉTRICOS

Los métodos no-paramétricos más utilizados en esta última década son los siguientes:

- 1) **Estadísticos no paramétricos de Hühn.**
- 2) **Método de “consistencia de comportamiento”.**
- 3) **Método de Kang.**
- 4) **Método de “ordenamientos estratificados”.**

5.1. Estadísticos no paramétricos de Hühn

Los estadísticos no-paramétricos se basan en las ordenaciones (r_{ij}) de los valores Y_{ij} (valores medios genotípicos en cada ambiente). Siguiendo la misma nomenclatura que propuso Hühn (1979), los estadísticos propuestos por él son los siguientes:

$$s_i^{(1)} = \frac{2\sum_{j < j'} |r_{ij} - r'_{ij}|}{e(e-1)} \quad (30)$$

donde r_{ij} es el orden del genotipo i en el ambiente j , y

e es el número de ambientes del ensayo.

Este primer estadístico nos calcula la media de las diferencias absolutas de los órdenes del genotipo i en todos los ambientes. La varianza de los órdenes es:

$$s_i^{(2)} = \frac{\sum_j (r_{ij} - \bar{r}_i)^2}{(e-1)} \quad (31)$$

donde \bar{r}_i (ordenación media marginal = $\sum_j r_{ij}/e$) puede ser interpretado como el esperado para cada r_{ij} bajo la hipótesis de máxima estabilidad (órdenes iguales). Es decir, la expresión de máxima estabilidad sería cuando todas las ordenaciones de un genotipo en los distintos ambientes fuera la misma y por tanto, igual a la ordenación media marginal.

Otros dos estadísticos de Hühn son:

$$s_i^{(3)} = \frac{\sum_j (r_{ij} - \bar{r}_i)^2}{\bar{r}_i} \quad (32)$$

$$s_i^{(6)} = \frac{2\sum |r_{ij} - \bar{r}_i|}{\bar{r}_i} \quad (33)$$

$S^{(3)}_i$ y $S^{(6)}_i$ representan la suma de las desviaciones al cuadrado y de las desviaciones absolutas, respectivamente, de las ordenaciones respecto de la estabilidad máxima, expresada en unidades de ordenaciones medias marginales.

Ambos combinan rendimiento y estabilidad, pero según Kang y Pham (1991), $S^{(3)}_i$ parece dar más peso a estabilidad que $S^{(6)}_i$.

Como los genotipos se ordenan separadamente en cada ambiente, los efectos ambientales no influyen sobre la estabilidad definida por estos estadísticos. Sin embargo, las diferencias entre genotipos podrían afectar a estas medidas de estabilidad y conducir a aceptar diferencias de estabilidad aun cuando no exista interacción **GxE**.

Para superar esta ambigüedad se “corrigen” los datos originales Y_{ij} con el efecto genotípico estimado como la diferencia entre la media marginal del genotipo i y la media total del ensayo, es decir:

$$Y_{ij}^* = Y_{ij} - (\bar{Y}_{i.} - \bar{Y}_{..}) \quad (34)$$

De esta forma las desviaciones de la hipótesis nula (H_0 : que todos los genotipos tengan igual estabilidad) se puede atribuir ahora solamente a interacciones **GxE**.

Se han desarrollado para los estadísticos $S^{(1)}_i$ y $S^{(2)}_i$ pruebas aproximadas de significación basadas en una distribución normal. Así, se calculan los dos estadísticos siguientes:

$$Z^{(m)}_i = [S^{(m)}_i - E(S^{(m)}_i)] / \sqrt{V(S^{(m)}_i)}, \text{ para } m = 1, 2 \quad (35)$$

que podría aproximarse a una distribución chi-cuadrado con un grado de libertad y,

$$S^{(m)} = \sum_i Z^{(m)}_i, \text{ para } m = 1, 2 \quad (36)$$

que podría aproximarse a una distribución chi-cuadrado con **G** (número de genotipos) grados de libertad y donde $E(S^{(m)}_i)$ es la media de $S^{(m)}_i$ y $V(S^{(m)}_i)$ es la varianza de $S^{(m)}_i$.

Nassar y Hühn (1987) nos permiten calcular estos estadísticos por medio de unas sencillas fórmulas:





$$E(S^{(1)}_i) = (G^2 - 1) / (3G) \tag{37}$$

$$E(S^{(2)}_i) = (G^2 - 1) / 12 \tag{38}$$

$$V(S^{(1)}_i) = (G^2 - 1) [(G^2 - 4)(E + 3) + 30] / [45G^2E(E - 1)] \tag{39}$$

$$V(S^{(2)}_i) = (G^2 - 1) [2(G^2 - 4)(E - 1) + 5(G^2 - 1)] / [360E(E - 1)] \tag{40}$$

Con la prueba de $S^{(m)}$ con G grados de libertad deducimos si existen diferencias significativas en estabilidad entre los genotipos. Para probar la estabilidad de un sólo genotipo i , se aplica el estadístico $Z^{(m)}_i$ como una chi-cuadrado con un grado de libertad.

Se pueden seguir algunas recomendaciones para una mejor aplicación de estos estadísticos no-paramétricos. Las podemos sintetizar en tres puntos:

a) Para un análisis de estabilidad fenotípica se deben de utilizar los datos corregidos Y^*_{ij} si la queremos estimar independientemente de los efectos del nivel de rendimiento.

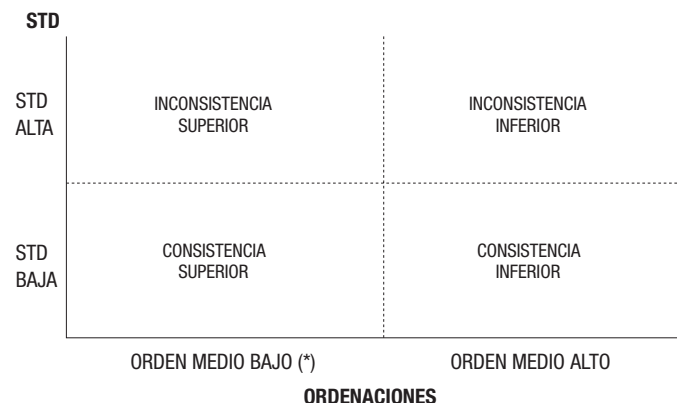
b) Para una estimación cuantitativa de estabilidad fenotípica es preferible $S^{(1)}_i$, por su sencillez de cálculo y de interpretación, existiendo para éste parámetro la prueba eficiente de significación descrita más arriba.

c) Si se está interesado en considerar simultáneamente estabilidad y rendimiento, los estadísticos no-paramétricos a aplicar son $S^{(3)}_i$ y $S^{(6)}_i$, los cuales miden la estabilidad en unidades de rendimiento. Pero deben utilizarse los datos originales Y_{ij} porque la transformación $Y_{ij} - Y^*_{ij}$ elimina el efecto genotípico de los datos.

5.2. Método de consistencia de comportamiento

Ketata y col. (1989) propusieron recientemente otro método basado en la utilización, para

Figura 7. Método de Ketata y col. 1989



(*) Orden medio bajo = rendimientos altos.

cada genotipo, del orden medio y de su desviación estandar (std) para construir una gráfica, la cual puede dividirse en cuatro zonas (Fig. 7). Como se puede ver en la Fig. 7 se seleccionarán como más estables y de alto rendimiento aquellos genotipos que esten en la zona de "consistencia superior".

También se puede construir la gráfica de medias genotípicas frente a su desviación estándar, aunque hay que tener en cuenta que los genotipos seleccionados podrían serlo desde un punto de vista del concepto estático o biológico de estabilidad.

5.3. Método de Kang

Kang (1988) ideó otro método no-paramétrico para seleccionar genotipos de alto rendimiento y estables. Para ello, se ordenan los rendimientos medios genotípicos en orden decreciente de tal forma que el genotipo de mayor rendimiento medio tenga el valor de orden=1. Paralelamente, y de forma similar, se establecen los órdenes de los genotipos atendiendo a la varianza de estabilidad de Shukla (σ^2_i), dando el valor de orden 1 al genotipo que tenga la menor. Posteriormente se suman ambos órdenes; de esta forma el genotipo que tenga menor orden-suma será el más deseable.

5.4. Método de ordenamientos estratificados

Este sencillo método no paramétrico fué ideado por Fox y col., (1990) y evalúa la proporción de ambientes en los que un genotipo tiene posiciones superiores, medias o bajas. Un genotipo que cuente con mayor proporción en el primer tercio se considerará como más adaptado.





5.5. Aplicación práctica

Los resultados obtenidos de la aplicación de cada uno de estos métodos a nuestros datos son los siguientes:

5.5.1. Método no paramétrico de Hühn

Las Tablas 9 y 10 presentan los valores para los estadísticos de Hühn. Para cada genotipo se ha calculado $Z^{(1)}_i$ y $Z^{(2)}_i$ basados sobre los ordenes de los datos corregidos y no corregidos.

Como puede verse en la Tabla 9, $\Sigma Z^{(1)}_i = 10.88$ y $Z^{(2)}_i = 14.33$. Ambos estadísticos son menores que el valor de la chi-cuadrado para el 5% y 11 grados de libertad ($\chi^2_{0.05,11} = 19.70$). Se puede concluir que no existen diferencias en estabilidad entre las 11 genotipos. De esta forma se acepta la hipótesis nula (es decir: no existen diferencias de estabilidad entre los genotipos del ensayo), por lo que no es necesario continuar con una comparación múltiple entre los valores $S^{(m)}_i$ ($m=1,2$).

A pesar de que no es necesario continuar con una comparación múltiple, podemos ver como en la Tabla 9, inspeccionando los valores individuales de $Z^{(m)}_i$, se encuentra que el genotipo 3 es significativo ($Z^{(m)}_i > \chi^2_{0.05,1} = 3.84$, $m = 1,2$). El error probabilístico esperado según la teoría de Probabilidades es:

$$(1) (0.05)^1 (0.95)^{10} = 0.33$$

mucho mayor que 0.05, que es el nivel elegido de probabilidad para que un genotipo muestre significación debido al azar.

De todo esto, podemos concluir que el genotipo 3 ha podido resultar significativo debido al azar y que el test o prueba global es más fidedigno que el individual indicándonos la no existencia de diferencias significativas de estabilidad entre genotipos.

Como para los estadísticos $S^{(m)}_i$ ($m = 3,6$) no se dispone de ninguna prueba de significación, los genotipos más estables serían aquellos que mostrasen los valores de $S^{(m)}_i$ más pequeños. Así, los genotipos más estables son: 9, 10, 6, 1 y 7 (Tabla 9).

Los resultados de estos estadísticos tomando como base de la ordenación los datos originales, distan mucho de lo ya expuesto (Tabla 10).

Así, $\Sigma Z^{(1)}_i = 82.56$ y $\Sigma Z^{(2)}_i = 47.76$ son altamente significativos, dado que $\Sigma Z^{(1)}_i$ y $\Sigma Z^{(2)}_i$ son mayores que el valor de $\chi_{0.05,11} = 19.7$, por lo que se puede decir que existen diferencias en estabilidad entre los 11 genotipos. Por tanto, procedemos a inspeccionar los valores individuales de $Z^{(m)}_i$, los cuales son significativos para los genotipos 1, 4, 5, 6, 7, 8, 9 y 10 para $S^{(1)}_i$, y 4, 5, 6, 7, 8, 9 y 10 para $S^{(2)}_i$, siendo altamente significativos para $S^{(1)}_i$ los genotipos **7** y **4**, ya que el valor de $\chi_{0.001,1} = 10.82$ es menor que $Z^{(1)}_4 = 12.142$ y $Z^{(1)}_7 = 24.811$.

En cuanto a los estadísticos $S^{(m)}_i$ ($m = 3,6$), el genotipo 7 resulta ser el más estable, siguiéndole 5 y 6.

Comentemos brevemente los efectos que la transformación de los datos originales Y_{ij} a los corregidos Y^*_{ij} ha ocasionado sobre los resultados de los distintos estadísticos no-paramétricos. Huehn (1990a,b) demostró que para los

Tabla 9. Cálculo de los estadísticos $S^{(m)}_i$ ($m = 1,2,3,6$) Datos Y^*_{ij} (transformados).

GENOTIPO	$s^{(1)}_i$	$s^{(2)}_i$	$z^{(1)}_i$	$z^{(2)}_i$	$s^{(3)}_i$	$s^{(6)}_i$
1	3.311	7.788	0.278	0.487	7.537	2.537
2	4.666	15.777	2.794	3.331	15.777	4.000
3	4.866	18.322	3.985*	6.910**	20.873	4.430
4	3.733	10.044	0.024	0.002	10.512	2.884
5	4.111	12.011	0.593	0.403	11.623	3.333
6	3.511	8.488	0.041	0.227	7.958	2.500
7	4.022	11.433	0.392	0.205	10.394	2.747
8	4.022	12.055	0.392	0.421	12.764	3.294
9	3.022	6.844	0.993	0.993	6.695	2.174
10	2.911	6.322	1.385	1.349	6.393	2.179
11	3.644	9.733	0.002	0.007	9.954	2.954
$s^{(m)} = \Sigma z^{(m)}_i$			10.88	14.33		

Tabla 10. Cálculo de los estadísticos $S^{(m)}_i$ ($m = 1,2,3,6$). Datos Y_{ij} (originales).

GENOTIPO	$s^{(1)}_i$	$s^{(2)}_i$	$z^{(1)}_i$	$z^{(2)}_i$	$s^{(3)}_i$	$s^{(6)}_i$
1	2.400	4.044	4.024*	3.539	3.500	1.538
2	3.233	8.247	0.427	0.306	7.385	2.636
3	2.488	5.511	3.466	2.010	8.552	3.034
4	1.488	2.322	12.142***	5.882*	4.098	1.882
5	2.233	3.692	5.183*	3.971*	3.006	1.330
6	1.844	2.722	8.454**	5.285*	2.130	1.043
7	0.566	0.336	24.811***	9.318**	0.221	0.358
8	2.111	3.211	6.125*	4.598*	4.737	2.492
9	1.977	3.211	7.243**	4.598*	4.070	1.887
10	1.922	2.581	7.736**	5.493*	2.716	1.578
11	2.577	4.733	2.951	2.767	4.392	1.855

$s^{(m)} = \sum z^{(m)}_i$

82.56* 47.76*****

parámetros clásicos de estabilidad (varianza ambiental, ecovalencia, coeficiente de regresión y S^2_{di}) los resultados obtenidos eran idénticos tanto para los datos originales Y_{ij} como para los transformados Y^*_{ij} . De esta forma, esta transformación sólo afecta a los estadísticos no-paramétricos, cambiando sus ordenaciones y haciendo que éstas sean muy diferentes según se opere con Y_{ij} o con Y^*_{ij} .

El gran efecto de la transformación queda demostrado observando los resultados de las pruebas de significación en las Tablas 10 y 11. Esta transformación reduce drásticamente los niveles de significación individuales (significaciones basadas sobre $Z^{(1)}_i$ y $Z^{(2)}_i$) y globales (significación para diferencias en estabilidad entre todos los genotipos). Por tanto, el utilizar los datos originales supone un mayor nivel de discriminación (es decir, se tienen mas posibilidades de encontrar mas diferencias significativas entre genotipos), que si se opera con los datos transformados.

De esta forma, la reducción en significación para los datos corregidos puede ser considerada como un aumento de fiabilidad si se está interesado en una caracterización de la estabilidad de los genotipos independientemente del nivel de rendimiento.

5.5.2. Método de "consistencia de comportamiento"

Un genotipo se considera estable si sus ordenaciones son relativamente consistentes en todos los ambientes. En la Figura 8 se representan las ordenaciones medias de cada genotipo en todos los ambientes en función de las desviaciones típicas de dichas ordenaciones y en la Figura 9 los rendimientos genotípicos medios frente a sus desviaciones típicas. Los resultados obtenidos para nuestros datos pueden verse en la Tabla 11, base para las representaciones gráficas de las Figuras 8 y 9.

Figura 8. Consistencia del comportamiento (órdenes)

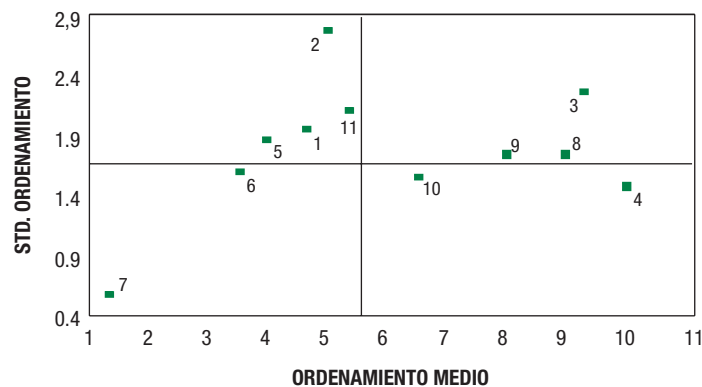
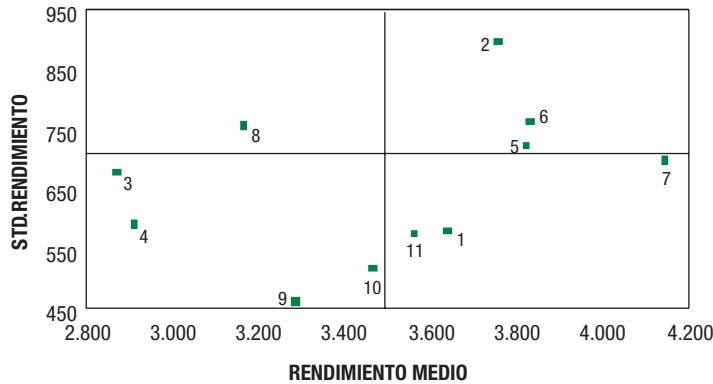


Figura 9. Consistencia del comportamiento (rendimiento)



En la Fig. 8 los genotipos que muestran una adaptación general se encontrarán en el cuadrante denominado “Consistencia Superior” (ordenamiento bajo y desviación típica baja), mientras que genotipos que muestran una adaptación específica se encuentran en el cuadrante de “Inconsistencia Superior”. Así, los genotipos seleccionados para amplia adaptación son el 7 y el 6, mientras que los genotipos 1, 2, 5 y 11 pueden ser seleccionados para ambientes específicos.

En la Fig. 9 los genotipos seleccionados por poseer un mayor rendimiento y estabilidad son 1, 7 y 11 que se sitúan en el cuadrante inferior derecha al poseer una desviación típica pequeña y un rendimiento superior a la media del ensayo.

Como se puede observar este método tiene la ventaja de su sencillez en los cálculos y su poder resolutivo en las gráficas. En éstas ya podemos estimar rápidamente qué genotipos

son más estables y de mayor rendimiento. Nos combina, pues, en una gráfica, estabilidad y rendimiento.

5.5.3. Método de Kang

En éste método se asigna un número de orden a cada valor de rendimiento medio, dando el valor de 1 al rendimiento más alto. También se asignan las ordenaciones a los valores de σ^2_i (varianza de Shukla; ver Capítulo 3), dando el valor de 1 a la varianza σ^2_i más pequeña. Las dos ordenaciones se suman para cada genotipo. El valor más pequeño, resultante de esta suma, daría el genotipo más deseable. La misma operación se puede efectuar con s^2_i , de tal forma que podemos sumar ambos valores resultantes de $\sigma^2_i + \text{rendimiento}$ y de $s^2_i + \text{rendimiento}$, obteniendo el orden final para cada genotipo. Los resultados obtenidos de la aplicación del método de Kang para nuestros datos se presentan en la Tabla 12.

Tabla 11. Resultados según el método de la consistencia de comportamiento.

GENOTIPOS	REND. MEDIO	STD. REND.	RANGO MEDIO	STD. RANGO
1	3699.38	581.971	4.60	2.011
2	3825.07	898.333	4.95	2.872
3	2877.50	678.528	9.20	2.347
4	2920.35	593.609	9.90	1.524
5	3896.25	723.460	3.95	1.921
6	3907.16	764.045	3.50	1.649
7	4241.11	700.386	1.35	0.579
8	3193.40	758.453	8.90	1.791
9	3320.56	463.358	7.90	1.792
10	3517.99	518.824	6.45	1.606
11	3617.71	578.907	5.30	2.175

STD.: Desviación típica.

Tabla 12. Método de Kang

GENOTIPO	σ^2_i	s^2_i	Rend.	$\sigma^2_i + \text{Rend.}$	$s^2_i + \text{Rend.}$	Suma	Orden
1	2	3	5	7	8	15	3
2	10	10	4	14	14	28	9
3	11	11	11	22	22	44	11
4	7	8	10	17	18	25	8
5	8	9	3	11	12	23	7
6	6	5	2	8	7	15	2
7	5	6	1	6	7	13	1
8	9	7	9	18	16	34	10
9	3	2	8	11	10	21	6
10	1	1	7	8	8	16	4
11	4	4	6	10	10	20	5

Por medio de la columna " σ^2_i +rendimiento" (Tabla 12) se pueden seleccionar los genotipos 1, 7, y 10 (en los cuales σ^2_i no es significativamente distinta de cero, ver Tabla 5) y en la columna " s^2_i +rendimiento" los genotipos 1, 6, y 10. Y en conjunto (columna "Orden"), los genotipos seleccionados serían el **7, 6, 1 y 10** (por este orden de preferencia). Como vemos los genotipos seleccionados difieren según utilicemos una columna u otra en un sólo genotipo, debido a las diferencias de significación de σ^2_i y s^2_i . Los genotipos que se consideraban estables (s^2_i no significativa) después de la utilización del covariante (índice ambiental) incluyen el genotipo 6 y excluyen el 7 (ver Capítulo, 3, Tabla 5).

Resulta ser, por tanto, un método útil para seleccionar simultáneamente para rendimiento y estabilidad de rendimiento. Además, se trata de un método realmente polivalente. Así, si estamos interesados en dar más peso al rendimiento o a la estabilidad, se podría hacer sin más que multiplicar por un coeficiente los órdenes de rendimiento o de estabilidad. Por ejemplo, si estamos interesados en dar el doble de peso al efecto rendimiento que a la estabilidad lo que hacemos es multiplicar por 2 el orden obtenido para los rendimientos medios y le sumamos el orden correspondiente a la varianza de Shukla (= $2 \times \text{orden rendimiento} + \text{orden } \sigma^2_i$).

5.5.4. Método de ordenaciones estratificados

Es un método muy simple que consiste en dibujar un histograma en el que quede reflejado el número de ambientes para los cuales un genotipo ha mostrado posiciones altas, medias o bajas. Los valores obtenidos según este método para cada genotipo pueden verse en la Tabla 13, y su representación gráfica en las Figuras 10, 11, y 12.

Un genotipo con mayor porcentaje de números de orden en la clase 1-3 será considerado como relativamente bien adaptado a todos los ambientes del ensayo de nuestro ejemplo. Sin embargo, con éste método no se identifican los genotipos adaptados específicamente a un cierto ambiente.

De la visualización de las Figuras 10, 11 y 12 podemos concluir que los genotipos seleccionados por su amplia adaptación serían los representados en la Fig. 10, es decir, los genotipos 7, 6 y 2, los cuales se caracterizan por contener la mayor parte de sus números de orden en la clase 1-3.

Este método resulta, como hemos visto, simple y sencillo y, como todos los métodos no-paramétricos, independiente de la escala, además de dar igual peso a cada ambiente.

Figura 10. Genotipos 2-6-7 (órdenes bajos)

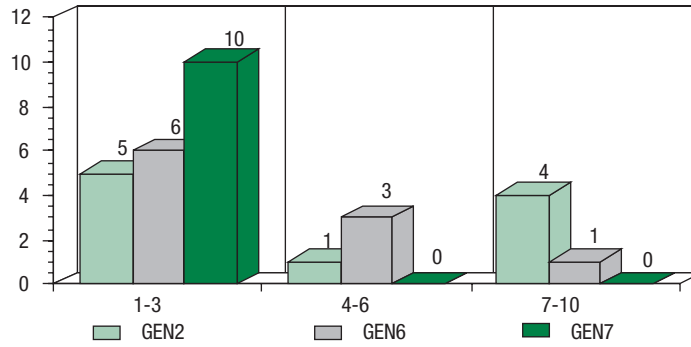


Tabla 13. Estratificación de rangos.

GENOTIPOS	CLASES DE ÓRDENES (Posiciones)			ORDEN FINAL
	1 - 3(*)	4 - 6	7 - 10	
1	3	5	2	5
2	5	1	4	3
3	0	2	8	8.5
4	0	1	9	10.5
5	3	7	0	4
6	6	3	1	2
7	10	0	0	1
8	0	1	9	10.5
9	0	2	8	8.5
10	0	5	5	7
11	3	3	4	6

(*) Número de ambientes en los cuales cada genotipo tiene un número de orden comprendido entre 1 y 3.

Figura 11. Genotipos 10-11-1-5 (órdenes medios)

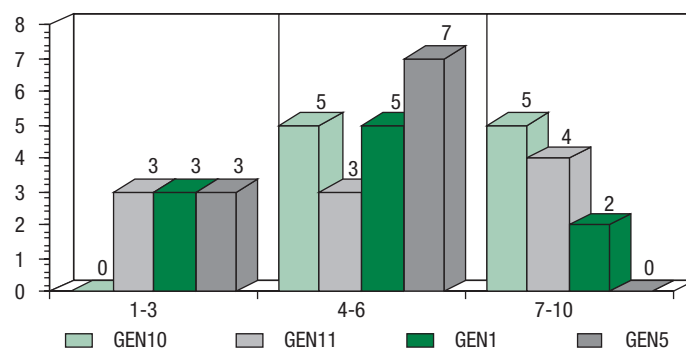
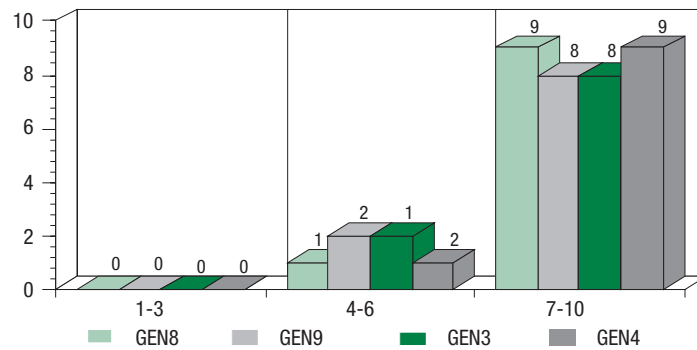


Figura 12. Genotipos 8-9-3-4 (órdenes altos)

REFERENCIAS CITADAS

- FOX, P.N., B. SKOVMAND, B.K. THOMPSON, H.J. BRAUN and R. CORMIER. 1990: Yield and adaptation of hexaploid spring triticale. *Euphytica* 47, 57-64.
- HUEHN, M. 1990b: Nonparametric measures of phenotypic stability. Part 2: Applications. *Euphytica* 47, 195-201.
- HUEHN, M. 1990a: Nonparametric measures of phenotypic stability. Part 1: Theory. *Euphytica* 47, 189-194.
- HÜHN, VON M. 1979: Beiträge zur erfassung der phänotypischen stabilität. *EDV in Medizin und Biologie* 10 (4), S. 112-117.
- KANG, M.S. 1988: A Rank-sum method for selectig high -yielding, stable corn genotypes. *Cereal Res.Comm.*16, 113-115.
- KANG, M. S, and H. N. PHAM. 1991:..Simultaneous Selection for High Yielding and Stable Crop Genotypes. *Agronomy journal.* 83, 161-165.
- KETATA, H., S.K. YAN and M. NACHIT. 1989: Relative consistency performance across environments. *Int. Sym. on Physiology and Breeding of Winter Cereals for stressed Mediterranean Environments.* Montpellier. July 3-6, 1989.
- NASSAR, R., and M. HÜHN, 1987: Studies on estimation of phenotypic stability: Test of significance for nonparametric measures of phenotypic stability. *Biomet.* 43, 45-53.

OTRAS REFERENCIAS ACONSEJADAS

- KANG, M.S., D.P. GORMAN and H.N. PHAM. 1991: Application of stability statistic to international maize yield trials. *Theor. Appl. Genet.* 81, 162-165.
- SHUKLA, G.K., 1972: Some Statistical aspects of partitioning genotype-environmental components of variability. *Heredity* 29, 237-245.

6

Σ α Π Ω ~~X~~ \pm ϑ
MÉTODO
DE ESKRIDGE

6. MÉTODO DE ESKRIDGE

La estabilidad de un genotipo la fundamentó Eskridge (1990) en cuatro diferentes modelos, según los parámetros de estabilidad incluidos en el análisis. Así, los parámetros incluidos en cada modelo son:

- 1) La varianza ambiental (S^2_{xi}) (3.4).
- 2) El coeficiente de regresión (b_i) de Finlay y Wilkinson (2.2.3).
- 3) La varianza (σ_i) de Shukla (3.1).
- 4) El coeficiente de regresión (b_i) de Finlay y Wilkinson y el cuadrado medio residual (S^2_{di}) de Eberthart y Russell (2.2.3).

Denominaremos respectivamente EV, FW, SH y ER los valores de Eskridge resultantes. Estos cuatro diferentes estadísticos de estabilidad se incorporan a la ecuación de Kataoka (1963):

$$\bar{Y}_i - Z_{(1-\alpha)} (V_i)^{1/2} \tag{41}$$

donde:

- \bar{Y}_i es la media genotípica del genotipo i .
- $Z_{(1-\alpha)}$ es el percentil $(1-\alpha)$ de una distribución Normal.
- V_i es una medida de estabilidad del genotipo i .

Veamos con más detalle como estas cuatro medidas de estabilidad se incorporan a la ecuación [41]:

1) En términos estadísticos, la verdadera media del genotipo i -ésimo respecto a todos los ambientes es μ_i con varianza ambiental σ^2_{io} . Los valores de μ_i y de σ^2_{io} se estiman por medio de Y_i (media genotípica en todos los ambientes) y S^2_{xi} . Estas estimaciones sustituyen respectivamente a Y_i y V_i en [41], dando el índice denotado por EV ($EV = Y_i - Z_{(1-\alpha)} (S^2_{xi})^{1/2}$). Utilizando este índice se podrán seleccionar genotipos con medias genotípicas mayores y varianzas ambientales menores.

2) Según éste modelo se selecciona un cultivar con media grande y pendiente (b_i) próxima a uno. La ecuación [41] podría utilizarse para sopesar la importancia del rendimiento medio según el coeficiente de Finlay y Wilkinson si se define una medida de varianza que nos cuantifi-

que cuanto se desvía de 1 la pendiente de un genotipo. Se demuestra (Eskridge, 1990) que la media y la varianza poblacionales se estiman por los valores de Y_i y de $(b_i-1)^2 S^2_y (1-1/e)$, respectivamente. Estos valores se sustituyen en [41] dando el índice FW ($FW = Y_i - Z_{(1-\alpha)} [(b_i-1)^2 S^2_y (1-1/e)]^{1/2}$). Cuando todos los genotipos tienen pendiente 1, el índice FW tiende al rendimiento medio.

3) Utilizando la varianza de Shukla, la media poblacional y varianza del genotipo i son μ_i y $\sigma^2_e + \sigma^2_i$, donde σ^2_e y σ^2_i son la varianza de los efectos debidos al azar de los ambientes y la varianza poblacional de Shukla, respectivamente (Eskridge, 1990). La media poblacional se estima por Y_i , σ^2_i por la varianza de Shukla y σ^2_e por medio del Cuadrado Medio del error en un ANOVA. $\sigma^2_e + \sigma^2_i$ sustituye a V_i en [41], resultando el índice SH ($SH = Y_i - Z_{(1-\alpha)} [\sigma^2_e + \sigma^2_i]^{1/2}$).

4) Siguiendo un razonamiento similar al utilizado para la obtención del índice FW, se demuestra (Eskridge, 1990) que la media y la varianza poblacionales según el método de Eberhart y Russell pueden estimarse por los valores de Y_i y $(b_i-1)^2 S^2_y (1-1/e) + S^2_{di}$, respectivamente. Esta varianza representa la proporción de la varianza total del genotipo i -ésimo, que es debida por un lado a la desviación de la pendiente de 1 y por otro a la desviación de la regresión lineal. Estos valores sustituidos en [41] resultan en el índice ER ($ER = Y_i - Z_{(1-\alpha)} [(b_i-1)^2 S^2_y (1-1/e) + S^2_{di}]^{1/2}$).

La forma de calcular algunos parámetros antes comentados es la siguiente:

$$S^2_y = \Sigma (Y_{.j} - Y_{..})^2 / (e-1) \tag{42}$$

la varianza de los efectos ambientales (σ^2_e) se calcula teniendo en cuenta nuestro modelo mixto, donde el efecto ambiental es aleatorio, así, la varianza se despeja de la siguiente forma (ver Tabla 1):

$$\sigma^2_e = [CM(E) - CM(RE)] / gr \tag{43}$$

$$CM(E) = (\Sigma Y^2_{.j} / gr - FC) / (e-1) \tag{44}$$

$$CM(RE) = [\Sigma \Sigma Y^2_{.jk} / g - SC(E) - FC] / [e(r-1)] \tag{45}$$

Un valor concreto de α (o de $Z(1-\alpha)$) indica el grado subjetivo del mejorador para aceptar bajos rendimientos expresado en términos de probabilidad. Se utilizan valores pequeños de α



cuando la obtención de rendimientos bajos presupone una serie de costes sociales altos (banca rota, hambre, ..etc), como sería el caso de una agricultura de subsistencia. Por ello hemos elegido un $\alpha = 0.05$ [$Z(0.95) = 1.645$], que significaría que de 20 rendimientos medios existe la probabilidad del 5% (o sea, una sola vez) de obtener un rendimiento bajo en un ambiente determinado.

6.1. Caso práctico

Los valores obtenidos para los índices del método de “**primera seguridad**”, de Eskridge se dan en la Tabla 14. Como vemos existen diferencias en las ordenaciones de los genotipos de un índice a otro, aunque el mejor genotipo (el 7) es el mismo para los 4 índices.

Estos cambios en las ordenaciones genotípicas de los índices pueden ser debidos a que el ordenamiento de los genotipos ha resultado ser muy sensible al nivel α especificado y a la elección del parámetro de estabilidad introducido en la fórmula [41]. Un $\alpha = 0.05$ indica que existirían problemas graves si se diera una bajada de rendimiento. Pueden utilizarse otros valores de α . Idealmente, el mejorador podría determinar el valor medio de α para la situación agrícola que es objeto del programa de mejora. Si no se pudiera determinar directamente, podría deducirse a partir de investigaciones previas. Por ejemplo, para agricultores de subsistencia las preferencias de riesgo han mostrado tener valores de $Z(1-\alpha)$ de 0.5 y 1.5, correspondiendo a valores de α comprendidos entre 0.3 y 0.05.

Tabla 14. Valores para los cuatro índices de “primera-seguridad” y las ordenaciones respectivas (*)

GENOTIPO	EV.	FW.	SH.	ER.	REND. MED.
1	2742.03(2)	3667.55(3)	2775.56(4)	3330.74(3)	3699.37(5)
2	2347.31(8)	3433.53(6)	2645.19(6)	3054.59(7)	3825.07(4)
3	1761.32(11)	2337.94(11)	1306.66(11)	1632.88(11)	2877.50(11)
4	1943.86(10)	2809.92(10)	1878.11(10)	2330.93(10)	2920.35(10)
5	2706.15(3)	3775.47(2)	2847.34(3)	3296.54(4)	3896.24(3)
6	2650.30(6)	3667.06(4)	2899.60(2)	3386.96(2)	3907.15(2)
7	3088.97(1)	4117.00(1)	3249.35(1)	3739.01(1)	4241.11(1)
8	1945.74(9)	3006.04(9)	2138.49(9)	2588.31(9)	3193.40(9)
9	2558.33(7)	3114.28(8)	2393.39(8)	2958.32(8)	3320.56(8)
10	2664.52(5)	3410.56(7)	2621.36(7)	3221.83(5)	3517.98(7)
11	2665.40(4)	3549.35(5)	2653.94(5)	3165.60(6)	3617.71(6)

(*) Ordenaciones: números entre paréntesis.

Como se observa en la Tabla 14, la elección de una determinada definición de estabilidad tiene un gran impacto sobre el ordenamiento de los genotipos. Así:

a) **La estabilidad de tipo 1** (S^2_{xi}) depende de la diversidad de ambientes del experimento. Si los ambientes son muy distintos entonces la estabilidad de tipo 1, y en consecuencia el índice EV, no resulta muy útil. Sin embargo, si los ambientes del ensayo son muy parecidos, entonces la estabilidad de tipo 1 y, por tanto, el índice EV, podrían ser útiles.

b) **La estabilidad de tipo 2** (b_i y α^2_i) considera un cultivar como estable si su respuesta ambiental es paralela a la respuesta media de todos los cultivares del ensayo. Este tipo de esta-

bilidad puede ser útil cuando el ensayo se realiza en un grupo de ambientes diversos. Consecuentemente los índices FW y SH pueden ser útiles para evaluar un grupo de genotipos entre sí en un amplio margen de ambientes, pero se debe extremar la precaución a la hora de realizar extrapolaciones a genotipos no incluidos en el ensayo.

c) **La estabilidad de tipo 3** (S^2_{di}) mide irregularidades impredecibles de la respuesta de un genotipo a los diversos ambientes (en contraste con la de tipo 2, que mide la parte predecible de la respuesta). El índice ER incluye ambos tipos de estabilidad, tipos 2 y 3, siendo por ello un índice más completo que FW y que SH.

Como resumen, se puede decir que la utilización de los índices del método de “primera

seguridad” tiene ciertas ventajas. Están basados sobre la razonable suposición de que el mejorador tiene como principal preocupación eliminar todo tipo de riesgos para una posible elección de un cultivar con elevada probabilidad en dar rendimientos bajos. Asimismo, poseen un interés intrínseco por ser, estos índices, límites inferiores de confianza. En suma, pondera la importancia relativa del dúo estabilidad-rendimiento. Finalmente, este método es útil para ser aplicado a los tres tipos de estadísticos univariantes de estabilidad definidos por Lin y col. 1986 (véase Capítulo 1).

Sin embargo, la aplicación de éstos índices puede estar limitada por una serie de desventajas. Así, puede resultar difícil encontrar un valor realista de α , sobre todo si hay que tener en cuenta una situación agrícola concreta. Por otro lado, los índices están basados sobre varianzas estimadas, las cuales pueden mostrar, por sí mismas, un gran error en su estimación. Unas estimaciones imprecisas de varianzas pueden producir ordenaciones de genotipos de valor dudoso o nulo. Finalmente, según Eskridge, se pueden desarrollar otros índices alternativos a los expuestos por él, siguiendo a Roy (1952) y Telser (1955).



REFERENCIAS CITADAS

• EBERHART, S.A. and W.A. RUSSELL. 1966: Stability Parameters for Comparing Varieties. *Crop Sci.* 6, 36-40.

• ESKRIDGE, K.M. 1990: Selection of stable cultivars using a safety-first rule. *Crop Sci.* 30, 369-374.

• FINLAY, K.W. and G.N. WILKINSON, 1963: The analysis of adaptation in a plant breeding programme. *Aust. J. Agric. Res.* 14, 742-754.

• KATAOKA, S. 1963: A stochastic programming model. *Econometrica* 31, 181-196.

• LIN, C.S., M.R. BINNS, and L.P. LEFKOVITCH, 1986: Stability analysis: Where do we stand?. *Crop Sci.* 26, 894-900.

• ROY, A.D. 1952: Safety-first and the holding of assets. *Econometrica* 20, 431-449.

• SHUKLA, G.K., 1972: Some Statistical aspects of partitioning genotype-environmental components of variability. *Heredity* 29, 237-245.

• TELSER, L.G. 1955: Safety-first and holding. *Rev. Econ. Stud.* 23, 1-16.

OTRAS REFERENCIAS ACONSEJADAS

• DILLON, J.L. and P.L. SCANDIZZO. 1978: Risk attitudes of subsistence farmers in northeast Brazil: a sampling approach. *Am. J. Agric. Econ.* 60, 425-435.

• ESKRIDGE, K.M., and B.E. JOHNSON. 1991: Expected utility maximization and selection of stable plant cultivars. *Theor. Appl. Genet.* 81, 825-832.

• HAZELL, P.B.R. 1982: Application of risk preference estimates in firm-household and agricultural sector models. *Am. J. Agric. Econ.* 64, 384-390.



7

Σ α Π **MÉTODO ESTRELLA** ϑ

7. MÉTODO ESTRELLA

El método Estrella (Flores, 1993) consiste en dibujar una estrella con tantos radios como ambientes, correspondiendo cada radio a un ambiente. En cada radio situamos, a una escala conveniente, el rendimiento medio en ese ambiente para un genotipo dado. Uniendo cada uno de éstos puntos, se obtiene una estrella representativa del comportamiento de cada variedad (o genotipo, en general).

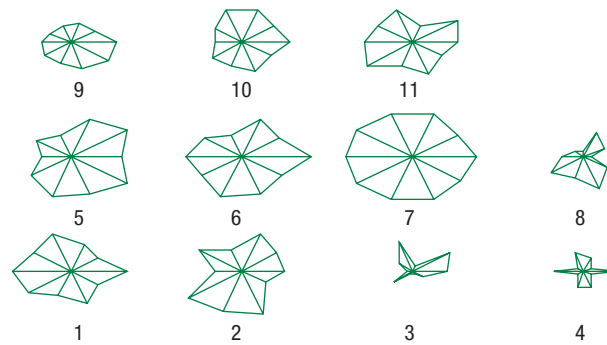
El rendimiento medio será proporcional al área de la estrella y cuanto más tienda la forma de ésta a la de una circunferencia, más estable será el comportamiento de la variedad

correspondiente. Así pues, la estabilidad puede estimarse de una forma visual en éste método.

Los dibujos de las estrellas se presentan en la Fig. 13. En ella podemos observar como los genotipos 7, 6, 5, 2 y 1 son los de mayor área, siendo, entre ellos, el genotipo 7 el que más tiende a la forma circular, siendo, por tanto, según éste método, el más estable.

La principal ventaja es la sencillez y rapidez de resultados, y aún más cuando algunos genotipos sobresalen de entre los demás. Como desventaja, no existe una prueba de significación satisfactoria.

Figura 13. Método "Estrella"



REFERENCIAS CITADAS

- FLORES, F. 1993: Interacción Genotipo-Ambiente en *Vicia faba* L. Tesis Doctoral. E.T.S.I.A.M. Córdoba.



8

ANÁLISIS DE COMPONENTES PRINCIPALES

8. ANÁLISIS DE COMPONENTES PRINCIPALES

El análisis de componentes principales (ACP) parte de una tabla de datos compuesta por **e** caracteres (**ambientes** en nuestro caso) observados en **g** individuos (**genotipos**). Estos individuos se consideran como puntos en un espacio de **e** dimensiones, determinado por unos ejes que representan a los caracteres.

El análisis consiste en transformar el conjunto de variables Y_1, Y_2, \dots, Y_e en un nuevo conjunto X_1, X_2, \dots, X_e en el que cada X_i es una combinación lineal de las Y_j , o sea:

$$X_i = a_{i1}Y_1 + a_{i2}Y_2 + \dots + a_{ie}Y_e \quad (46)$$

siendo la suma de los cuadrados de los coeficientes a_{ij} igual a la unidad, por ser las a_{ij} los cosenos directores de la recta representada por la ecuación (46):

$$\sum_{j=1}^e a_{ij}^2 = 1 \quad (47)$$

Las componentes principales se obtienen diagonalizando la matriz de covarianzas **C**:

$$\mathbf{C} = \mathbf{A} \mathbf{D}_\lambda \mathbf{A}' \quad (\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I}) \quad (48)$$

siendo \mathbf{D}_λ la matriz diagonal $\mathbf{D}_\lambda = \text{diag}(\lambda_1, \dots, \lambda_e)$ que contiene los valores propios de **C**. Estos valores se obtienen resolviendo la ecuación $|\mathbf{C} - \lambda_i \mathbf{I}| = 0$.

Se demuestra que λ nunca puede ser negativo y que $\sum_i \lambda_i = e$. Como cada λ lleva asociado un vector propio, y estos son ortogonales entre sí, hemos obtenido unas nuevas variables en función de las antiguas:

$$X_i = \sum_j a_{ij} Y_j \quad (49)$$

que están descorrelacionadas entre sí.

Se demuestra asimismo que la varianza de X_i es λ_i , lo que implica que cada valor propio explica un porcentaje $\lambda_i^* (100/e)$ de la variación total registrada y además la primera componente principal X_1 es la que tiene mayor varianza, siendo X_2 la que tiene mayor varianza de todas las nuevas variables no correlacionadas con X_1 , y X_3 la que tiene mayor varianza de todas las nuevas variables no correlacionadas con X_1 y X_2 , ... etc. hasta definir el conjunto de nuevas variables. Para el tratamiento matemático que conduce a la breve descripción anterior, véase Kendall and Stuart (1961) y Cuadras (1981).

Con pocos vectores, pues, podemos explicar, en ciertos casos, gran cantidad de la variación total. Estudiando los componentes de los nuevos ejes puede establecerse cuáles han sido los caracteres que más han contribuido a conseguir la separación mostrada por el eje en cuestión; estos caracteres serán aquellos para los cuales el coeficiente correspondiente ($a_{ij}\sqrt{\lambda_i}$) es mayor en valor absoluto. El caso ideal es aquel en que el coseno que forma la variable antigua con la nueva ($a_{ij}\sqrt{\lambda_i}$) es próximo a la unidad, lo que indica que la nueva variable coincide en dirección con la antigua, diciéndose entonces que aquella tiene el "sentido físico" dado por ésta.

Las componentes principales se pueden obtener también partiendo de la matriz de correlaciones **R**. Las componentes principales extraídas de la matriz **C** son distintas de las obtenidas de la matriz **R**, es decir, que el método *no* es invariante respecto a la matriz de partida.

La elección entre **C** ó **R** es un tema controvertido. Si las unidades de medida de las variables son distintas (años, kilos, metros,..etc.), es preferible el uso de **R**, porque equivale a utilizar variables reducidas y, por tanto, sin dimensión física. Pero si las unidades de medida son las mismas (como es nuestro caso) o razonablemente comensurables, es preferible realizar el análisis sobre **C**, que es menos artificial

Como resumen, la finalidad de las componentes principales es la de simplificar la estructura de los datos, sin obedecer a un modelo fijado "a priori", para poder explicar, en pocas componentes, la mayor parte de la información que contienen las variables. El análisis de componentes principales (ACP) es uno de los métodos más antiguos de ordenación indirecta. Representamos la matriz de datos inicial como una matriz **GxE**, **G** unidades u observaciones (genotipos) y **E** variables (ambientes). El propósito de éste análisis es producir una ordenación de las observaciones en un pequeño número de dimensiones, las cuales explican la mayor cantidad de variación original.

Se pueden realizar cuatro análisis de componentes principales en el estudio de la Interacción **GxE**:

1) ACP en el que los genotipos son las variables y los ambientes las observaciones, partiendo de los datos originales de rendimiento.

2) Lo mismo que en 1 pero partiendo de datos transformados (Y_{ij}^*):



$$Y_{ij}^* = \bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} \quad (50)$$

es decir, eliminando del dato original los valores medios genotípicos y ambientales.

3) ACP en el que los ambientes son las variables y los genotipos las observaciones, partiendo de los datos originales.

4) Lo mismo que en 3 pero con datos transformados de forma similar a 2.

Estos análisis, para nuestro ejemplo, se han realizado con el programa NTSYS de la siguiente forma:

a) Partimos de la matriz de datos 10 (ambientes) x 11 (genotipos) para los casos 1) y 3) anteriores y de la matriz 10 x 11 transformada según [50] para la obtención de los componentes, que representan ciertamente a la interacción **GxE**, para los casos 2) y 4) (ver Apéndice).

b) Calculamos la matriz de covarianzas **C** con la instrucción "Simil Interval", dándole en el Coeff=VARCOV.

c) A partir de **C** se calculan sus valores propios y vectores propios asociados, con la instrucción EIGEN.

d) Por último, se realiza la proyección de los vectores propios (con la instrucción PROJ) obteniéndose la gráfica de ambientes (Figs. 14 y 15) si hemos partido de genotipos como variables (casos 1 y 2) y la de genotipos (Figs. 16 y 17) si se ha partido de ambientes (casos 3 y 4).

8.1. Caso práctico

Los resultados de los cuatro análisis de Componentes Principales para nuestros datos se presentan a continuación. Dado que en este último decenio la utilización del análisis de Componentes Principales se ha realizado como componente del análisis AMMI, no se tratará con mucha profundidad en este capítulo.

8.1.1. ACP de Genotipos. Datos originales de rendimiento

Los primeros componentes explican, respectivamente el 76.91 y 14.08 % de la varianza. La proyección de las coordenadas ambientales para los dos primeros componentes principales se representan en la Figura 14.

Se comprueba que para el primer componente, el ordenamiento de las coordenadas ambientales se corresponde exactamente con el obtenido con los rendimientos medios ambientales (Fig. 14: en la parte derecha están los ambientes de alto rendimiento, y en la parte izquierda los de bajo rendimiento). Así pues, este componente representa el efecto ambiental. En cuanto al segundo componente no se aprecia un patrón obvio de respuesta.

8.1.2. Datos transformados

Los dos primeros componentes explican el 68.24 % y el 10.73 % de la varianza, respectivamente. Las coordenadas ambientales para los dos primeros componentes principales se han representado en la Figura 15.

Figura 14. ACP datos sin transformar (ambientes)

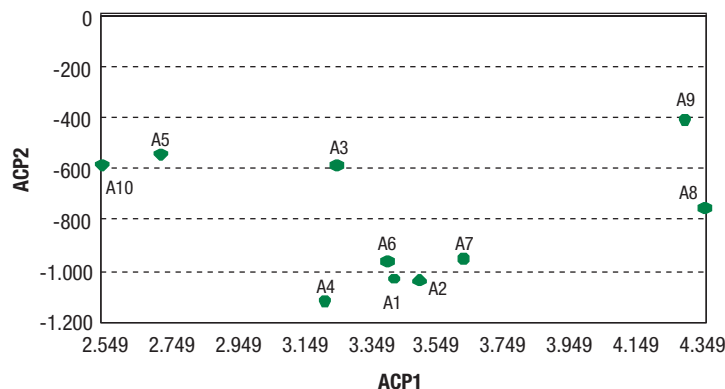
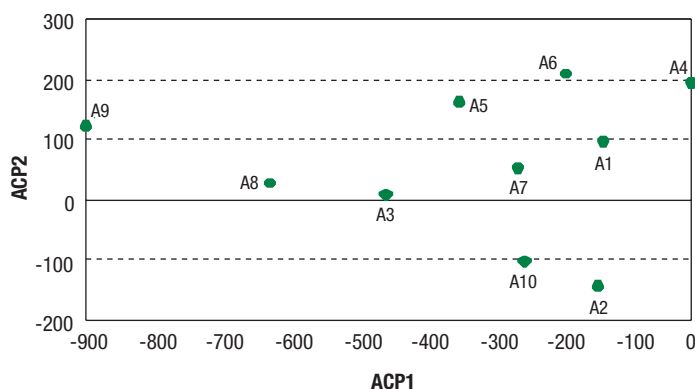


Figura 15. ACP datos transformados (ambientes)



Partir de los datos transformados tiene una justificación: que de alguna forma el análisis de componentes principales nos ayude a definir qué ambientes y qué genotipos muestran un grado menor de interacción. Es decir: identificar, sobre todo, aquellos genotipos que engrosan en menor cuantía la interacción **GxE** total. Por tanto, en la Figura 15 podemos esperar que la distribución de los ambientes obedezca a una gradación en el aporte a la interacción **GxE**. Cuando veamos el análisis A.M.M.I. en el capítulo siguiente recordaremos la coincidencia de resultados con los obtenidos en este apartado. Se puede observar como en la Fig. 15 el ordenamiento de los ambientes según el primer componente principal ACP1 (de izquierda a derecha: A9, A8, A3, ..., A4) coincide completamente con el ordenamiento que se puede realizar en la Fig. 18 respecto del eje de las coordenadas IACP1 ambientales (de la parte superior de la gráfica a la inferior: A9, A8, A3, ..., A4).

Como vemos existe una relación entre interacción GxE y la distribución de los ambientes dada por el ACP (Fig. 15), pero, así como en el análisis AMMI (ver Capítulo siguiente) el aporte que cada

ambiente ejerce sobre la Interacción GxE viene dado por el valor y signo de su coordenada IACP, por contra, en el análisis de Componentes Principales no resulta tan obvio determinar la cuantía del aporte de cada ambiente a la Interacción GxE.

De esta forma nos limitaremos en este Capítulo a exponer brevemente los resultados de los ACP's y dirigir la atención del lector al estudio del análisis AMMI.

8.1.3. ACP de Ambientes. Datos originales de rendimiento

Los dos primeros componentes explican el 77.41 % y 11.68 % respectivamente. Las coordenadas genotípicas para estos dos primeros componentes han sido representadas en la Fig. 16.

Se puede observar en la Fig. 16 como se sigue el mismo patrón que en el caso del ACP de genotipos (caso 8.1). El primer componente ordena los genotipos según su rendimiento medio, mientras que del segundo componente no podemos extraer ninguna información clara relacionada con estabilidad.

Figura 16. ACP datos sin transformar (genotipos)

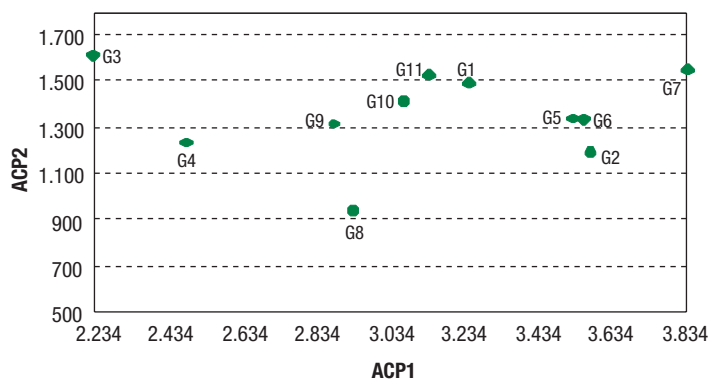
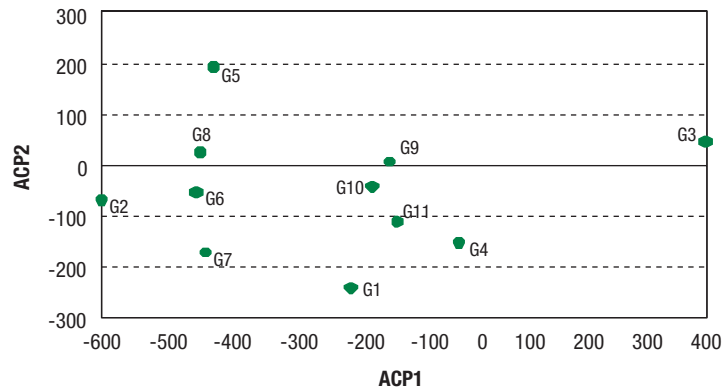


Figura 17. ACP datos transformados (genotipos)

8.1.4. Datos transformados

En la Figura 17 se representan las coordenadas genotípicas de los dos primeros componentes principales, el primero extrae un 68.24 %, y el segundo un 10.73 % de la variación total.

En esta gráfica podemos distinguir claramente tres grupos compuestos por los siguientes genotipos:

Grupo I: 2, 5, 6, 7 y 8.

Grupo II: 1, 4, 9, 10 y 11.

Grupo III: 3.

Se supone que como hemos partido de los datos transformados estos grupos obtenidos deben estar relacionados de alguna manera con la Interacción **GxE**. Este aspecto será dilucidado en el Capítulo siguiente donde obtenemos mas información, respecto a estabilidad, sobre los grupos formados. Recordamos al lector lo comentado en el

apartado 8.2 de este Capítulo, se tratará con más profundidad la relación existente entre estos grupos formados y la Interacción GxE en el análisis AMMI (Capítulo 9), dada la imposibilidad de poder realizarla, con cierta sencillez, con el ACP. Con el ACP sólo hemos sido capaces de determinar la existencia de tres grupos bien diferenciados, pero no podemos enlazar cada grupo con un determinado grado de estabilidad.

Según Freeman y Dowker (1973), cuando en el ACP de genotipos partimos de los datos originales de rendimiento, y el primer componente principal explica un elevado porcentaje de la varianza y además éste se corresponde con la ordenación de rendimientos medios, el segundo componente y los sucesivos representan la interacción **GxE**. Pero este no ha sido nuestro resultado, lo que demuestra, una vez más, que no caben generalizaciones en la interpretación de los análisis de este tipo.

REFERENCIAS CITADAS

- CUADRAS, C.M. 1981: Métodos de análisis multivariante. Ed. por Eunibar. Barcelona.
- FREEMAN, G.H. and B.D. DOWKER. 1973: The analysis of variation between and within genotypes and environments. *Heredity* 30, 97-109.
- KENDALL, M.G. and A. STUART. 1961: The advanced of statistics. Vol. 2, 3 rd ed. Griffin and Co. Ltd. London.

OTRAS REFERENCIAS ACONSEJADAS

- KEMPTON, R.A. 1984: The use of biplots in interpreting variety by environment interactions. *J. Agric. Sci., Camb.* 103, 123-135.
- MUNGOMERY, V.E., R. SHORTER and D.E. BYTH, 1974: Genotype x environment interactions and environmental adaptation. I Pattern analysis - Application to soya bean populations. *Aust. J. Agric. Res.* 25, 59-72.

9

“ADDITIVE MAIN EFFECTS AND MULTIPLICATIVE INTERACTION”: A.M.M.I.

9. “ADDITIVE MAIN EFFECTS AND MULTIPLICATIVE INTERACTION”: A.M.M.I.

Recuérdese que los ensayos de comparación de variedades se realizan con un gran número de variedades sembradas en distintos ambientes, y dispuestas en varias repeticiones. Recuérdese, asimismo, que la suma total de cuadrados para los datos de rendimiento puede separarse generalmente en tres partes:

- Efecto principal genotípico.
- Efecto principal ambiental.
- Interacción genotipo x ambiente.

Por definición, los efectos principales son aditivos, siendo la interacción no-aditiva. Si la interacción se expresa en términos individuales de genotipo y ambiente, la partición aditiva es imposible. La alternativa más simple para el análisis es una suma de términos multiplicativos.

El AMMI empieza con un análisis de la varianza (ANOVA) para calcular los efectos aditivos de genotipo y ambiente, y aplica a continuación un Análisis de Componentes Principales (ACP) para analizar los efectos no aditivos de la interacción. Como matriz de entrada para el ACP no se toman los datos originales de rendimiento, sino los datos transformados ($Y_{ij}^* = Y_{ij} - Y_{i.} - Y_{.j} + Y_{..}$) ($Y_{i.}$, $Y_{.j}$, $Y_{..}$ son, como siempre, medias marginales) que representan indudablemente a la interacción, realizándose un ACP (extrayéndose los componentes principales de la matriz de covarianzas y no de la de correlaciones) para genotipos y otro para ambientes, obteniéndose así las coordenadas ambientales y genotípicas respectivamente.

El modelo es:

$$Y_{ge} = \mu + \alpha_g + \beta_e + \sum \lambda_n \tau_{gn} \delta_{en} + \theta_{ge} \quad (51)$$

donde: Y_{ge} es el rendimiento del genotipo **g** en el ambiente **e**, μ es la media general, α es la desviación media genotípica, β es la desviación media ambiental, λ_n es el autovalor del eje n del análisis de Componentes Principales, τ_{gn} y δ_{en} son los vectores propios unitarios genotípicos y ambientales, respectivamente, asociados a λ_n .

Se ha seleccionado la siguiente escala para el cálculo de las coordenadas del ACP: las coordenadas genotípicas y ambientales se expresan como el producto de la raíz del autovalor λ_n por su vector propio asociado. Multiplicando las

coordenadas ambientales ($\sqrt{\lambda_n} \delta_{en}$) por las genotípicas ($\sqrt{\lambda_n} \tau_{gn}$) obtenemos directamente la estimación de la interacción ($\lambda_n \tau_{gn} \delta_{en}$), de un determinado genotipo en cualquier ambiente. El conjunto de estos valores constituyen así, la matriz de datos de la interacción estimada por el método AMMI.

Los grados de libertad para cada componente del ACP retenido se calcula por el método de Gollob (1968):

$$df = G + E - 1 - 2n. \quad (52)$$

Cálculo del autovalor.

1º) Los autovalores de una matriz **Y** descompuesta por **SVD** (Singular Value Decomposition) pueden obtenerse a partir de la siguiente expresión (Gabriel, 1971):

$$\|Y\|^2 = \lambda^2_1 + \dots + \lambda^2_e \quad (53)$$

haciendo la matriz **Y** igual a la matriz de datos de la interacción, la expresión anterior nos dice que la suma de cuadrados de la interacción la podemos descomponer en **e** sumas de cuadrados, representando cada una de ellas la suma de cuadrados de cada uno de los ejes del ACP.

2º) Por tanto, si obtenemos el porcentaje de variación que retiene el primer eje del ACP (ver Capítulo 8) podremos calcular la suma de cuadrados del primer eje, sin más que multiplicarlo por la suma de cuadrados de la interacción **GxE** que obtenemos del ANOVA y dividir el resultado entre el número de repeticiones del ensayo.

3º) Según [53] la suma de cuadrados del primer eje es igual a λ^2_1 ; podemos pues calcular la raíz del primer autovalor del AMMI, que será igual a la suma de cuadrados del primer eje del ACP elevado a 1/4. Veamos cómo:

Siendo:

SS_{ge} = suma de cuadrados interacción **GxE**

k_i % el porcentaje de variabilidad retenido por el eje **i**, y r el número de repeticiones

si $SS_{eje i} = \lambda^2_1$, según [53], y

$$SS_{eje i} = [SS_{ge} (k_i/100)]/r, \text{ entonces} \quad (54)$$

$$\sqrt{\lambda} = (SS_{eje i})^{1/4} \quad (55)$$



9.1. La validez Predictiva

9.1.1. **La validez predictiva** se calcula por medio del procedimiento de Confirmación Cruzada (C.C.) descrito por Gauch y Zobel (1988). La matriz de datos sin transformar se divide en dos submatrices; una de ellas representa los datos para el **modelo** y la otra los datos para la **confirmación**. Para cada tratamiento de nuestro ejemplo (por ejemplo, para cada combinación genotipo y ambiente) se han seleccionado dos repeticiones al azar para el modelo y al que aplicaremos el análisis AMMI, habiéndose reservado la tercera como conjunto de observaciones de confirmación (así, tenemos $11 \times 10 \times 1 = 110$ observaciones para confirmación).

El método de Confirmación Cruzada se utiliza para seleccionar el número óptimo de ejes retenidos en el ACP. Usando este método, cada observación se estima mediante el modelo utilizado para el AMMI con dos repeticiones sin necesidad de tener que ampliar nuestra matriz de datos.

9.1.2. Posteriormente se ajustan varios modelos: **AMMI0**, el cual sólo estima los efectos aditivos de genotipo y ambiente sin retener o incluir ningún eje del ACP en el modelo; **AMMI1**, que combina los efectos aditivos del anterior con el efecto interacción estimado por el primer eje del ACP; y así sucesivamente hasta completar el total del número de ejes que será el menor de **g** y **e**. El modelo completo (**AMMI10**) se denomina **CELL MEANS** según la nomenclatura más utilizada.

9.1.3. Los valores estimados para cada modelo se comparan con los datos de confirmación (véase parágrafo 9.1.1). Por ejemplo, el valor estimado por cualquier de estos modelos para el genotipo 1 en el ambiente 1 se compara con el dato correspondiente de confirmación. La diferencia entre el valor estimado y este último se eleva al cuadrado, se le suman todas las diferencias restantes de todos los genotipos en todos los ambientes (véase 9.1.3.1). Esta suma de cuadrados se divide por el número de observaciones de confirmación (véase 9.1.3.2) y se calcula su raíz cuadrada (véase 9.1.3.3) dando lo que se denomina **RMS PD (“Root Mean Square Predictive Difference”)** según la nomenclatura usual de este método. Un peque-

ño valor del RMS PD indica un buen ajuste del modelo. Así, sea $X_{1,1}$ el valor estimado para el genotipo 1 en el ambiente 1 e $Y_{1,1}$ el dato reservado de confirmación para la misma combinación de genotipo-ambiente. Como se ha explicado, la forma de proceder para el cálculo del estadístico RMS PD es la siguiente:

9.1.3.1.)

$$(X_{1,1} - Y_{1,1})^2 + (X_{1,2} - Y_{1,2})^2 + \dots + (X_{g,e} - Y_{g,e})^2 = \text{SS PD}$$

9.1.3.2.)

$$\text{SS PD}/n^{\circ} \text{ datos confirmación} = \text{MS PD}$$

9.1.3.3.)

$$\sqrt{\text{MS PD}} = \text{RMS PD}$$

La diferencia entre el **(RMS PD)²** y el error obtenido del modelo ajustado **(RMS FE)²**, es decir, la varianza de los datos de confirmación, se utiliza para la estimación del error predictivo **(RMS PE)²** (siempre siguiendo la nomenclatura usual de este método).

Se estima que, en general, unos 10 modelos diferentes elegidos al azar de la matriz de datos son suficientes para la realización de la confirmación cruzada y comprobar si el análisis AMMI es más preciso que un modelo de simples efectos aditivos.

9.1.4. El análisis AMMI se completa con la construcción de un **“biplot”**. En abscisas se representan los efectos principales (medias genotípicas y ambientales) y en ordenadas la coordenadas en el primer eje del ACP.

9.2. Aplicación práctica

9.2.1. Procedimientos

Los resultados obtenidos para nuestros datos del ejemplo se exponen a continuación. Se han utilizado dos procedimientos:

a) Para la obtención de las coordenadas genotípicas y ambientales se ha ejecutado una corriente de control para el paquete estadístico SAS (ver Apéndice).

b) En el caso de que no se disponga de la corriente de control anterior otra forma válida es la siguiente (véanse en el Apéndice los programas necesarios):

1) Se ha diseñado un programa en lenguaje BASIC que a partir de la matriz de datos nos genera diez modelos totalmente al azar.

2) Con el paquete estadístico NTSYS se calculan los vectores propios genotípicos y ambientales (de **INT**) sin el peso de la raíz cuadrada del autovalor *para cada modelo*.

3) Con un programa en lenguaje BASIC (véase en el Apéndice) se calculan el ANOVA, los autovalores y las coordenadas genotípicas y ambientales. Previamente habrá habido que introducir, como ficheros de texto, los vectores propios obtenidos por el NTSYS.

Hasta aquí hemos completado los resultados obtenidos en el apartado a), pero el análisis AMMI va más allá: debemos calcular cual es el modelo más adecuado, es decir, debemos calcular el RMSPD de los modelos al azar generados en el apartado 1.

4) Con otro programa en BASIC calculamos los RMSPD para el AMMI0, AMMI1, AMMI2, ..., AMMI10 de cada modelo anterior.

5) Calculamos la media de los valores del RMSPD para cada AMMI (AMMI0, AMMI1, ..., AMMI10) de los diez modelos generados al azar.

El valor más pequeño del RMSPD nos indicará el modelo AMMI más preciso.

9.2.2. Resultados

A continuación comentaremos los resultados obtenidos.

La prueba **F** para el análisis AMMI recomienda incluir los dos primeros ejes en el modelo, ya que el tercero resulta no significativo al 5% (Tabla 15).

El ANOVA del AMMI muestra que ambientes, genotipos e interacción **GxE** son altamente significativos ($P < 0.0001$) representando el 71.37, 9.45 y el 8.67 % de la S.C. de la combinación total de tratamientos, respectivamente.

Como hemos comentado anteriormente según el criterio **postdictivo** del AMMI medido por el ANOVA se incluirían en el modelo los dos primeros ejes de la interacción (recuérdase que el tercero resulta significativo al 8%, Tabla 15). Sin embargo, desde un punto de vista **predictivo**, medido por el RMS PD, se selecciona el AMMI1 ya que tiene el valor más bajo para este estadístico (453.08 kg/ha) (Tabla 17), sus coordenadas genotípicas y ambientales se muestran en la Tabla 16. Recordemos que este modelo es el que incluye sólo el primer eje de la interacción.

Con este modelo bastan 37 g.l.: 9 para ambientes más 10 para genotipos y 18 para el

Tabla 15. Efectos aditivos e interacción del ANOVA incluyendo los 3 primeros ejes del ACP de la interacción

F. de V.	GL	SC	R ²	F	Pr>F
Total	329	22751.6			
Ambiente	9	9412.6	41.37	19.48	.0001
Rep./Amb.	20	1073.7	4.7	3.92	.0001
Genotipo	10	5571.2	24.48	12.68	.0001
Gen.xAmb.	90	3953.5	17.37	3.21	.0001
Int. ACP1	18	2697.8	68.24	10.93	.0001
Int. ACP2	16	424.3	10.73	1.93	.0200
Int. ACP3	14	325.5	8.23	1.69	.0810
Resd.(ACP1)	72	1255.7	31.76	1.27	.1002
Resd.(1-2)	42	505.8	21.03	0.86	.6980
Error	200	2740.0	12.04		

- **Resd.(ACP1):** Componente residual cuando solo introducimos en el modelo el primer eje.
- **Resd.(1-2):** Componente residual cuando se han introducido en el modelo los dos primeros ejes.
- **% Varianza explicada por los tres primeros componentes:** Int.ACP1=68.24%; Int.ACP2 = 10.73 %; Int.ACP3 = 8.23 %.

Tabla 16. Coordenadas genotípicas y ambientales del primer componente del AMMI.

GENOTIPO	AMBIENTE	RENDIMIENTO	Int. ACP1
1		3699.38	-2.557
2		3825.07	23.068
3		2877.50	-38.996
4		2920.35	-13.547
5		3896.25	11.667
6		3907.16	13.422
7		4241.11	12.456
8		3193.40	12.945
9		3320.56	-6.482
10		3517.99	-4.749
11		3617.71	-7.228
	1	3607.44	-13.513
	2	3686.88	-12.896
	3	3304.04	8.480
	4	3433.34	-22.456
	5	2776.58	1.194
	6	3561.86	-9.681
	7	3784.10	-4.811
	8	4422.35	20.237
	9	4259.10	38.932
	10	2633.84	-5.486

primer eje ACP de la interacción. Es 1.4 (=53/37; ver Tabla 17) veces más “parsimonioso” (esto es, necesita de pocos g.l.) que el AMMI2, que es el que hubiera habido que seleccionar desde el punto de vista del criterio **postdictivo** al recomendar la prueba F incluir los dos primeros ejes en el modelo, como ya se ha comentado (Tabla 15).

En la Tabla 15 vemos como el modelo AMMI1 captura el 77.71 % de la S.C. del total del ensayo $[(9412.6+5571.2+2697.8)/22751.6]$ como valor *predictivo*. Por tanto, el procedimiento Confirmación Cruzada (C.C.) (Tabla 17) indica que el 22.29 % restante de la S.C. total (100-77.71) se corresponde con un residuo de valor **no predictivo**.

9.2.3. Estimación de las repeticiones necesarias

Un cálculo interesante es determinar el número de repeticiones necesarias, dentro de ambientes, para igualar la precisión del modelo

AMMI1. Para ello, empleando la nomenclatura de Crossa y col. (1990), la cantidad de mayor interés es el MSE(modelo) (error medio cuadrático del modelo) que puede estimarse como la diferencia entre el MSE(modelo - confirmación) y la varianza de los datos de confirmación.

El MSE (modelo - confirmación) es el cuadrado del RMS PD para el AMMI1 (453.08 kg/ha, Tabla 17). La varianza de los datos de confirmación es simplemente el cuadrado medio del error del ANOVA del AMMI (173,350 ignorando repeticiones dentro de ambientes, es decir, sin tener en cuenta, al realizar el ANOVA, la suma de cuadrados del término Rep./Amb. que quedaría incluida en la suma de cuadrados del error), el cual, expresado en términos de su raíz cuadrada, es 416.35 kg/ha (RMS FE, Tabla 18).

Así, para el AMMI1, el MSE(modelo) es estimado como $[(453.08)^2-(416.35)^2]$, el cual expre-

Tabla 17. RMS PD (Kg ha^{-1}) para siete modelos basados en datos de rendimiento.

MODELO	G.L.	RMS PD
AMMI0	19	499.14
AMMI1	37	453.08
AMMI2	53	476.68
AMMI3	67	494.04
AMMI4	79	500.56
MODELO COMPLETO	329	509.14

Tabla 18. Estimación de parámetros del AMMI para su evaluación predictiva para rendimiento medio de 11 genotipos en 10 ambientes.

PARÁMETROS	ESTIMADO
MODELO Y ERROR	
Modelo	77.71 %
Variación debida al azar	22.29 %
EVALUACIÓN DEL MODELO	
Modelo postdictivo	2 IACP
G. L. postdictivo	53
Modelo predictivo	AMMI1
G. L. predictivo	37
Parsimonia	1.4
PREDICCIÓN	
Observaciones modelo ajustado	220
Datos de confirmación	110
RMS PD	453.08 kg/ha
RMS FE	416.35 kg/ha
RMS PE	178.70 kg/ha
RMS PE relativo a la media total	5.04 %
GANANCIA CON EL AMMI	
R_{AMMI}	5.43
Beneficio en repeticiones (RB)	3.43
Observaciones adicionales (AO)	377
Factor de Ganancia (GF)	2.72

sado en términos de su raíz cuadrada es 178.70 kg/ha (RMS PE en Tabla 19). Esto representa un error de tan sólo el 5% (178.70/3546) relativo a la media total del ensayo.

La varianza de los datos de confirmación es el cuadrado medio del error de una sola repetición, mientras el MSE(modelo) está basado sobre R_m repeticiones (donde R_m =número de repeticiones utilizados para el modelo).

De acuerdo con Gauch y Zobel (1988), el cálculo del número de repeticiones (R_L) que hacen que la predicción basada sobre el modelo completo tenga igual precisión que el mejor modelo predictivo basado en R_m repeticiones, se realiza de una forma aproximada como

$$R_L = [\text{Var}(\text{datos confirmación})/\text{MSE}(\text{modelo})] = [\text{RMS FE}/\text{RMS PE}]^2 .$$

El número aproximado de repeticiones necesarios para que el modelo completo iguale al AMMI1, por tanto, es $R_L = R_{AMMI} = (416.35/178.70)^2 = 5.43$.

Como el mejor modelo predictivo está basado sobre $R_m=2$ repeticiones, el factor de ganancia aproximado se puede definir como $FG = R_L / R_m = 5.43/2 = 2.72$ (Tabla 18).

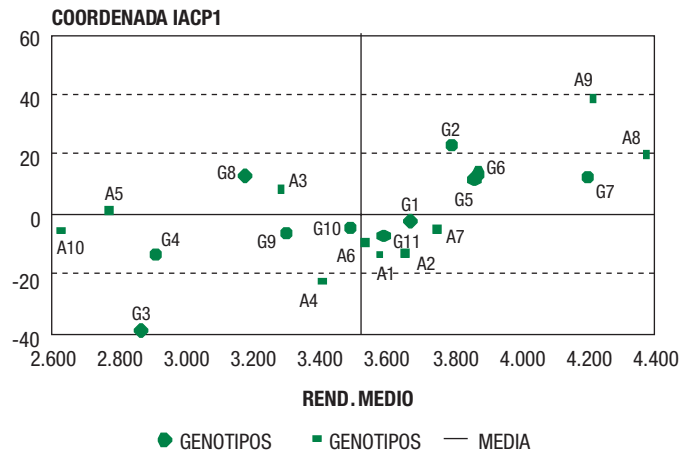
De este modo el modelo AMMI1 basado sobre dos repeticiones es, en el caso de los datos aquí analizados, tan preciso como el modelo completo basado en 5 repeticiones. Puede decirse, pues, que los valores ajustados por AMMI1 son más precisos que los del modelo completo para estimar el rendimiento.

9.2.4. El “biplot”

Una forma de plasmar visualmente los resultados del AMMI es realizando la gráfica o “**biplot**” de rendimiento medio en función de las coordenadas del primer eje del ACP (IACP1) de genotipos y de ambientes (Fig. 18). Los genotipos (o ambientes) con coordenadas elevadas para IACP1 (tanto positivas como negativas) tienen una aportación mayor a la interacción **GxE**, mientras que los genotipos (o ambientes) con IACP1 próximo a cero aportan mínimamente a la interacción, resultando éstos más estables.

Como puntualizaron Zobel y col. (1988), el rendimiento esperado, según el modelo AMMI1, para cualquier combinación genotipo-ambiente puede calcularse a partir de la Fig. 18. Veamos un ejemplo de como realizar esta estimación y cómo el modelo AMMI1 resulta más ajustado que el AMMI0. Para ello necesitamos los datos siguientes:

Figura 18. Análisis AMMI



- Rendimiento medio del genotipo 7 = 4241.11 kg/ha (ver Tabla 16).
- Rendimiento medio de todos los genotipos en el ambiente 9 = 4259.1 kg/ha (ver Tabla 16).
- Rendimiento medio del ensayo = 3,546.95 kg/ha.

De la matriz de datos que podemos ver en el Apéndice recogemos el siguiente extracto:

AMBIENTE	GENOTIPO	REPETICIÓN	RENDIMIENTO (kg/ha)
9	7	1	6,016.70
9	7	2	5,250.00
9	7	3	5,070.80

El rendimiento estimado por el método AMMI0 sería:

$$\begin{aligned} \text{Efectos aditivos} &= \\ &= 4,241.11 + 4,259.1 - 3,546.95 = 4,953.26 \text{ kg/ha.} \end{aligned}$$

El rendimiento estimado por el método AMMI1 sería:

$$\begin{aligned} \text{Efectos aditivos} + \text{efectos multiplicativos} \\ \text{extraídos del primer eje del ACP (ver Tabla 16)} &= \\ &= 4,953.26 + (12.45 * 38.93) = 5,437.93 \text{ kg/ha.} \end{aligned}$$

Como vemos todos estos cálculos también podemos realizarlos de una forma aproximada con los datos que nos ofrece el biplot de la Figura 18, pues disponemos de los valores medios tanto de genotipos como de ambientes para el cálculo de los efectos aditivos y de las coordenadas del IACP 1 para el cálculo de los efectos multiplicativos.

Para estimar que modelo AMMI1 se ha ajustado más a los datos de partida que el modelo AMMI0 debemos tener en cuenta que, como se ha comentado anteriormente el modelo aditivo AMMI0 parte simplemente de los efectos aditivos, es decir, de la suma de las medias genotípica y ambiental menos la media total. La interacción se estima multiplicando las coordenadas genotípicas por las ambientales del ACP1. Estos dos resultados se suman para producir el valor esperado del modelo AMMI1.

De una forma práctica se procede de la siguiente forma: del extracto anterior deducimos que el rendimiento medio del genotipo 7 en el ambiente 9 es de 5,445.8 kg/ha, a cuyo valor se ajusta más la estimación del modelo AMMI1 (5,437.93 kg/ha) que la del modelo AMMI0 (4,953.26 kg/ha). Existe una diferencia de 484.67 kg/ha entre ambos métodos, un 13.6 % respecto de la media del ensayo, dada por los efectos multiplicativos (efectos de la Interacción Gx E) extraídos del primer eje del ACP.

9.2.5. Agrupaciones en el biplot

Genotipos y ambientes con igual signo en IACP1 producen una interacción positiva, en tanto que si tienen distinto signo la producen negativa.

Se observan tres grupos de **genotipos** en la Fig. 18:

- 1) **Grupo 1**: incluye los genotipos 2, 5, 6, 7 y 8. Los cinco se caracterizan por tener valores intermedios de signo positivo en sus coordenadas IACP1. Para estos genotipos, el modelo



AMMI1 predice rendimientos genotípicos iguales a los dados por el modelo AMMI0 en ambientes con coordenadas IACP1 próximas a cero, menores que los dados por el AMMI0 en ambientes con coordenadas IACP1 negativas y mayores en ambientes con coordenadas IACP1 positivas.

2) **Grupo 2:** constituido por los genotipos 1, 4, 9, 10 y 11 que tienen unos valores intermedios de signo negativo en sus coordenadas IACP1.

3) **Grupo 3:** constituido por el genotipo 3 que tiene las coordenadas IACP1 de signo negativo más elevadas.

En cuanto a los **ambientes**, no resulta tan clara la posibilidad de establecer una serie de grupos como hicimos con los genotipos. Aún así podemos decir que muestran una mayor variabilidad en el efecto principal (media) que en el efecto de la interacción (IACP1).

9.2.6. Comentarios finales

Como podemos observar, hemos obtenido unos resultados muy similares a los obtenidos por el análisis de Componentes Principales en cuanto a la ordenación de los genotipos y ambientes. Sin embargo, *el análisis AMMI va más allá estimándonos el aporte de cada genotipo y ambiente al efecto de la interacción.*

En la Tabla 18 resumimos los resultados dados por el AMMI para nuestro ensayo. En ella podemos ver como el modelo AMMI1 extrae el 77.71 % de la S.C. del total del ensayo, con 37 g.l. (10 de genotipos, 9 de ambientes y 18 del IACP1), y 22.29 % como variación no predictiva debida al azar, con 72 g.l.

En dicha Tabla también podemos ver como el número aproximado de repeticiones (R_{AMMI}) necesarios para que el modelo completo iguale la precisión del AMMI1 es de 5.43. Es decir, el modelo AMMI1 basado en dos repeticiones es tan preciso como el modelo completo con 5.43 repeticiones, de esta forma existe un factor de ganancia teórico en precisión de 2.72. El beneficio en repeticiones es de 3.43 o lo que es lo mismo un beneficio de 377 observaciones.

9.3. El “ruido” en el Análisis AMMI

Un aspecto importante de los ensayos multi-locales en Mejora es estimar los rendimientos

genotípicos con la máxima exactitud. Un obstáculo en ésta estimación son el “ruido” o “noise” y el error en los datos de campo.

De forma práctica, el ruido en el análisis AMMI se cuantifica por la S.C. residual después de ajustar el mejor modelo predictivo, y el error se estima por las diferencias entre unidades individuales experimentales (repeticiones) dentro del mismo tratamiento (combinaciones **GxE**).

En la Tabla 15, el término “Residual (IACP1)” se correspondería con el ruido, el cual es extraíble por el análisis AMMI de la S.C. de la interacción **GxE**. En la misma Tabla 15, el término “error” nos viene dado por el error de variación extraído del diseño de bloques al azar. Esto no implica, sin embargo, que éstas u otras fuentes de variación estén libres de error y ruido.

Desde un punto de vista práctico, es difícil una distinción precisa entre “ruido” y “error”. La desigualdad en el riego, la aplicación de productos agroquímicos, los daños por insectos y animales y las enfermedades contribuyen a la existencia de unas variaciones indeseables de tipo “ruido”, tanto como las diferencias en suelo y las imprecisiones en las operaciones de campo. Estadísticamente, el grado en que éstos pueden ser cuantificados por medio de repeticiones, bloques incompletos, error, residual, u otros términos en un análisis de varianza dependen del diseño experimental, y del análisis estadístico realizado.

Existen tres tipos básicos de estrategias para controlar el error e incrementar la precisión de los rendimientos genotípicos estimados:

- 1) Partición del error.
- 2) Partición de la variación genotípica, y
- 3) partición de la variación **GxE**.

El primero utiliza un diseño de bloques incompletos en forma de “lattices” o “lattices” generalizados, el cual sustrae la variación entre bloques del error experimental. El segundo incluye un método espacial (análisis “neighbor”, análisis en cuadros latinos, análisis de parcelas vecinas) para ajustar los patrones de variación de la fertilidad del suelo.

Ambas estrategias se basan en el diseño experimental. Por el contrario, el modelo AMMI extrae el término residual o ruido de la interacción **GxE**.

Dado que estos tres tipos de estrategias de control del error se aplican a fuentes de variación diferentes y ortogonales (error, genotipos, e interacción) se pueden utilizar independiente o simultáneamente.

Una ventaja significativa del análisis AMMI es que las estimas en un ambiente dado se realizan utilizando también la información de los otros ambientes del ensayo, mientras que en las otras estrategias de estudio de interacción GxE, un ambiente no tiene influencia en las estimas realizadas en cualquier otro ambiente del ensayo. Gauch (1990) sugiere que ésta es una de las razones por la cual las estimas realizadas con el AMMI son más precisas predictivamente que las realizadas por medio del modelo de tratamientos medios, es decir, cuando la estima del rendimiento se realiza como la media de las **R** repeticiones de cada genotipo en cada ambiente. El modelo AMMI considera todas las observaciones **GxE_iR**, mientras que el modelo de tratamientos medios se enfoca exclusivamente sobre las **R** repeticiones y descarta las **R(GxE-1)** observaciones restantes.

Desde el punto de vista biológico, un ambiente perjudicado por algún tipo de estrés (por ejemplo, niveles altos de sodio en el suelo) entre otros ambientes sin éste problema podría ser anulado

por el efecto conjunto de los restantes ambientes. Según Crossa y col. (1991) la mayoría de las causas biológicas asociadas a la interacción **GxE** podrían ser relegadas al término residual del análisis AMMI.

En resumen, como ya se comentó en la Introducción y como podemos deducir de lo expuesto en este capítulo, el análisis AMMI tiene las siguientes ventajas sobre los demás métodos de estudio de la interacción GxE:

1) Se trata de un método potente de diagnóstico, ya que nos permite la posibilidad de poder estudiar submodelos representativos de los datos totales del ensayo.

2) Nos aporta una gran cantidad de información sobre la naturaleza de la interacción GxE, es decir, nos permite clasificar tanto los genotipos como los ambientes según sus patrones de comportamiento.

3) Un aspecto que ha quedado muy claro en este capítulo es la mejora de la precisión en las estimas del rendimiento. Se ha demostrado que se obtiene la misma precisión en la estima de rendimiento con el análisis AMMI que con otros tipos de análisis con un número de repeticiones de 2 a 5 veces mayor.

REFERENCIAS CITADAS

• BESAG, J. and R. KEMPTON. 1986: Statistical analysis of field experiments using neighboring plots. *Biometrics* 42, 231-251.

• CROSSA, J., H.G. GAUCH, Jr., and R.W. ZOBEL. 1990: Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. *Crop Sci.* 30, 493-500.

• CROSSA, J., P.N. FOX, W.H. PFEIFFER, S. RAJARAM and H.G. GAUCH, Jr. 1991: AMMI adjustment for statistical analysis of an international wheat yield trial. *Theor. Appl. Genet.* 81, 27-37.

• GABRIEL, K.R. 1971: The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453-467.

• GAUCH, H.G.Jr. 1990: Full and reduced models for yield trials. *Theor Appl. Genet.* 80, 153-160.

• GAUCH, H.G.Jr., and R.W. ZOBEL. 1988: Predictive and postdictive success of statistical analyses of yield trials. *Theor Appl. Genet.* 76, 1-10.

• GOLLOB, H.F. 1968: A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* 33, 73-115.

- PATTERSON, H. D., E. R. WILLIAMS, and E. A. HUNTER. 1978: Block Designs for Variety Trials. *J. Agric. Sci.* 90, 395-400.

- WILKINSON, G. N, S. R. ECKERT, T. W. HANCOCK and O. MAYO. 1983: Nearest Neighbour (N N) Analysis of Field Experiments. *J. R. Statist. Soc.B.* 45, 151-211.

- YATES, F. 1936: A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci.* 26, 424-455.

- ZOBEL, R.W., MADISON J. WRIGHT and HUGH G. GAUCH, Jr. 1988: Statistical analysis of a yield trial. *Agron. J.* 80, 388-393.

**OTRAS REFERENCIAS
ACONSEJADAS**

- BRADU, D. and K.R. GABRIEL. 1978: The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 20, 47-68.

- CROSSA, J., 1990: Statistical analyses of multilocation trials. *Adv. Agr.* 44, 55-85.

- DIGBY, P.G.N. and R.A. KEMPTON. 1987: *Multivariate analysis of ecological communities.* Chapman and Hall (Ed.)

- GAUCH, H.G. 1982: *Multivariate analysis in community ecology.* 1st Ed. Cambridge Univ. Press. London and New York.

- GAUCH, H.G. 1985: Mimeo 85-7. Cornell Univ. Ithaca, New York.

- GAUCH, H.G. 1988: Model selection and validation for yield trials with interactions. *Biometrics* 44, 705-715.

- GAUCH, H.G.Jr., and R.W. ZOBEL. 1989: Accuracy and selection succes in yield trial analyses. *Theor. Appl. Genet.* 77, 473-481.

- GAUCH. H. G, Jr. and R. W. ZOBEL. 1990: Imputing Missing Yield Trial Data. *Theor. Appl. Genet.* 79, 753-761.

- KEMPTON, R.A. 1984: The use of biplots in interpreting variety by environment interactions. *J. Agric. Sci., Camb.* 103, 123-135.

- KRZANOWSKI, W.J. 1983: Cross-validatory choise in principal components analysis-some sampling results. *J. Statist. Computation and simulation* 18, 299-314.

- WOLD, S. 1978: Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 20, 397-405.





10

ANÁLISIS DE GRUPOS

10. ANÁLISIS DE GRUPOS

El problema de la clasificación consiste básicamente en hallar una partición tal que satisfaga a algún criterio discriminante o de discriminación óptima. Dicho criterio se puede dar en forma de relación funcional que refleje los niveles deseables a elegir entre varias particiones o agrupamientos. A esta relación funcional se la llama **función objeto**.

El problema intrínseco de la clasificación es la elección del criterio de discriminación óptima o de la **función objeto**. Esta nos dice cuándo se ha obtenido una partición deseable.

Los procedimientos de clasificación jerárquicos se basan en que los dos individuos más próximos quedan automáticamente agrupados y tratados en adelante como un solo grupo. Así, el número de individuos se habrá reducido a n-1: un grupo de 2 individuos y n-2 grupos de un individuo cada uno. El proceso se continúa hasta que todos los individuos son agrupados en un grupo que los contiene a todos.

Se necesita por un lado una medida de homogeneidad interna del grupo (**medida de similitud**) y por otro, un criterio que dé la disparidad entre dos grupos distintos (**algoritmo de clasificación**). Y por último, una relación jerárquica como resultado de una clasificación, representada gráficamente por un **dendrograma** que es la representación gráfica de una distancia ultramétrica sobre un conjunto finito.

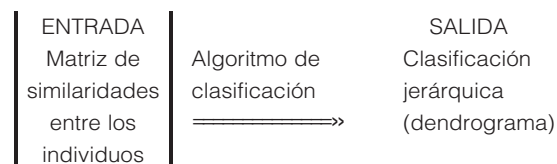
Cuando se trabaja en espacios métricos (eso sucede cuando las observaciones están caracterizadas por caracteres cuantitativos y en algunos casos cuando lo están por cualitativos) para medir la proximidad existente entre pares de elementos suele utilizarse una **distancia**, que es una medida particular de disimilaridad.

Hay que señalar, sin embargo, que el problema de determinar la distancia ultramétrica más próxima a una distancia definida (o a una disimilaridad) no está resuelto, existiendo un gran número de algoritmos que conducen, a partir de un cuadro de distancias o de disimilaridades, a un cuadro de distancias ultramétricas y en consecuencia a una jerarquización.

De la utilización de distintos algoritmos, también denominados **estrategias de agregación**, resultarán generalmente distintas clasifi-

caciones jerárquicas. Este aspecto constituye el punto débil más importante de la clasificación jerárquica.

El esquema de una clasificación jerárquica es el siguiente:



Este tipo de análisis se ha utilizado como una técnica alternativa para descomponer un gran número de genotipos, cuya interacción con el ambiente es significativa, en subgrupos de genotipos que presentan respuestas similares, sin ninguna restricción en cuanto al tipo de modelo de respuesta.

La dificultad principal que se encuentra para aplicar éste método es la de elegir una medida de similitud y un algoritmo de agrupamiento entre la amplia gama citada por los distintos autores que han tratado el tema, sin que exista un criterio objetivo que ayude a tal elección. Dada la naturaleza multivariante del Análisis de Grupos, el establecimiento de los diferentes grupos vendrá determinado tanto por el peso específico de la observación (genotipo), es decir, por el efecto genético (rendimiento medio genotípico), como por el perfil que presenta el genotipo en relación a las variables observadas (ambientes). Por tanto, resulta difícil determinar el grado en que cada uno de los factores citados contribuye al establecimiento de los diferentes grupos.

Ante tal situación y dado que desconocemos a priori la naturaleza de la interacción **GxE**, se ha utilizado aquí, en primer lugar, una medida de similitud convencional: **el cuadrado de la distancia euclídea entre genotipos**. Esta medida tiene en cuenta, simultáneamente, el efecto genético y el efecto de la interacción GxE.

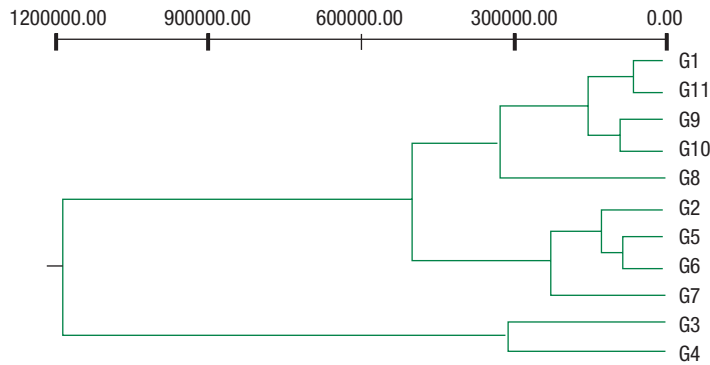
Si Y_{ij} es el valor observado del genotipo i -ésimo en el ambiente j -ésimo, la distancia euclídea entre dos genotipos (i, i') se define como:

$$d^2(i, i') = \sum_{j=1}^e (Y_{ij} - Y_{i'j})^2 \quad (56)$$

Como método aglomerativo se ha utilizado primeramente el **método de ligamiento medio**



Figura 19 Análisis de grupos Genotipos (UPGMA-DIST. EUCL.)



entre grupos o método **UPGMA**. Este algoritmo define la distancia entre dos grupos como la media de las distancias entre todos los pares de casos posibles, donde cada uno de los miembros de cada par pertenece a cada uno de los grupos.

En base a esta distancia y a este algoritmo se ha procedido a efectuar la clasificación de genotipos y ambientes, obteniéndose como resultado los dendrogramas de las Figuras 19 y 20. Las representaciones gráficas obtenidas a partir de los procedimientos de clasificación terminan necesariamente con la realización de un corte en el dendrograma. En ellos, el nivel al que se ha efectuado el corte viene dado por el momento del ciclo en el cual se produce un salto cuantitativo. A pesar de que dicho corte puede parecer objetivo, en la mayor parte de los casos se hace por el investigador de forma subjetiva. No podemos olvidar de ningún modo que al realizar el corte en el árbol de relaciones (esto es, en el dendrograma), entra en juego el criterio del investigador para determinar la altura a la cual

debe efectuarse el corte. Existe, por tanto, una clara interacción entre el procedimiento estadístico y el conocimiento que se posee sobre los datos, por lo que incluso pueden estar justificados truncamientos a distinto nivel.

Para la realización de los análisis de grupos que vamos a comentar a continuación se ha utilizado el paquete estadístico de métodos multivariantes NTSYS:

1º) A partir de la matriz rectangular de medias 10 x 11 o 11 x 10 dependiendo de si queremos obtener el dendrograma de genotipos o de ambientes, respectivamente. En primer lugar se elige el tipo de distancia en el subprograma "Similarity Interval". En este caso se ha escogido "DISTSQ" (distancia euclídea al cuadrado), dando como resultado la matriz de distancias (la matriz de partida se da en el Apéndice).

2º) Una vez obtenida la matriz de distancias, el análisis de grupos sigue con la elección del

Figura 20 Análisis de grupos Ambientes (UPGMA-DIST. EUCL.)

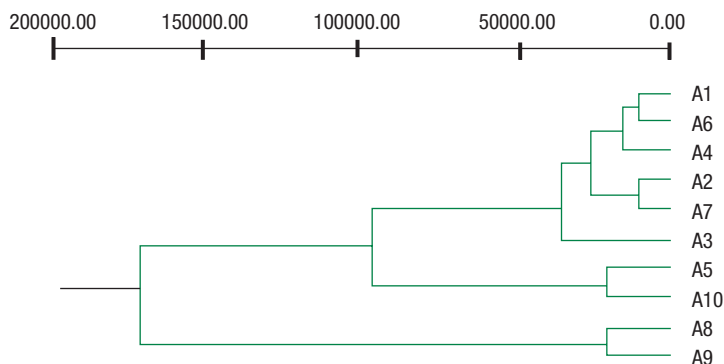
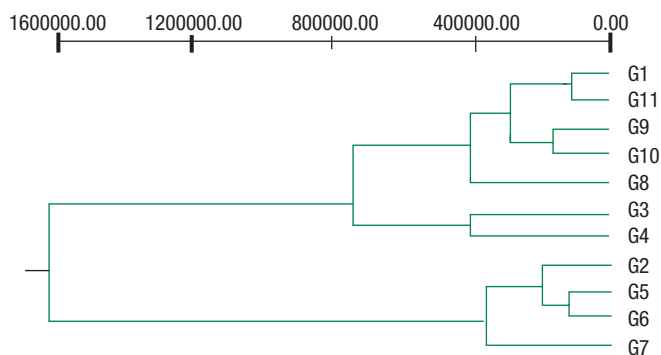


Figura 21 Análisis de grupos Genotipos (WPGMA-DIST. EUCL.)



Σ

α

Π

Ω

\times

\pm

ϑ

algoritmo de fusión. Para ello, seleccionamos el subprograma "SAHN Clustering" que nos da a elegir entre varios algoritmos de fusión (UPGMA, WPGMA, SINGL, COMPL, ...) dando como resultados los árboles de relaciones o dendrogramas de las Figuras 19 a la 26.

Como se puede observar, en las Figuras 19 y 20 los grupos formados son:

GENOTIPOS:

- Grupo 1: 2, 5, 6 y 7 genotipos de mayor rendimiento;
- Grupo 2: 1, 8, 9, 10 y 11 genotipos de rendimientos intermedios y
- Grupo 3: 3 y 4 genotipos de bajo rendimiento

AMBIENTES:

- Grupo 1: 8 y 9 son en los que se han obtenido mayores rendimientos medios;
- Grupo 2: 1, 2, 3, 4, 6 y 7 donde se obtienen rendimientos intermedios y

Grupo 3: 5 y 10 donde se han obtenido los rendimientos más bajos,

Por tanto, el efecto **GxE** queda totalmente absorbido por el efecto genético, de modo que los grupos que se obtienen se diferencian única y exclusivamente en base a sus rendimientos medios.

De la Figura 21 a la 26 se muestran los dendrogramas de los análisis de grupos realizados con la misma medida de similaridad pero con distintos algoritmos de fusión WPGMA, SINGL y COMPL. En ellas observamos cómo los resultados son muy similares a los comentados anteriormente.

Puesto que estos resultados no aportan nada nuevo a lo ya conocido para los datos del ejemplo, resulta interesante incidir más sobre la naturaleza de la interacción **GxE**. Ello exige buscar una medida de similaridad que, además de

Figura 22 Análisis de grupos Ambientes (WPGMA-DIST. EUCL.)

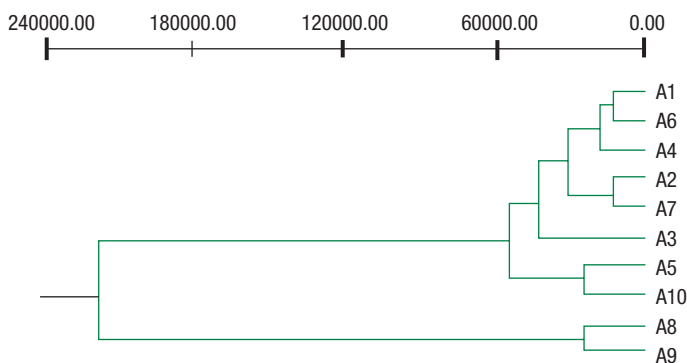
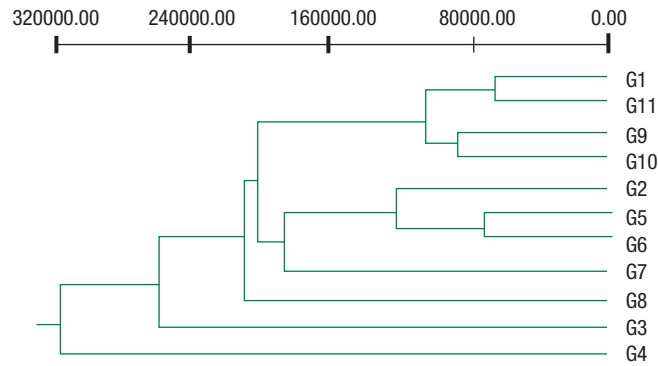


Figura 23 Análisis de grupos Genotipos (SINGL-DIST. EUCL.)



hacer más hincapié en el perfil de los genotipos, esto es, en el comportamiento de los genotipos a través de los ambientes, corrija el enorme peso que tiene el propio efecto del rendimiento medio.

Por las razones expuestas anteriormente, sin olvidar el hecho de que medidas y algoritmos distintos conducen a grupos diferentes, podemos utilizar otras alternativas propuestas por distintos autores. Así: **1)** La medida de similaridad entre dos genotipos y el algoritmo de clasificación propuestos por **Lin (1982)**. Dicha medida de similaridad está basada únicamente en el efecto de la interacción **GxE**; por ello, cabe esperar una mayor intervención del perfil que muestra el comportamiento de cada genotipo en los diferentes ambientes más que su rendimiento medio. **2)** Otra alternativa es el Análisis de Grupos siguiendo el modelo propuesto por **Fox y Rosielle (1982)**; éste último está siendo muy utilizado en los últimos años.

10.1. Análisis de grupos según Lin

Lin (1982) define la disimilaridad entre dos genotipos en base a las diferencias entre los efectos medios de los genotipos en los diferentes ambientes y hace uso del método aglomerativo de pares de grupos sin ponderar. Una vez definida la disimilaridad entre dos genotipos, consiste básicamente en ir construyendo grupos de tal forma que el cuadrado medio dentro del hipotético nuevo grupo sea mínimo en cada ciclo de fusión.

La medida de disimilaridad propuesta por Lin es:

$$d(i,i') = \frac{\sum_{j=1}^e (l_{ij} - l_{i'j})^2}{2(e-1)} \tag{57}$$

donde:

$$l_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} \tag{58}$$

Figura 24 Análisis de grupos Ambientes (SINGL-DIST. EUCL.)

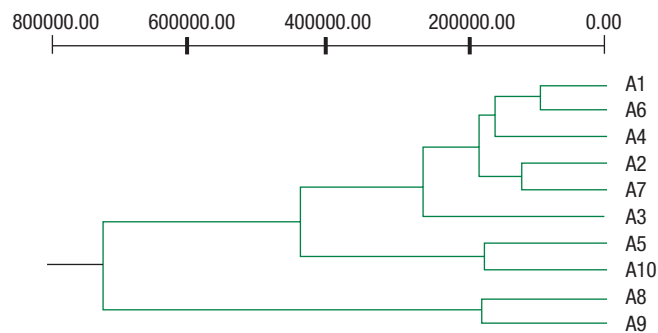
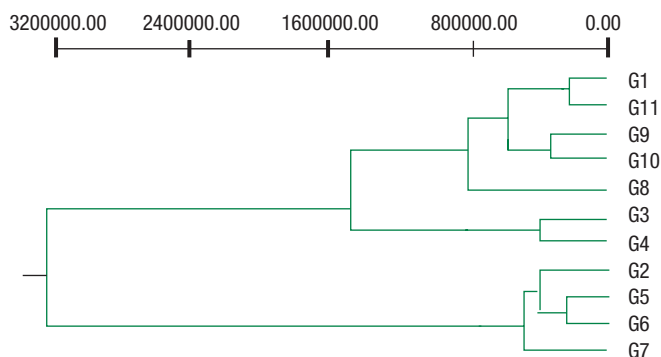


Figura 25 Análisis de grupos Genotipos (COMPL-DIST. EUCL.)



y e es el número de ambientes.

Esta medida de disimilaridad sólo tiene en cuenta el efecto de la interacción **GxE**. Como algoritmo de clasificación se utilizó el método de Sokal y Michener (1958), en el que el nuevo índice de disimilaridad para un subgrupo de r genotipos es:

$$d(1, 2, \dots, r) = \frac{2 \sum_{1 \leq i < i' \leq r} d(i, i')}{r(r-1)} \quad (59)$$

Así, se van formando aquellos nuevos grupos que presenten el mínimo valor para el nuevo índice lo que, al coincidir con el **CM** de la interacción **GxE** de los r genotipos, tal y como demostró Lin, implica ir construyendo grupos dentro de los cuales el efecto de la interacción **GxE** es homogéneo.

También proporciona un método objetivo para cortar el dendrograma, pues como el índice coincide con el **CM** de la interacción (ver Lin, 1982), ésta puede compararse con el **CM** del error experimental, y el corte se efectúa cuando esta prueba **F** resulta significativa.

Los resultados obtenidos tras aplicar el método de Lin a los genotipos y ambientes del ejemplo se muestran en las Figuras 27 y 28 y en las Tablas 19 y 20, donde, por ejemplo, $d(9,10,4)$ se calcula de la siguiente forma:

$$d(9,10,4) = 1/3 [d(9,10) + d(9,4) + d(10,4)] = 1/3 (30109.68 + 43837.32 + 47904.38) = 40617.12$$

las distancias para todos los pares de genotipos $d(9,10)$, $d(9,4)$, $d(10,4)$, ..etc. se calculan basándose en la ecuación [57].

Figura 26 Análisis de grupos Ambientes (COMPL-DIST. EUCL.)

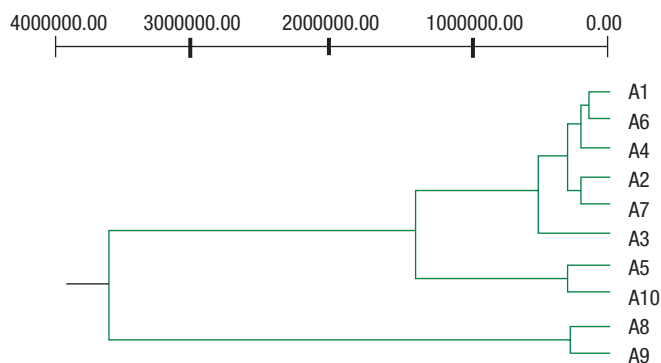


Figura 27 Análisis de grupos Genotipos (MÉTODO LIN 1982)

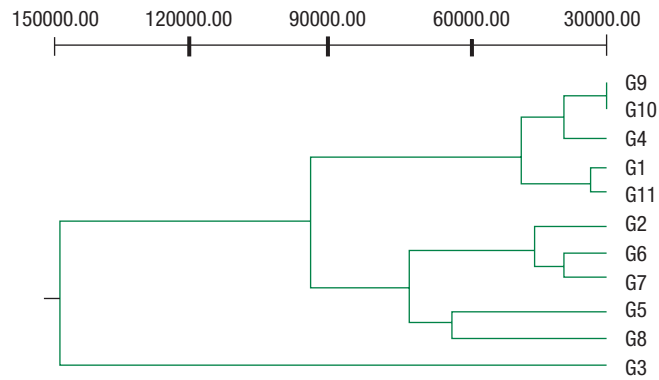


Tabla 19 Análisis de grupos de los genotipos, según el método de Lin.

CICLO DE FUSIÓN	G.L.(#)	GENOTIPOS AGRUPADOS	C.M. DE LA GxE(&)
1	9	(9,10)	30109.68
2	18	(1,11)	30906.41
3	27	(6,7)	37882.38
4	36	(9,10),4	40617.12
5	45	(6,7),2	45900.30
6	54	(9,10,4),(1,11)	49254.10
7	63	(5,8)	67447.26
8	72	(5,8),(6,7,2)	74344.10
9	81	(5,8,6,7,2),(9,10,4,1,11)	97580.30 **
10	90	(5,8,6,7,2,9,10,4,1,11),3	146425.48 ***

(#): Grados de libertad dentro de grupos = N.º ciclo x (N.º ambientes - 1)
 (&): Se contrastan con $\sigma^2_o = 57791$ con 220 g.l.
 * P<0.05 ** P<0.01 *** P<0.001

Como se puede advertir, en el caso de los genotipos (Tabla 19) el C.M. de la interacción **GxE** es significativo en el ciclo de fusión 9, constituyéndose según el criterio de Lin tres grupos homogéneos en cuanto a la interacción **GxE** (Fig. 27):

GRUPO	GENOTIPOS
G1	2,5,6,7,8
G2	1,4,9,10,1
G3	3

que resultan ser los mismos grupos formados tanto por el ACP (Capítulo 8) como por el análisis A.M.M.I. (Capítulo 9).

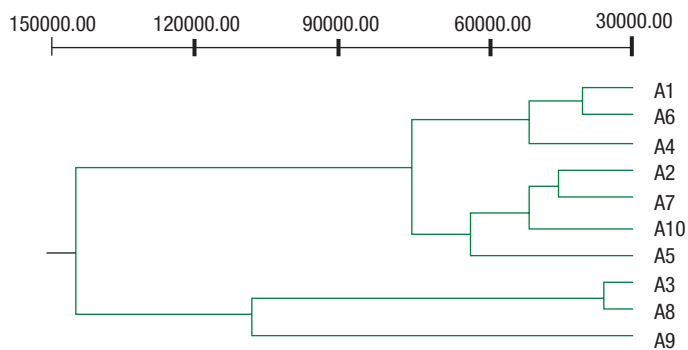
Respecto a los ambientes se puede decir algo similar. Según el método de Lin se obtienen cuatro grupos, ya que el C.M. de la interacción **GxE** es significativo en el ciclo 7 (Tabla 20 y Fig. 28).

Tabla 20 Análisis de grupos de los ambientes, según el método de Lin.

CICLO DE FUSIÓN	G.L.(#)	GENOTIPOS AGRUPADOS	C.M. DE LA GxE(&)
1	10	(1,6)	43307.00
2	20	(3,8)	43402.00
3	30	(2,7)	49635.00
4	40	(2,7),10	51687.80
5	50	(1,6),4	52836.20
6	60	(2,7,10),5	65442.00 *
7	70	(2,7,10,5),(1,6,4)	93440.60 **
8	80	(3,8),9	104200.00 ***
9	90	(2,7,10,5,1,6,4),(3,8,9)	146425.48 ***

(#): Grados de libertad dentro de grupos = N.º ciclo x (N.º genotipos - 1)
 (&): Se contrastan con $\sigma^2_o = 57791$ con 220 g.l.
 * P<0.05 ** P<0.01 *** P<0.001

Figura 28 Análisis de grupos Ambientales (MÉTODO LIN 1982)



Σ
 α

Π

Ω

\times

\pm

ϑ

Los grupos formados son los siguientes (Fig. 28):

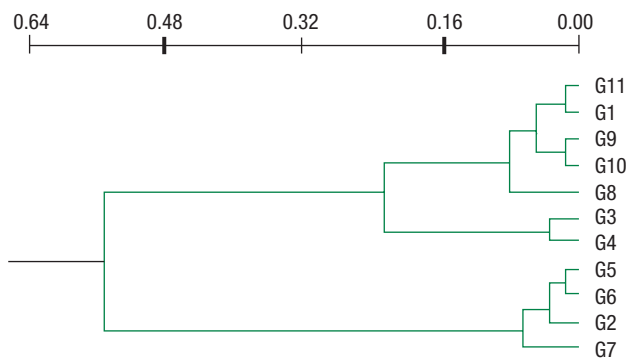
GRUPO	AMBIENTES
A1	1,6,4
A2	2,7,10,5
A3	3,8
A4	9

En los métodos anteriores no pudimos agrupar los ambientes según un patrón relacionado con la interacción **GxE**. Ahora bien, si observamos el **biplot** de la Fig. 18, conjuntamente con el agrupamiento obtenido por el método de Lin vemos que realmente existe una gran similitud y que las agrupaciones han sido realizadas según su contribución a la interacción **GxE**. Así, el grupo A1 formado por los ambientes 1, 6 y 4 son los que en el **biplot** de la Fig 18 muestran una mayor contribución de signo negativo a la interacción **GxE**. El grupo A2 formado por los ambientes 2, 7, 10 y 5 muestra una escasa contribución de signo negativo a la interacción **GxE**. El grupo A3 formado por los

ambientes 3 y 8 resultan ser los de intermedia contribución de signo positivo y el grupo A4, formado tan sólo por el ambiente 9, es precisamente el que más contribuye a la interacción **GxE**.

Como hemos podido comprobar, el método de análisis de grupos de Lin resulta útil para identificar **fácilmente** la estructura de la interacción tanto en genotipos como en ambientes. El criterio para identificar el patrón de respuesta en éste método (y en el contexto de un ensayo **GxE**) es el tamaño de los cuadrados medios de la interacción **GxE**. Se considera que los genotipos dan la misma respuesta frente a todos los ambientes si los C.M. de la interacción **GxE** no son significativamente diferentes del error estimado. De esta forma Lin pensó que “dado que todos los índices, definidos o construidos, son los C.M. de la interacción **GxE** para un correspondiente genotipo, las agrupaciones basadas en el valor crítico de F podrían satisfacer este criterio y resultar en grupos aproximadamente homogéneos” (Lin, 1982).

Figura 29 Análisis de grupos Genotipos (MÉTODO FOX-ROSIELLE)



10.2. Análisis de grupos según Fox y Rosielle

El método de Fox y Rosielle (1982) parte de datos estandarizados para ambientes cuando queremos agrupar ambientes, y datos sin estandarizar para genotipos. La medida de similitud es la **distancia euclídea al cuadrado**, y la estrategia de fusión es el **método Ward de mínimas varianzas o "Incremental Sum of Squares"**, el cual minimiza la suma de cuadrados dentro de grupos para todas las particiones.

Los resultados del análisis de grupos según Fox y Rosielle se presentan en las Tablas 21 y 22 y en las Figuras 29 y 30. Para su realización se dispone de una corriente de control para el paquete estadístico SAS (ver Apéndice).

En la Figura 29 vemos como las agrupaciones obtenidas para los genotipos coinciden con

las obtenidas por el método aglomerativo UPGMA, es decir, los genotipos han sido agrupados fundamentalmente atendiendo al efecto principal (rendimiento medio) quedando enmascarado el efecto interacción.

En cuanto a las agrupaciones obtenidas para los ambientes (Fig. 30) vemos como resulta ser muy parecida a la obtenida por el método de Lin, la diferencia estriba en que con el método de Fox y Rosielle se ha obtenido un desglose mayor del grupo que se caracteriza por tener una contribución intermedia de signo negativo a la interacción GxE, constituyéndose así el grupo 3 (ambientes 5 y 7) que resultan ser los ambientes de menor contribución a la interacción. Por otro lado se forma el grupo 4 constituido por todos los ambientes que contribuyen con signo positivo que en el método de Lin quedaban agrupados en los grupos 3 y 4.

Figura 30 Análisis de grupos Ambientes (MÉTODO FOX-ROSIELLE)

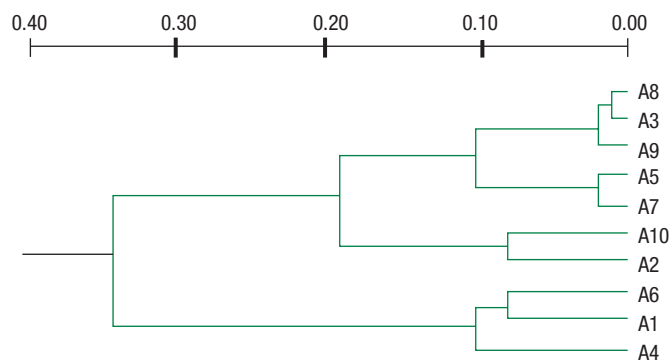


Tabla 21. Análisis de grupos de Fox y Rosielle para genotipos.

N.º CLUSTER	GRUPOS	UNIÓN	FRECUENCIA NUEVO CLUSTER	R ² SEMI-PARCIAL
10	G11	G1	2	0.0098
9	G5	G6	2	0.0140
8	G9	G10	2	0.0146
7	G2	CL9	3	0.0224
6	CL8	CL10	4	0.0375
5	CL7	G7	4	0.0452
4	G3	G4	2	0.0500
3	G8	CL6	5	0.0720
2	CL3	CL4	7	0.2363
1	CL2	CL5	11	0.4978

Tabla 22 Análisis de grupos de Fox y Rosielle para ambientes.

N.º CLUSTER	GRUPOS	UNIÓN	FRECUENCIA NUEVO CLUSTER	R ² SEMI-PARCIAL
9	A8	A3	2	0.0133
8	CL9	A9	3	0.0235
7	A5	A7	2	0.0259
6	A10	A2	2	0.0875
5	A6	A1	2	0.0900
4	CL5	A4	3	0.1145
3	CL7	CL8	5	0.1162
2	CL3	CL6	7	0.1916
1	CL2	CL4	10	0.3372



Como resumen podemos decir que las agrupaciones obtenidas para ambientes por ambos métodos es muy similar de forma que el efecto de la interacción ha podido aflorar a pesar de la importancia del efecto principal.

Método de Lin

- Grupo 1: 1, 6, 4
- Grupo 2: 2, 7, 10, 5
- Grupo 3: 3, 8
- Grupo 4: 9

Método de Fox y Rosielle

- Grupo 1: 1, 6, 4
- Grupo 2: 2, 10
- Grupo 3: 5, 7
- Grupo 4: 3, 8, 9

Diversos autores han obtenido resultados positivos con este método para sus datos, pudiendo agrupar los genotipos en base a su respuesta frente a los ambientes, y por tanto, establecer qué grupos de genotipos son más estables. En nuestro ejemplo, por el contrario, este método no supera al de Lin a la hora de explicar los datos experimentales.

REFERENCIAS CITADAS

- FOX, P.N. and A.A. ROSIELLE. 1982: Reducing the influence of environmental main-effects on pattern analysis of plant breeding environments. *Euphytica* 3, 645-656.
- LIN, C.S. 1982: Grouping genotypes by a cluster method directly related to genotype-environment interaction mean square. *Theor. Appl. Genet.* 62, 277-280.

- SOKAL, R.R. and C.D. MICHENER. 1958: A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* 38, 1409-1438.

OTRAS REFERENCIAS ACONSEJADAS

- ABOU-EL-FITTOUH, H.A., J.O. RAWLINGS, and P.A. MILLER. 1969: Classification of environments to control genotype by environment interactions with an application to cotton. *Crop Sci.* 9, 135-140.
- BROWN, J.S. 1991: Principal component and cluster analyses of cotton cultivar variability across the U.S. cotton belt. *Crop Sci.* 31, 915-922.

- BURR, E.J. 1970: Cluster sorting with mixed character types. II. Fusion strategies. *Aust. Compt. J.* 2, 98-103.

- BYTH D.E., R.L. EISEMANN and I.H.DE LACY. 1976: Two-way pattern analysis of a large data set to evaluate genotypic adaptation. *Heredity* 37, 215-230.

• CUADRAS, C.M. 1981: Métodos de análisis multivariante. Ed. por Eunibar. Barcelona.

• CHARMET, G., F. BALFOURIER, A. BION. 1990: Agronomic Evaluation of a Collection of French Perennial Ryegrass Populations: Multivariate Classification Using Genotype x Environment Interactions. *Agronomie* 10, 807-823.

• FOX, P.N., B. SKOVMAND, B.K. THOMPSON, H.J. BRAUN and R. CORMIER. 1990: Yield and adaptation of hexaploid spring triticale. *Euphytica* 47, 57-64.

• GHADERI, A., M.W. ADAMS and A.W. SAETTLER. 1982: Environmental response patterns in commercial classes of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 63, 17-22.

• GHADERI, A., E.H. EVERSON and C.E. CRESS. 1980: Classification of environments and genotypes in wheat. *Crop Sci.* 20, 707-719.

• HAIR, J.F., Jr., R.E. ANDERSON, and R.L. TATHAM. 1987: *Multivariate data analysis with readings*. Mcmillan Publ. Co., New York.

• HANSON.W.D. 1970: Genotypic stability. *Theor. Appl. Genet.* 40, 226-231.

• HAYWARD, M.D., I.H. DeLACEY, B.F. TYLER and D.W. DRAKE. 1982: The application of pattern analysis for the recognition of adaptation in a collection of *Lolium multiflorum* populations. *Euphytica* 31, 383-396.

• JOHNSON, G.R. 1977: Analysis of genotypic similarity in terms of mean yield and stability of environmental response in a set maize hybrids. *Crop Sci.* 17, 837-842.

• JUDEZ, L. 1989: Técnicas de análisis de datos multidimensionales. Publ. M^o de Agricultura, Pesca y Alimentación.

• LIN, C.S., M.R. BINNS, and L.P. LEFKOVITCH, 1986: Stability analysis: Where do we stand? *Crop Sci.* 26, 894-900.

• RAMEY, T.B., and A.A. ROSIELLE. 1983: HASS cluster analysis: a new method of grouping genotypes or environments in plant breeding. *Theor. Appl. Genet.* 64, 131-133.

• SHORTER, R. and R.J. NORMAN. 1983: Cultivar x environment interactions for kernel yield in virginia type peanuts in queensland. *Aust. J. Agric. Res.* 34, 415-426.

11

ANÁLISIS DE COORDENADAS PRINCIPALES

11. ANÁLISIS DE COORDENADAS PRINCIPALES

Westcott (1986) criticó duramente todos las metodologías univariantes disponibles en 1986 para el estudio de la interacción **GxE**. Como alternativa, propuso técnicas de análisis multivariante, aunque no consideró apropiados todos los métodos disponibles (entre otros, tanto el Análisis de Grupos como el Análisis de Componentes Principales) en base a las razones comentadas en la Introducción. Sin embargo, sugirió la utilización de otros métodos multivariantes hasta ese momento prácticamente inéditos en este campo, tales como el Análisis de Coordenadas Principales y el Análisis de Correspondencias (que veremos en el capítulo siguiente).

El análisis de coordenadas principales (CCPP) parte de una matriz de similitud y intenta efectuar la representación de las observaciones en un espacio euclídeo, mediante una distancia compatible con la similitud, tal como la denominada **distancia de Gower**. Estas medidas de similitud pueden ser muy diversas, no existiendo un criterio absoluto para elegir una de ellas. Westcott (1987) propuso la utilización de un coeficiente de similitud que permite que todos los valores propios de la matriz transformada a partir de la matriz de similitud sean positivos, y en caso de que existan algunos negativos, que éstos sean de escasa relevancia. Esto es una condición necesaria para poder representar las distancias en un espacio euclídeo.

En un ambiente **j**, si L_j y S_j denotan el mayor y el menor rendimiento del grupo de genotipos, la similitud definida por Westcott entre dos

genotipos **i** e **i'** se determina de la siguiente forma, siendo Y_{ij} e $Y_{i'j}$ los rendimientos medios de los genotipos **i** e **i'** en el ambiente **j** :

$$S(i,i') = \frac{L_j - \frac{(Y_{ij} + Y_{i'j})}{2}}{L_j - S_j} \quad (60)$$

donde si $i=i'$, entonces $S(i,i')=1$.

Cuando hay más de un ambiente, la similitud que ha de emplearse es la similitud media. Esto se realiza por ciclos sucesivos de cálculo de índices de similitud introduciendo en cada ciclo un nuevo ambiente (Tabla 23).

Para ello:

1º) Se ordenan los ambientes en función de sus valores medios, existiendo la posibilidad de dos ciclos, uno por extremo. Si se empieza por el ambiente de valor máximo, se tiene el ciclo H1, y si se comienza por el ambiente de valor mínimo, el ciclo L1.

2º) Se calcula la matriz de similitud, según [60], para cada ambiente.

3º) Comenzando por H1 o por L1, se van añadiendo ambientes sucesivamente. Así, el ciclo H2 estaría constituido por los dos ambientes de valores máximos y el ciclo L2 por los dos ambientes de valores mínimos, y así sucesivamente.

4º) Se calcula la matriz de similitud base o matriz de similitudes medias según los ambientes que constituyan el ciclo, ya que como hemos comentado antes, *cuando hay más de un*

Figura 31 Coordenadas Principales (CICLO L1) (dos dimensiones)

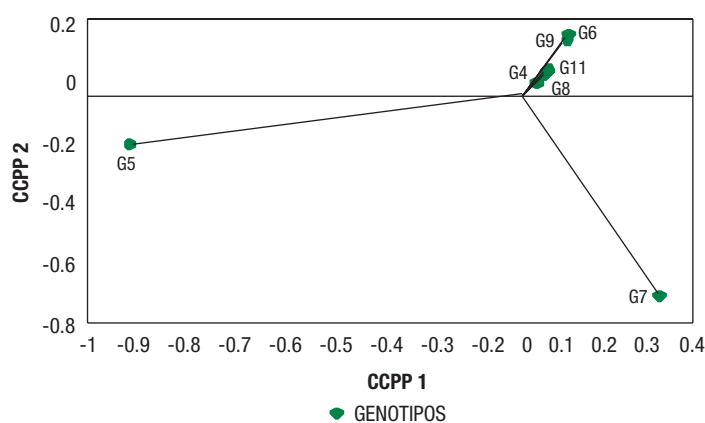
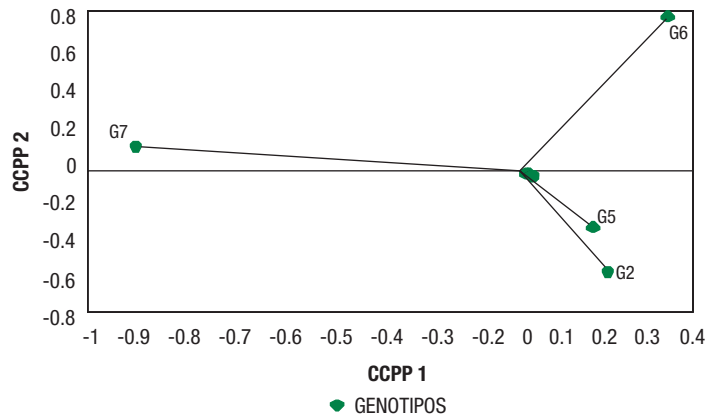


Figura 32 *Coordenadas Principales (CICLO H1) (dos dimensiones)*



ambiente, la similaridad que ha de emplearse es la similaridad media. Así por ejemplo, para el ciclo H2, se obtendría de la media de las matrices de similaridad de los dos ambientes de mayor rendimiento medio.

Esta medida de similaridad determina que, en la representación bidimensional (Figuras 31-32) o tridimensional (Figuras 33-42), aquellos genotipos que estén muy por encima de la media sean muy disimilares y por tanto quedarán representados por puntos muy alejados entre sí. Por ejemplo, para el ciclo L1 (Fig. 33) los genotipos más alejados son el 5, 7, 6 y 9, siendo por tanto estos genotipos los más productivos en este ambiente; en la matriz de similaridad (DISTL1.TXT, ver Apéndice) sus distancias recíprocas deben ser pequeñas en relación a la que tienen respecto de otros genotipos menos productivos en este ambiente. Así, la distancia entre los genotipos 5 y 7 es $S(5,7)=0.068$, mientras que la distancia entre el genotipo 5 y el menos productivo (el 4) es

$S(5,4)=0.50$. Concluyendo, pequeños valores en la matriz de similaridad nos indican que el genotipo en cuestión es muy productivo y se representará, por tanto, en la gráfica de coordenadas principales como un genotipo alejado; valores grandes nos indican que dicho genotipo es muy poco productivo en el ambiente o grupo de ambientes del ciclo en cuestión y, por tanto, se representará como un punto central en la gráfica. De este modo, los “buenos” genotipos son aquellos que están muy alejados del centro, siendo inmediata su identificación. Los genotipos estables son aquellos que, consistentemente, son “buenos” en los ciclos sucesivos conforme se van incorporando nuevos ambientes.

En la Tabla 23 se muestra la composición de ambientes para cada ciclo del análisis de coordenadas.

Una vez obtenida la similaridad, las coordenadas principales se obtienen como sigue:

Figura 33 *Coordenadas Principales (CICLO L1)*

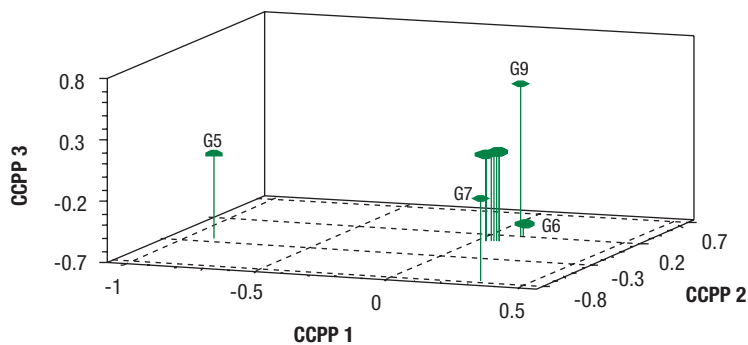
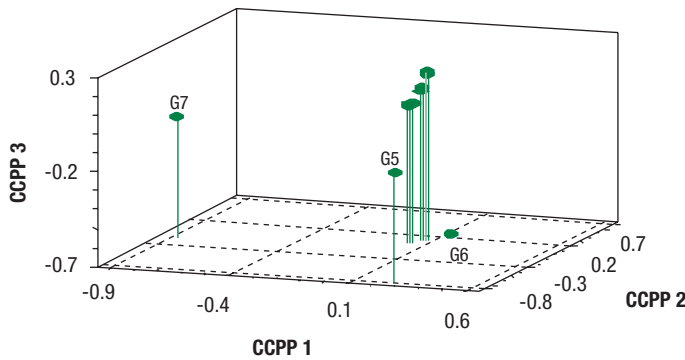


Figura 34 Coordenadas Principales (CICLO L2)



1) La matriz de similaridad **S(i,j)** se transforma en una nueva matriz **T(i,j)**, de la siguiente forma:

$$t_{ij} = s_{ij} - \bar{s}_{i.} - \bar{s}_{.j} + \bar{s}_{..} \quad (61)$$

2) Se diagonaliza la matriz **T**:

$$T = B D_{\mu} B' = B D^{1/2}_{\mu} D^{1/2}_{\mu} B' = W W' \quad (62)$$

donde:

B = matriz de los vectores propios de **T** normalizados.

D = matriz diagonal de los valores propios de **T**, y

W = $B D^{1/2}_{\mu}$ es la matriz de coordenadas principales, por lo que para una representación en dimensión **d**, se toman las **d** primeras columnas de **W**.

Para comprobar la variabilidad total, representada por la dimensión **d**, se suman los valores propios hasta el valor propio de la dimensión **d**, y se divide por la suma de todos los valores propios, tal y como se realiza en el Análisis de Componentes Principales.

En definitiva, el Análisis de Coordenadas Principales permite representar objetos respecto a una distancia deducida del coeficiente de similitud.

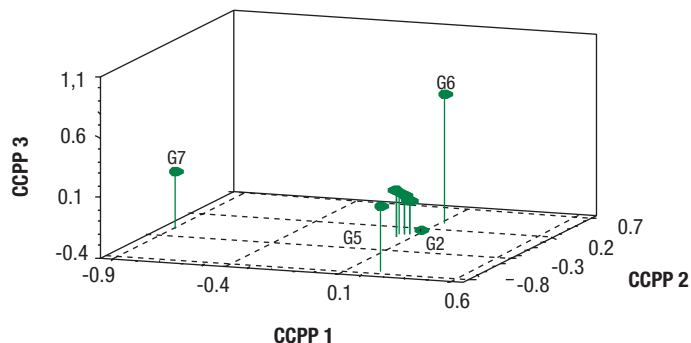
Árbol de mínima dispersión. Gower y Ross (1969) proponen calcular un árbol de mínima dispersión, para corregir la pérdida de información que conlleva pasar de un espacio de **e** dimensiones a uno bidimensional (se representa en las Figuras 31-32).

Dados **g** puntos en un espacio de **e** dimensiones, siempre es posible trazar tal árbol

Tabla 23 Ciclos H y L para el análisis de Coordenadas Principales.

CICLO	AMBIENTE AÑADIDO	REND. MEDIO MEDIO	AMBIENTES QUE FORMAN EL CICLO
L1	10	2633	10
L2	5	2776	10,5
L3	3	3304	10,5,3
L4	4	3433	10,5,3,4
H1	8	4422	8
H2	9	4259	8,9
H3	7	3784	8,9,7
H4	2	3686	8,9,7,2
H5	1	3607	8,9,7,2,1
H6	6	3561	8,9,7,2,1,6

(1) Ambientes por debajo de la media total del ensayo.
 (2) Ambientes por encima de la media total del ensayo.

Figura 35 *Coordenadas Principales (CICLO L3)*

mediante líneas rectas que unen pares de puntos, si se verifican las siguientes condiciones:

- En ningún caso se forman polígonos cerrados.
- Por cada punto pasa al menos una línea.
- El árbol no presenta discontinuidades.

Existen muchos algoritmos para llevar a cabo la construcción de un árbol de mínima dispersión. Se ha utilizado el primer algoritmo de Gower y Ross (1969). Este consiste en un proceso iterativo que comienza uniendo los puntos más próximos entre sí y continúa de modo que siempre se verifiquen las condiciones anteriores. Para nuestro ejemplo, para el cálculo de las longitudes, del árbol de mínima dispersión, entre genotipos hemos utilizado el subprograma MST (Minimum Spanning Tree) del paquete estadístico NTSYS.

11.1. Caso práctico

Los resultados de aplicar el análisis de coordenadas principales a los datos del ejemplo que estamos siguiendo se han obtenido de la siguiente forma:

1º) Mediante un programa en BASIC (ver Apéndice) hemos calculado la matriz de distancias para cada ambiente, utilizando la fórmula [60].

2º) Por medio de otro sencillo programa en BASIC (ver Apéndice) se calculan las matrices de distancias para cada ciclo. Consiste en calcular las medias para cada ciclo según los ambientes que lo forman.

3º) Con el paquete estadístico NTSYS realizamos un doble centrado, es decir, calculamos la matriz **T** según [61]. Para ello se ha utilizado el subprograma "Double Center".

4º) Calculamos los vectores propios de la matriz **T**, con el subprograma "Eigenvectors", y los representamos en dos y tres dimensiones, con el subprograma "Matrix Plot - graphics" y "3-D Model - graphics" respectivamente, dándonos como resultado las Figuras 31-42.

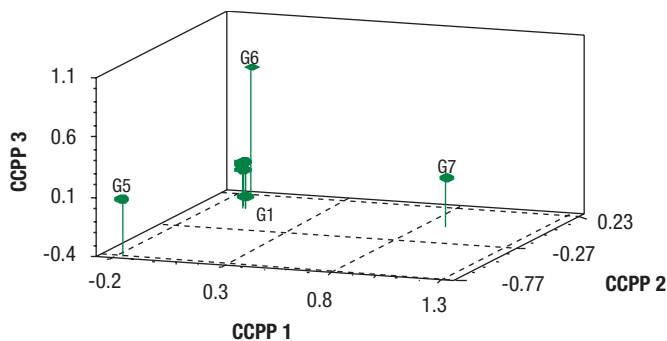
5º) Como ya se ha dicho, para las longitudes del árbol de mínima dispersión se ha utilizado el subprograma "MST" del paquete estadístico NTSYS.

Los tres primeros ejes del análisis de coordenadas, para todos los ciclos, contribuyen aproximadamente, en nuestro caso, con el 44 % de la variación total, lo que representa un porcentaje similar al obtenido por otros autores en diferentes materiales.

Las representaciones gráficas de los análisis de coordenadas principales se muestran de la Figura 31 a la 42. Las dos primeras (Fig. 31-32) representan las coordenadas principales en los dos primeros ejes mostrándose también el **árbol de mínima dispersión** cuyas longitudes obtenidas para todos los ciclos son las que refleja la Tabla 24. De ésta forma resulta más fácil identificar aquellos genotipos más disimilares (o de mayor rendimiento) en cada ciclo.

Como puede observarse en la Tabla 24, se han representado los valores para los seis primeros genotipos en cada ciclo. Tanto en la Tabla 24 como en las Figuras 31-42 se observa que

Figura 36 Coordenadas Principales (CICLO L4)



cuatro genotipos son los más disimilares y, por tanto, los de mayor rendimiento en cada ciclo: 7,6,5 y 2.

Aunque para los datos de nuestro ejemplo resulta claro definir el orden de los genotipos

más disimilares, cuando tengamos datos en los que resulte difícil determinar dicho orden podemos calcular la media de las longitudes del árbol de mínima dispersión (AMD) para cada genotipo por ciclos (Tabla 25).

Tabla 24 Árbol de mínima dispersión para los primeros genotipos de cada ciclo.

CICLO	GENOTIPOS	LONGITUD	CICLO	GENOTIPOS	LONGITUD
L1	G4-G5	50.0	H2	G3-G7	50.0
	G4-G7	43.2		G3-G2	47.3
	G4-G6	33.8		G3-G6	45.4
	G4-G9	33.0		G3-G5	43.8
	G4-G11	23.0		G3-G1	31.7
	G4-G10	23.0		G3-G8	29.1
L2	G4-G7	46.6	H3	G3-G7	54.9
	G4-G5	39.5		G3-G2	52.7
	G4-G6	33.2		G3-G6	46.9
	G4-G11	27.0		G3-G5	45.7
	G4-G2	26.8		G3-G1	38.7
	G4-G1	26.7		G3-G11	37.5
L3	G4-G7	51.6	H4	G4-G7	54.4
	G4-G5	44.4		G4-G2	50.5
	G4-G6	39.8		G4-G5	50.4
	G4-G2	36.0		G4-G6	46.7
	G4-G1	31.2		G4-G11	41.4
	G4-G10	30.5		G4-G1	38.2
L4	G4-G7	55.5	H5	G4-G7	57.1
	G4-G5	43.2		G4-G5	51.9
	G4-G6	39.3		G4-G6	51.7
	G4-G1	38.1		G4-G2	50.5
	G4-G2	36.4		G4-G11	43.6
	G4-G10	36.0		G4-G1	42.8
H1	G3-G7	50.0	H6	G3-G7	58.9
	G3-G6	47.1		G3-G6	52.6
	G3-G2	45.6		G3-G5	49.4
	G3-G5	44.9		G3-G2	47.8
	G3-G1	27.2		G3-G1	45.7
	G3-G10	25.0		G3-G11	45.2

Figura 37 Coordenadas Principales (CICLO H1)

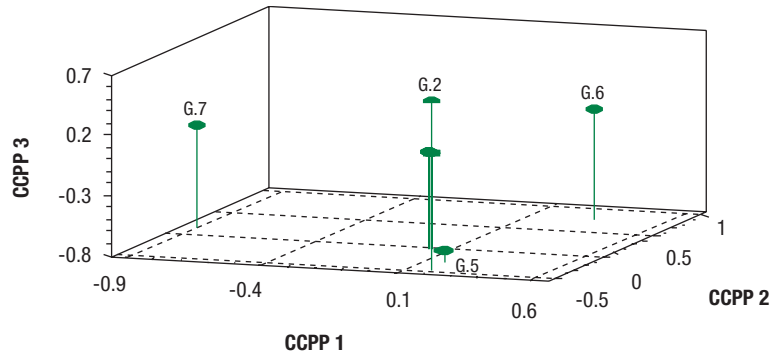


Tabla 25 Longitudes medias del AMD para los genotipos más disimilares.

GENOTIPO	CICLO L	CICLO H	RANGO#
7	49.22(1)*	54.21(1)	= 2 (1)
5	44.27(2)	47.68(4)	= 6 (2.5)
6	36.52(3)	48.40(3)	= 6 (2.5)
2	29.20(5)	49.06(2)	= 7 (3)
1	29.40(4)	37.38(5)	= 9 (4)
10	27.92(6)	32.13(7)	= 13 (5.5)
11	27.67(7)	36.61(6)	= 13 (5.5)

* Entre paréntesis, rango del genotipo dado en ese ciclo.

Suma de los rangos anteriores y número de orden final (entre paréntesis).

De la Tabla 25 podemos extraer las siguientes conclusiones:

1) Tres genotipos resultan ser los más estables, con diferencia de los demás, en el ciclo L (de bajo rendimiento): el 7, 5 y el 6.

2) El genotipo 7 resulta ser el más estable en los ambientes de alto rendimiento (Ciclo H), seguido de los genotipos 5, 6 y 2.

Por tanto, para los genotipos y ambientes analizados en este ejemplo, el análisis de CC.PP. origina unos resultados satisfactorios para la detección de genotipos de alto rendimiento que permanecen estables bajo diferentes condiciones ambientales, además de coincidentes con los obtenidos por métodos anteriores, puesto que resultan ser los genotipos agrupados en el Grupo 1 del A.M.M.I.

Figura 38 Coordenadas Principales (CICLO H2)

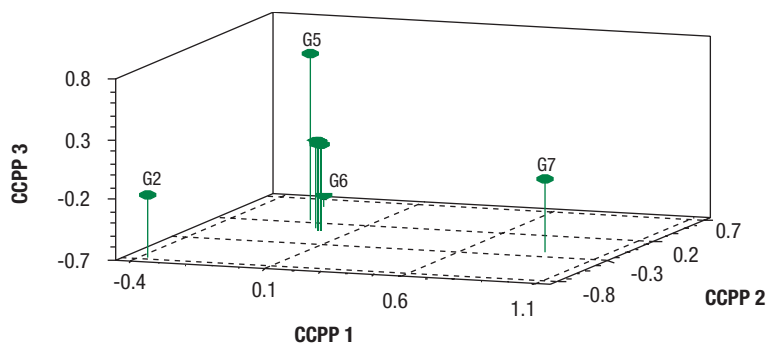
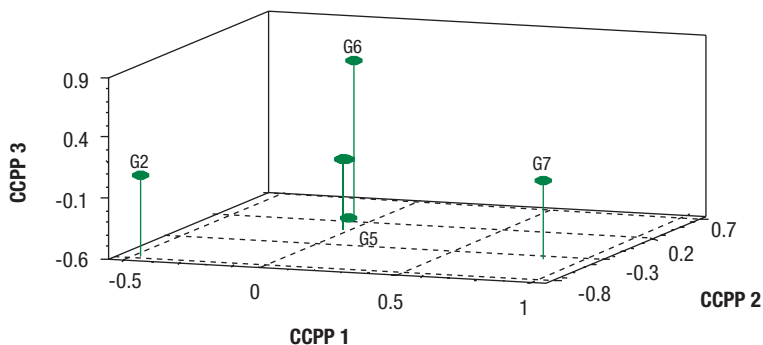


Figura 39 *Coordenadas Principales (CICLO H3)*



Como conclusión podemos enumerar las principales ventajas de éste método:

- 1) Es altamente fiable cuando se utiliza para analizar un grupo de datos que incluyen resultados de ambientes extremos de bajo y alto rendimiento.
- 2) Identifica fácilmente los genotipos estables en una secuencia gráfica generada por los distintos ciclos del análisis.
- 3) Se trata de un modelo “libre” (es decir, no se requiere ninguna condición respecto a la

distribución de las variables) y puede ser aplicado sin problemas a un pequeño grupo de ambientes.

De los resultados expuestos y del conocimiento que se posee acerca del comportamiento de los genotipos en base a análisis anteriores y que, como veremos mas adelante, quedan confirmados por los análisis de correspondencias y canónico, se deduce que cuando existen genotipos que cumplan las características deseadas se ponen en evidencia por éste método los más estables, **de ahí que alentemos a una mayor utilización de éste método.**

Figura 40 *Coordenadas Principales (CICLO H4)*

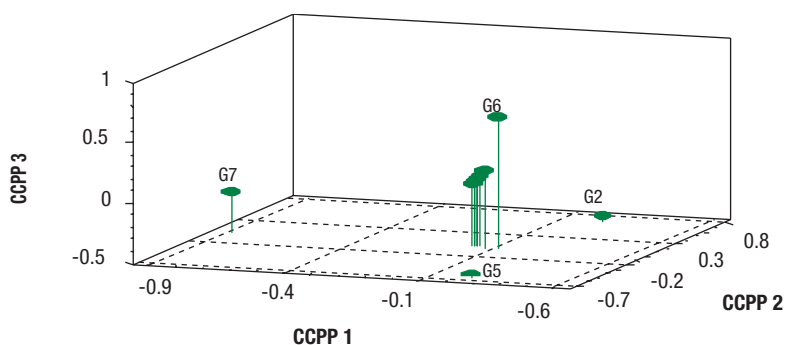


Figura 41 *Coordenadas Principales (CICLO H5)*

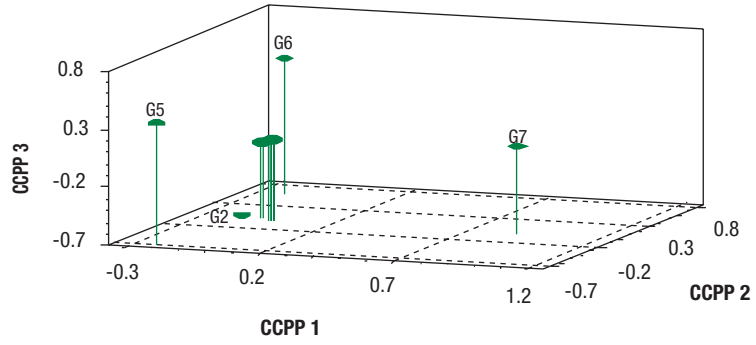
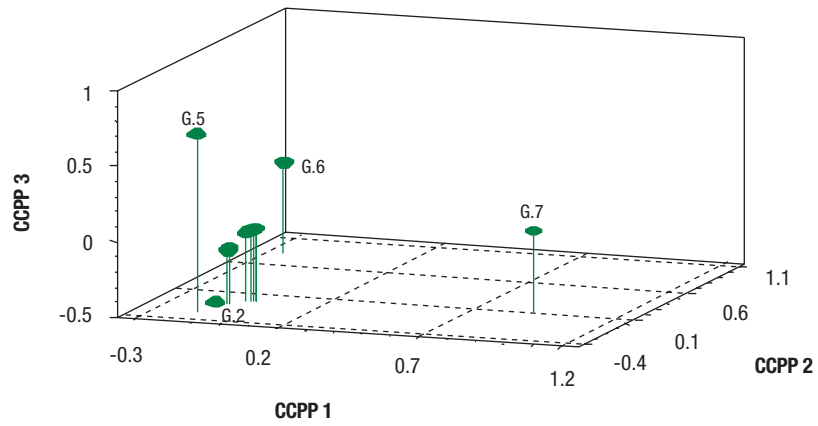


Figura 42 *Coordenadas Principales (CICLO H6)*



REFERENCIAS CITADAS

- GOWER J. C. and G. J. S. ROSS. 1969: Minimum Spanning Trees and Single Linkage Cluster Analysis. *Applied statistics*. 18, 54-64.
- WESTCOTT, B. 1986: Some methods of analysing genotype-environment interaction. *Heredity* 56, 243-253.
- WESTCOTT, B. 1987: A method of assessing the yield stability of crop genotypes. *J. Agric. Sci., Camb.* 108, 267-274.



OTRAS REFERENCIAS ACONSEJADAS

- CROSSA, J., 1988: A comparison of results obtained with two methods for assessing yield stability. *Theor Appl. Genet.* 75, 460-467.
- CROSSA, J., B. WESTCOTT and C. GONZALEZ. 1988: The Yield Stability of Maize Genotypes Across International Environments: Full Season Tropical Maize. *Expl. Agric.* 24, 253-263.
- CROSSA, J., B. WESTCOTT and C. GONZALEZ. 1988: Analysis in yield stability of maize genotypes using a spatial model. *Theor Appl. Genet.* 75, 863-868.
- CROSSA, J., B. WESTCOTT and C. GONZALEZ. 1989: The yield stability of CIMMYT'S maize germplasm. *Euphytica* 40, 245-251.
- CROSSA, J., 1990: Statistical analyses of multilocation trials. *Adv. Agr.* 44, 55-85.
- CROSSA, J., P.N. FOX, W.H. PFEIFFER, S. RAJARAM and H.G. GAUCH, Jr. 1991: AMMI adjustment for statistical analysis of an international wheat yield trial. *Theor. Appl. Genet.* 81, 27-37.
- GOWER, J.C. 1966: Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 3, 325-338.





12

ANÁLISIS FACTORIAL DE CORRESPONDENCIA

12. ANÁLISIS FACTORIAL DE CORRESPONDENCIAS

El Análisis Factorial de Correspondencias (AFC) es un tipo especial de Análisis de Componentes Principales aunque con propiedades que permiten una mejor representación de la estructura de ciertas tablas de datos (por ejemplo, cuando los datos representan frecuencias) basado en una distancia especial llamada *chi-cuadrado*, cuya principal propiedad viene dada por la posibilidad de efectuar una representación simultánea de las variables y de las poblaciones.

El análisis de correspondencias ha sido un método olvidado en el contexto de la interacción **GxE**. En principio parece lógico que así fuera, ya que es un método especialmente diseñado para el estudio de tablas de contingencia, en las cuales cada casilla contiene el número de individuos que cumplen simultáneamente las características deseadas. Sin embargo, también es aplicable a los casos en que se dispone de una matriz de observaciones y variables y se considera que los valores que se encuentran en cada casilla son el resultado de expresar una cantidad total (en este caso, el rendimiento) en función del genotipo (observación) y del ambiente (variable) considerado. Hill (1974) y Westcott (1986) propusieron investigar sus posibilidades en el estudio de la interacción **GxE**.

El análisis de correspondencias asigna a cada genotipo un peso específico igual a uno y se basa en el comportamiento de los genotipos en cada uno de los ambientes, es decir, *proyecta los genotipos de acuerdo con su perfil y no en función del nivel de rendimiento* (como nos ocurría en el análisis de componentes principales y en el análisis de coordenadas principales). El AFC utiliza el nivel de rendimiento sólo como identificativo del genotipo.

Teóricamente, es suficiente que las sumas de filas y columnas sean positivas. En nuestra matriz, todos los valores son positivos, y suponemos que el rendimiento total observado se reparte en función del genotipo y del ambiente considerados.

Sea **Y(I,J)** la matriz de datos originales, donde cada casilla contiene el valor observado Y_{ij} del genotipo *i*-ésimo en el ambiente *j*-ésimo. Se procede de la siguiente forma:

1) Se obtienen tres nuevas matrices a partir de **Y**:

$$X(I,J)=Y_{ij}/(Y_{i.}^{1/2} Y_{.j}) \quad (63)$$

$$M(I,J)=Y_{ij}/(Y_{i.}^{1/2} Y_{.j}^{1/2}) \quad (64)$$

$$Z(I,J)=Y_{ij}/(Y_{i.} Y_{.j}^{1/2}) \quad (65)$$

2) Se obtiene la matriz **M'** traspuesta de **M** y calculamos **P=M'M**

3) Diagonalizamos la matriz **P**: $P=T D_{\mu} T'$

donde:

T=matriz de los vectores propios de **P** normalizados,

D=matriz diagonal de los valores propios de **M**.

Las coordenadas de los genotipos en un espacio de dimensión **d**, son las **d** primeras columnas de: **A=X T**, y las coordenadas de los ambientes, en dimensión **d**, las columnas de: **B=Z M T D_μ^{-1/2}**.

Hay que añadir que en ambos casos se prescinde de la primera columna ya que en el proceso de diagonalización el primer elemento de la diagonal es 1. Por tanto, las coordenadas vendrán dadas por la 2ª, 3ª, ... columnas de las matrices **A** y **B**. Para comprobar la variabilidad total, representada por la dimensión **d**, se suman los valores propios (excepto el primero) hasta el valor propio de la dimensión **d**, y se divide por la suma de todos los valores propios, exceptuando nuevamente el primero.

12.1. Caso práctico

Para la realización del Análisis Factorial de Correspondencias se ha utilizado el subprograma "Correspondence Analysis" del paquete estadístico NTSYS.

Las coordenadas tanto de los genotipos como de los ambientes se pueden observar en la Tabla 26 y en la Fig. 43 podemos ver los resultados del análisis de correspondencias plasmados en forma de gráfica.



Figura 43 Análisis de Correspondencias

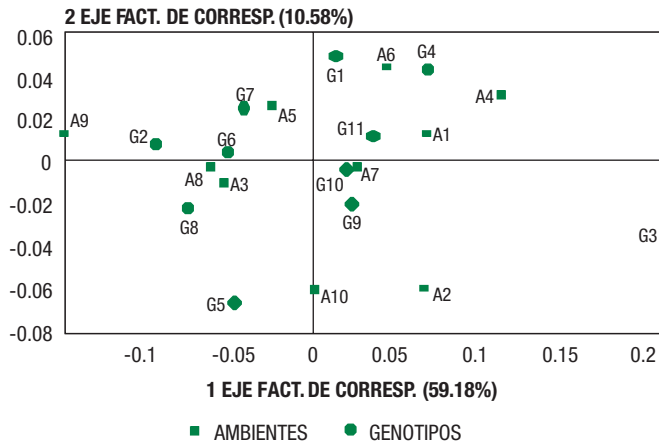


Tabla 26 Las tres primeras coordenadas del Análisis Factorial de Correspondencias.

GEN.	GENOTIPOS			AMB.	AMBIENTES		
	1ª	2ª	3ª		1ª	2ª	3ª
1	0.014	0.049	-0.011	1	0.067	0.013	0.017
2	-0.092	0.008	-0.030	2	0.066	-0.059	-0.026
3	0.209	-0.034	-0.024	3	-0.052	-0.010	0.029
4	0.068	0.043	0.069	4	0.111	0.031	-0.010
5	-0.046	-0.066	-0.002	5	-0.024	0.026	-0.086
6	-0.050	0.004	0.012	6	0.043	0.044	0.035
7	-0.040	0.025	-0.029	7	0.026	-0.003	-0.002
8	-0.073	-0.022	0.045	8	-0.060	-0.003	0.011
9	0.023	-0.020	0.030	9	-0.146	0.013	-0.004
10	0.020	-0.004	0.005	10	0.001	-0.060	0.022
11	0.036	0.012	-0.040				

%Varianza explicada por los tres primeros ejes factoriales **59.18** **10.58** **9.85**
%Varianza acumulada **59.18** **69.75** **79.60**

Figura 44 A. F. de Correspondencias (Genotipos)

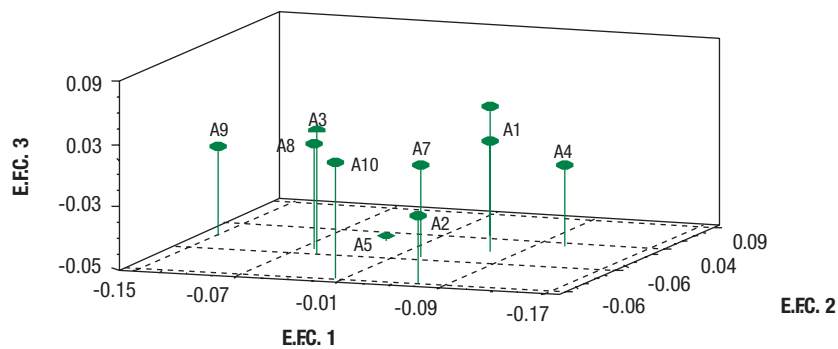
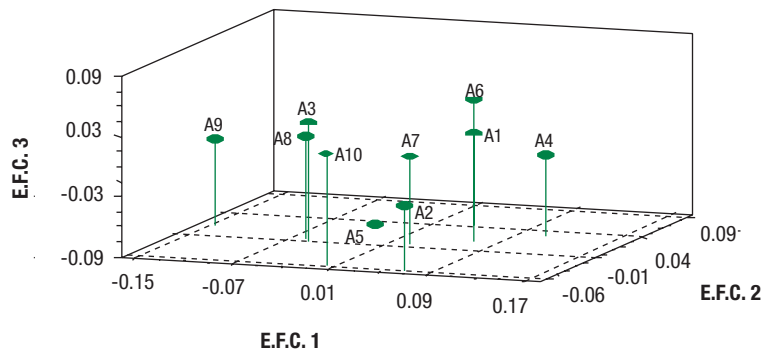


Figura 45 A. F. de Correspondencias (Ambientes)



En la Figura 43 podemos ver la proyección de los genotipos y de los ambientes en el primer plano factorial. Su estudio pone de manifiesto la existencia de asociación entre genotipos y ambientes, lo que se traduce en que aquellos no son independientes de estos, es decir: **existe interacción GxE**. Este hecho puede comprobarse fácilmente, ya que ni los ambientes ni los genotipos se han distribuido aleatoriamente en el plano de proyección, sino que se han formado unas determinadas agrupaciones, que pueden observarse mejor en la representación tridimensional de las Figuras 44 y 45. En el caso de independencia, tanto los genotipos como los ambientes se habrían distribuido aleatoriamente en forma de nube de puntos alrededor del origen.

Los tres primeros ejes factoriales explican el 59.18 %, 10.58 % y el 9.85 % de la variabilidad total, respectivamente, lo que supone en total un 79.60 %, por lo que puede decirse que sólo con éstas tres dimensiones se explica un buen porcentaje de la variabilidad total.

Si comparamos la Fig. 43 con la Fig. 18 del "biplot" del análisis AMMI, vemos como el primer eje factorial de correspondencias dispone tanto genotipos como ambientes en el mismo orden relativo que lo hacía el eje de coordenadas del

IACP1 del biplot, obteniéndose así las mismas agrupaciones entre genotipos y ambientes que las ya explicadas para el AMMI1.

Por otro lado, el AFC permite efectuar el estudio particular de cada uno de los genotipos, pudiendo extraer más información que con el análisis de grupos o en los casos en que el análisis de CCPP no ofrece conclusiones claras, ya que con el AFC podemos concluir qué genotipos son los más estables sin más que observar su situación en la gráfica respecto del origen: los genotipos más próximos al origen resultan ser los más estables, pudiéndose construir *regiones concéntricas de estabilidad a partir del origen*, resultando, por tanto, que los genotipos más estables por este método son el 9, 11, 10, 7 y 6.

Como conclusión, cabe decir, que este método se muestra potencialmente muy adecuado para el estudio de la interacción **GxE** pues permite visualizar de forma gráfica la presencia o ausencia de dicha interacción, a la vez que facilita la posibilidad de efectuar agrupaciones de genotipos en función de su comportamiento, es decir, de su aportación a la interacción **GxE**, como es nuestro caso: los que aporten menos, aquellos situados en regiones concéntricas más próximas al origen, serán los más estables.

Σ

α

Π

Ω

X

±

∂

REFERENCIAS CITADAS

- HILL . M. O. 1974: Correspondence Analysis: A Neglected Multivariate Method. Applied statistics 23, 340-354.
- WESTCOTT, B., 1986: Some methods of analysing genotype-environment interaction. *Heredity* 56, 243-253.

OTRAS REFERENCIAS ACONSEJADAS

- BENZECRI, J.P. 1970: L'analyse des données. II. L'analyse des correspondances. Dunod. París.
- BENZECRI, J.P. and F. BENZECRI. 1980: La pratique de l'analyse des données. I. Analyse des correspondances. Dunod. París.
- CUADRAS, C.M. 1981: Métodos de análisis multivariante. Ed. por Eunibar. Barcelona.
- GREENACRE, M.J. 1987: Theory and applications of correspondance analysis. Academic Press. London.
- JUDEZ, L. 1989: Técnicas de análisis de datos multidimensionales. Publ. M^o de Agricultura, Pesca y Alimentación.
- LEFEBVRE, J. 1983: Introduction aux analyses statistiques multidimensionales. Masson. París.

13

Σ α Π Ω **ANÁLISIS CANÓNICO** ϑ

13. ANÁLISIS CANÓNICO

Las técnicas de representación de datos desarrolladas anteriormente se han basado en distancias como la euclídea, la distancia de Gower, la distancia chi-cuadrado, etc. El análisis canónico (AC) utiliza **la distancia D de Mahalanobis**. El AC es un método de representación de grupos, a lo largo de ejes con máximo poder de discriminación en relación a la distancia D.

La distancia de Mahalanobis entre los grupos π_i y π_j es la distancia estadística entre sus individuos medios respectivos:

$$D^2(\pi_i, \pi_j) = (\bar{Y}_i - \bar{Y}_j)' S^{-1} (\bar{Y}_i - \bar{Y}_j) \quad (66)$$

siendo **S** la matriz de covarianzas.

Un aspecto interesante del análisis canónico es la posibilidad de construir y representar regiones de confianza para los individuos medios de los diferentes grupos (ver Figura 46).

Como en el caso del Análisis de Componentes Principales, este análisis define la posición de cada individuo sobre un nuevo conjunto de ejes, pero ahora la elección de éstos ejes no se realiza maximizando la variación total en cada eje, sino el cociente de las varianzas **entre** grupos y **dentro** de grupos.

El Análisis Canónico comienza con un conjunto predefinido de grupos (lo que obliga a haber aplicado antes otro tipo de análisis) y procede a investigar las interrelaciones entre estos grupos. Además, los grupos se representan en

unos ejes en los que sus individuos han sido proyectados según una proyección canónica (esto es, la transformación de la matriz Y de datos en otra matriz X (ver [67]), la cual contiene las coordenadas de una configuración euclídea en dimensión $m = \min(g, k - 1)$, siendo g el número de genotipos y k el número de grupos predefinidos), pudiéndose investigar la posición y orientación de unos grupos respecto de otros.

El AC, pues, nos permite pasar de la matriz de datos medios **Y** a la matriz de datos **X**, referidas a las variables canónicas V_i :

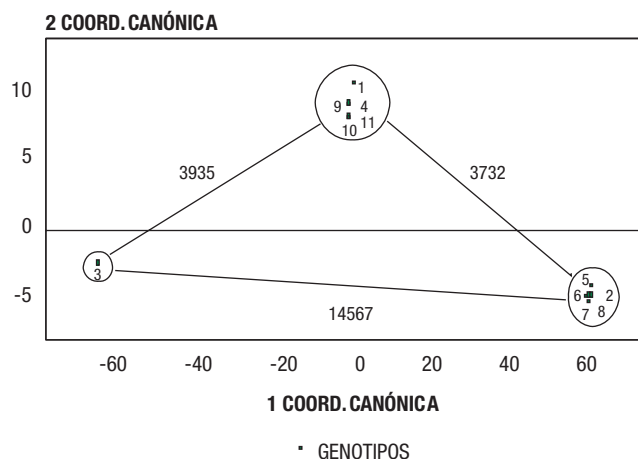
$$X = Y V \quad (67)$$

Procedimiento para la obtención de la matriz X.

Partimos de una matriz de medias **Y (gxe)**, de **g** individuos (genotipos) y **e** variables (ambientes), en la que los individuos se agrupan en **k** grupos; por tanto, la matriz **Y** se considera subdividida en **k** submatrices (Y_1, Y_2, \dots, Y_k).

Los datos se transforman primeramente restándole a cada uno el valor medio de cada columna; obtenemos así la matriz **H** y, por tanto, las submatrices H_1, H_2, \dots, H_k . La suma de cuadrados y productos total se obtendrá del producto matricial $T = H'H$. Dentro de cada grupo se obtiene también la matriz de suma de cuadrados y productos $W_i = H_i'H_i$, y sumando todas estas matrices obtenemos la matriz de suma de cuadrados y productos total *dentro de grupos* $W = \sum W_i$. De esta forma, la matriz de suma de cuadrados y productos *entre grupos* se calcularía por diferencia: $B = T - W$.

Figura 46 Análisis Canónico





El AC busca una combinación lineal de las \mathbf{e} variables de tal forma que se obtenga la mayor variación **entre** grupos en relación con la variabilidad **dentro** de grupos. La primera combinación lineal \mathbf{v}_1 se elige de forma que el cociente [variación entre grupos]/[variación dentro de grupos] sea máximo:

$$[\mathbf{v}'_1 \mathbf{B} \mathbf{v}_1] / [\mathbf{v}'_1 \mathbf{W} \mathbf{v}_1]$$



La segunda combinación lineal maximiza de nuevo el cociente anterior referido a \mathbf{v}_2 , es decir:

$$[\mathbf{v}'_2 \mathbf{B} \mathbf{v}_2] / [\mathbf{v}'_2 \mathbf{W} \mathbf{v}_2]$$

cumpliendo la condición $\mathbf{v}'_1 \mathbf{W} \mathbf{v}_2 = 0$, y así sucesivamente.



El proceso de extracción de variables canónicas se repetirá hasta que estas igualen a $\mathbf{m} = \min.(\mathbf{k} - \mathbf{1}, \mathbf{e})$ (\mathbf{k} =número de grupos), dado que a partir de \mathbf{m} las demás coordenadas son constantes.

De esta forma se obtendría la matriz \mathbf{V} del cambio de coordenadas; introduciéndola en [67] obtenemos la matriz de coordenadas canónicas \mathbf{X} .

Como resumen, en el AC se siguen los siguientes pasos:

1) Transformar las variables hasta que la matriz de covarianzas dentro de grupos sea una matriz identidad.

2) Calcular las medias de grupos sobre las variables transformadas.

3) Realizar un Análisis de Componentes Principales sobre las medias, ponderando cada media con el número de observaciones de cada grupo. Los valores propios serán igual al cociente obtenido al dividir la variación entre grupos por la variación dentro de grupos en la dirección de cada componente principal.

4) Transformar los componentes principales al espacio de las variables originales, obteniéndose así las variables canónicas.

Las principales características de la representación canónica son:

1) Las variables canónicas son de varianza unidad e independientes entre sí. Por lo tanto, trazaremos los ejes canónicos como ejes ortogonales.

2) Las distancias geométricas euclídeas de las representaciones canónicas de los grupos coinciden con las distancias de Mahalanobis, salvo la pérdida debida a la reducción de la dimensión.

3) En el caso de normalidad multivariante y homogeneidad de covarianzas, las proyecciones de los vectores de medias teóricas están contenidas en regiones confidenciales esféricas, que son círculos en el caso de dos dimensiones, de radios $R_e / \sqrt{N_t}$ ($t = 1, \dots, k = \text{grupos}$).

La utilización de estas **esferas de confianza** permite dar una visión precisa del grado de separación o confusión entre los diferentes grupos.

13.1. Caso práctico

Para nuestro ejemplo se han predefinido 3 grupos. Estos grupos son los resultantes de los análisis de ACP, AMMI, ANALISIS DE GRUPOS y AFC que, como vimos, eran coincidentes:

Grupo 1: genotipos 2, 5, 6, 7 y 8,
Grupo 2: genotipos 1, 4, 9, 10 y 11,
Grupo 3: genotipo 3.

Se ha utilizado para la realización del Análisis Canónico el subprograma CANDISC del paquete estadístico SAS (Sarle, 1985), creando para ello una sencilla corriente de control (ver Apéndice). Los principales datos obtenidos se muestran en las Tablas 27 y 28

En la Fig. 42 se muestra la gráfica de las dos primeras coordenadas canónicas, las cuales representan el 98.28 % y el 1.72 %, respectivamente, es decir, que el 100 % de las interrelaciones de estos grupos predefinidos queda explicado por estos dos ejes.

En esta misma figura también se han colocado las distancias entre grupos, comprobándose que el grupo 2 es un grupo intermedio entre los otros dos, como ya se ha comentado en los métodos anteriores, resultando estar compuestos precisamente por los genotipos considerados como más estables.

Como conclusión podemos decir que el análisis canónico nos ha distinguido claramente estos grupos predefinidos, sugiriendo que esta clasificación y ordenación **no tiene nada de arbitraria en nuestro caso**.



Tabla 27 Coordenadas canónicas para cada genotipo.

GENOTIPO	1.ª COORD. CANÓN.	2.ª COORD. CANÓN.
1	1.578	10.794
2	60.327	-4.674
3	-62.013	-2.375
4	0.450	9.367
5	60.542	-4.045
6	59.874	-5.152
7	59.550	-4.804
8	60.008	-4.703
9	0.433	9.275
10	0.229	8.259
11	0.161	8.407
%Varianza explicada por las dos primeras Coord. canónicas	98.28	1.72
%Varianza acumulada	98.28	100.00

Tabla 28 Coordenadas canónicas para el individuo medio de cada grupo.

GRUPO	1.ª COORD. CANÓN.	2.ª COORD. CANÓN.
1	60.060	-4.675
2	0.570	9.221
3	-60.631	-4.545

REFERENCIAS CITADAS

- SARLE, W.S. 1985: PROC CANDISC. In: SAS User's Guide: Statistics. Version 6. 4th Edition. SAS Institute, Inc. (Eds.), Cary, NC.

OTRAS REFERENCIAS ACONSEJADAS

- CUADRAS, C.M. 1981: Métodos de análisis multivariante. Ed. por Eunibar. Barcelona.

- DE PACE, C, A. FILIPPETTI and L. RICCIARDI. 1988: A Combined Univariate and Multivariate Approach for Selecting High Performing Genotypes of *Vicia faba* L. Plant. Breeding. 100, 199-208.

- DIGBY, P.G.N. and R.A. KEMPTON. 1987: Multivariate analysis of ecological communities. Chapman and Hall (Ed.)

- GHADERI, A., M.W. ADAMS and A.W. SAETTLER. 1982: Environmental response

patterns in commercial classes of common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. 63, 17-22.

- RAO, C.R. 1970: Multivariate analysis: An indispensable statistical aid in applied research. Sankhya 22, 317-338.

- WHITEHOUSE, R.N. 1970: Canonical analysis as an aid in plant breeding. pp. 269-282. In: R.A. Nilan (Ed.), Barley Genetics II. Washington State Univ. Press, Pullman, WA.



14

**MODELO SREG.
GGE BIPLLOT**

Σ α Π Ω \times μ ϑ

14. MODELO SREG. GGE BIPLLOT

Analizando el comportamiento de una serie de variedades de trigo de invierno procedentes de un ensayo realizado en varias localidades de Ontario, Yan *et al.* (2000) propusieron para dicho análisis, lo que ellos denominaron, el “GGE biplot”, biplot construido a partir de los dos primeros componentes principales (ACP1 y ACP2) extraídos de un análisis de componentes principales con los datos previamente centrados respecto las medias ambientales. El término “GGE biplot” procede de que se dibuja en el biplot tanto el efecto principal genotípico (G) como el de la interacción GxE, los cuales son, verdaderamente, las dos fuentes de variación más importantes, y por tanto, debemos considerarlas conjuntamente, en las evaluaciones del comportamiento de diversos genotipos en diferentes ambientes (Gauch y Zobel, 1996).

El biplot ya ha sido utilizado anteriormente por otros investigadores (Bradú y Gabriel, 1978; Cooper *et al.* 1997), pero no tal y como lo desarrollaron Yan *et al.* (2000) y Yan y Kang (2002), por medio del cual podemos obtener la siguiente información:

- a) Estudiar el comportamiento de los diferentes genotipos en un ambiente determinado.
- b) Determinar la relativa adaptación de un genotipo dado a los diferentes ambientes.
- c) Comparar cualquier pareja de genotipos en función de su comportamiento en los diferentes ambientes del ensayo.
- d) Determinar los mejores genotipos en cada ambiente.
- e) Determinar la existencia de mega-ambientes, en función del agrupamiento de ambientes producido en el biplot.
- f) Visualizar conjuntamente los efectos principales genotípicos y su estabilidad.

Aunque el método del “GGE biplot” fue desarrollado para el análisis de datos procedentes de ensayos multi-ambientales (EMA), puede aplicarse para todos los tipos de tablas de datos de dos vías, en los cuales se asume una estructura del tipo entrada x probador (siguiendo la nomenclatura de Yan *et al.* (2000): en los datos de un EMA, los genotipos serían las entradas y los ambientes serían los probadores).

Explicaremos brevemente el modelo SREG (“sites regresión”) y cómo generamos el “GGE biplot” correspondiente. El rendimiento medio de un genotipo *i* en el ambiente *j* se describe usualmente por el modelo lineal siguiente [68]:

$$Y_{ij} = \mu + g_i + e_j + gxe_{ij} \tag{68}$$

donde μ es la media general del ensayo, g_i es el efecto principal del genotipo *i*, e_j es el efecto principal del ambiente *j*, y gxe_{ij} es la interacción entre el genotipo *i* y el ambiente *j*. Suprimiendo del modelo g_i y/o e_j o todos los valores $\mu + g_i + e_j$, se consigue que la variación explicable por los términos suprimidos sea absorbida por el término gxe_{ij} . Es a esta matriz de valores gxe_{ij} la que sometemos a una descomposición singular (SVD). Precisamente, cuando aplicamos SVD a los valores gxe_{ij} en el modelo [68] tenemos el modelo AMMI (Zobel *et al.* 1988).

Cuando utilizamos el modelo de regresión de ambientes (SREG) obtenido a partir de suprimir el término g_i de la ecuación [68] y aplicar SVD a los valores gxe_{ij} (los cuales están ahora ambientalmente centrados) tenemos el siguiente modelo:

$$(g_i + gxe_{ij}) = Y_{ij} - \mu - e_j = \lambda_1 \xi_{i1} \eta_{j1} + \lambda_2 \xi_{i2} \eta_{j2} + \varepsilon_{ij} \tag{69}$$

donde Y_{ij} es el rendimiento medio del genotipo *i* en el ambiente *j*, e_j es el rendimiento medio del ambiente *j* frente a todos los genotipos, λ_1 y λ_2 son los valores singulares para ACP1 y ACP2, respectivamente, ξ_{i1} y ξ_{i2} son las coordenadas ACP1 y ACP2, respectivamente, para el genotipo *i*, η_{j1} y η_{j2} son las coordenadas ACP1 y ACP2, respectivamente, para el ambiente *j*, y ε_{ij} es el residual del modelo asociado con el genotipo *i* y el ambiente *j*.

Un biplot basado en la ecuación [69] contiene los efectos del genotipo G más los efectos de la interacción GxE, de ahí que se le haya denominado como “GGE biplot”; por el contrario, un biplot generado a partir de una descomposición singular de los valores gxe_{ij} procedentes del modelo [68] contiene sólo los efectos de la interacción GxE, de ahí que se le denomine “GE biplot”.

Para dibujar el GGE biplot de forma que las coordenadas ACP1 y ACP2 de genotipos y ambientes estén referidas a una escala simétrica,



podemos reescribir la ecuación anterior [69] en la forma siguiente:

$$Y_{ij} - \mu - e_j = \xi_{i1}^* \eta_{j1}^* + \xi_{i2}^* \eta_{j2}^* + \varepsilon_{ij} \quad (70)$$

donde $\xi_{i1}^* = \lambda^{0.5_1} \xi_{i1}$; $\xi_{i2}^* = \lambda^{0.5_2} \xi_{i2}$; $\eta_{j1}^* = \lambda^{0.5_1} \eta_{j1}$ y $\eta_{j2}^* = \lambda^{0.5_2} \eta_{j2}$. La escala del modelo [70] tiene la ventaja de que ACP1 y ACP2 tienen la misma unidad (raíz cuadrada de la unidad original, p. ej. t ha⁻¹ en términos de rendimiento).

El modelo SREG tiene una versión ambiental estandarizada:

$$(Y_{ij} - \mu - e_j)/z_j = \xi_{i1}^* \eta_{j1}^* + \xi_{i2}^* \eta_{j2}^* + \varepsilon_{ij} \quad (71)$$

donde z_j es:

- la desviación típica de las medias de los genotipos para el ambiente j , particularmente útil para analizar datos en los cuales los distintos probadores o "tester" están medidos en distintas unidades,
- el error estándar del genotipo medio para el ambiente j ; obviamente, este tipo de error sólo puede estimarse si nuestro diseño consta de varias repeticiones, por lo que este modelo sólo es posible cuando tenemos repeticiones en el ensayo, de forma que podamos ajustar cualquier heterogeneidad entre probadores.

Comentemos brevemente lo que podemos observar en un GGE biplot (Fig. 47). Se representan todos los genotipos por sus "marcas" (p. ej. G1) correspondientes; de ellos, los más alejados del origen se unen por una línea a trazos formando un polígono, por lo que el resto de genotipos se encontrarán en su interior. Se dibujan líneas que partiendo del origen sean perpendiculares a cada lado del polígono, de forma que el biplot estará dividido en varios sectores. La marca de cada ambiente se une con el origen mediante una línea, éstos quedaran repartidos entre los distintos sectores formados anteriormente. Los genotipos situados en los vértices del polígono serán los de mayor respuesta o mejor comportamiento en los ambientes incluidos en su sector (determinado por las líneas perpendiculares a cada lado que confluyen en dicho vértice). Una condición necesaria para poder identificar los mejores genotipos en función de sus efectos principales por medio del biplot, **es que las coordenadas ACP1 genotípicas tengan un coeficiente de correlación mayor de 0.95 con los efectos principales genotípicos**; generalmente, esto ocurre, cuando el primer eje del ACP extrae un gran porcentaje de la

variación total. De esta forma, un genotipo ideal sería aquel que tuviera grandes valores de la coordenada ACP1 (alto rendimiento medio, o sea, un alto efecto principal) y casi nulos de la coordenada ACP2 (más estabilidad). De la misma forma, un ambiente ideal sería aquel con una gran coordenada ACP1, o sea, capaz de discriminar entre los genotipos ensayados, es decir, facilitar la identificación de los genotipos con mejor adaptación general, y coordenada ACP2 casi nula, o sea, más representativo del ambiente medio. Si un ambiente tiene una gran coordenada ACP2 podría favorecer el comportamiento de algunos genotipos y al mismo tiempo, desfavorecer el de otros; así, un genotipo seleccionado bajo un ambiente con una alta coordenada ACP2 puede resultar altamente específico para ese ambiente.

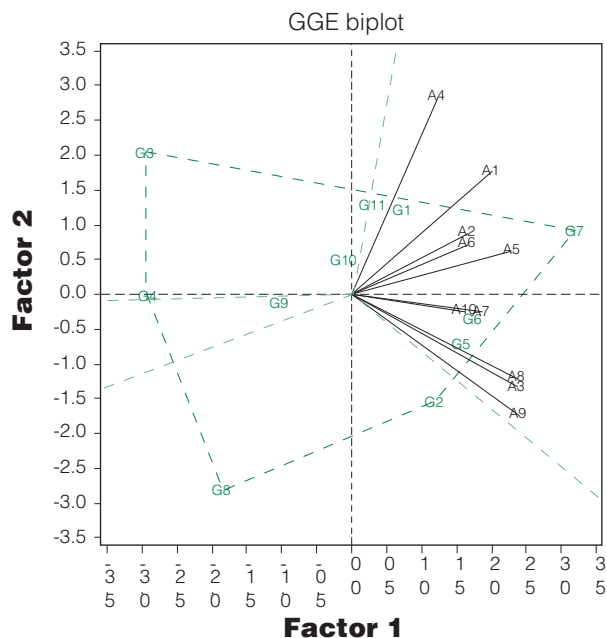
14.1. Caso práctico

Para ejemplo de éste modelo utilizaremos los datos del fichero MEDIAS.txt. Dado que se trata de un ensayo con repeticiones, seguiremos el modelo [71] en el cual Z_j es el error estándar del ambiente j . Utilizaremos la corriente de control específica para nuestros datos, que encontramos en el Apéndice de este Manual (Yan2000.sas), y necesitaremos, como software estadístico, el SAS. Una vez ejecutado el programa en SAS, nos generará el biplot de la Figura 47 y, la salida texto que podemos encontrar en el Apéndice.

En el GGE biplot de la Figura 47 podemos identificar los genotipos vértice como aquellos conectados por los lados del polígono. Estos genotipos son: **G7, G2, G8, G4 y G3**. Se han trazado las perpendiculares a cada lado del polígono dividiendo así el biplot en una serie de sectores, uno por cada genotipo vértice. Se demuestra que todos los ambientes incluidos en un sector dan el mismo genotipo como superior en cada ambiente, coincidiendo éste con el genotipo vértice.

En la Figura 47 vemos que todos los ambientes están incluidos en un único sector, cuyo genotipo vértice es G7, por lo cual podemos decir que el genotipo G7 ha sido el mejor en todos los ambientes del ensayo. El resto de genotipos vértice no han tenido ningún ambiente en sus sectores respectivos; lógicamente, no han sido los de mayor rendimiento en ninguno de los ambientes ensayados, ya que, como hemos visto anteriormente, el genotipo G7 ha sido el genotipo de mayor rendimiento en todos los ambientes. Los genotipos interiores al polígono, particularmente aquellos localizados cerca del origen (G10,G9), son los

Figura 47



genotipos de menor respuesta en los ambientes ensayados, siendo, por contra, los genotipos de los vértices los de mayor respuesta. Los genotipos vértice situados en los cuadrantes opuestos a aquellos en que se sitúan los ambientes, es decir, aquellos con coordenadas ACP1 de signo generalmente negativo, son realmente los genotipos de menor rendimiento medio. Dicho de otra forma, los que más **pobrementemente** han **rendido** en todos los ambientes; en nuestro caso, los genotipos **G3, G4 y G8**.

Estudiando simultáneamente rendimiento y estabilidad podemos comentar lo siguiente a la vista de la Figura 47. Vemos como el genotipo **G7** proporciona los **mayores rendimientos medios** (el mayor valor de la coordenada ACP1) y fue relativamente **estable** frente a todos los ambientes ensayados (pequeño valor absoluto de la coordenada ACP2); le siguen los genotipos **G6, G5 y G2** como un segundo grupo de genotipos con **alto rendimiento medio**, siendo G6 y G5 de similar o mayor estabilidad que G7 y el genotipo G2 ligeramente más inestable que los anteriores, probablemente, como veremos más adelante, porque en los ambientes A1 y A4 promedió por debajo de la media de cada ambiente. Los genotipos **G11 y G1** muestran similar rendimiento medio (**ligeramente superior a la media general** del ensayo, que vendría dada aproximadamente por el eje ACP2) y similar estabilidad a la mostrada por el genotipo G2. Por el contrario, vemos como el biplot nos muestra en el extremo opuesto tres genotipos no adaptados a los ambientes del ensayo (G8, G4 y G3), mostrando

rendimientos muy pobres en todos los ambientes; puede destacarse el genotipo G4 como el peor adaptado, dado que es el de menor rendimiento medio (la menor coordenada ACP1) y muy estable (baja coordenada ACP2). Finalmente, los rendimientos medios de los genotipos **G10 y G9** estuvieron **por debajo de la media general** del ensayo (ACP1<0) y ligeramente estables (pequeña coordenada ACP2).

El biplot también muestra el rendimiento de cualquier genotipo en cualquiera de los ambientes; por ejemplo, el genotipo G2 mostró un rendimiento medio por encima de la media general del ensayo (ACP1>0), pero en dos ambientes, A4 y A1, mostró un rendimiento inferior a la media de cada uno de estos ambientes. Según Yan *et al.* (2000) la forma de visualizar si un genotipo está por encima o por debajo de la media de un ambiente viene indicado por el ángulo virtual formado por las líneas que unen el origen con los respectivos genotipo y ambiente; así, si éste ángulo es obtuso dicho genotipo rinde por debajo de la media ambiental y si el ángulo es agudo el rendimiento está por encima.

Con respecto a los ambientes, el biplot muestra como **A8, A3 y A9** fueron los **ambientes más discriminantes**, tal y como indican sus mayores coordenadas ACP1. Sin embargo, debido a su gran coordenada ACP2, las diferencias observadas entre genotipos en estos ambientes puede no reflejar exactamente las diferencias en rendimiento medio de estos genotipos en todos los ambientes.

Por el contrario, los ambientes **A7 y A10** no fueron muy **discriminantes** pero las diferencias entre genotipos fueron altamente consistentes con los promedios obtenidos frente a todos los ambientes ensayados, ya que tomaron las menores coordenadas ACP2. En los ambientes con coordenadas ACP2 próximas a cero, los genotipos se ordenan básicamente de acuerdo con sus coordenadas ACP1 o a sus efectos principales (ya que, en nuestro caso, la correlación entre las coordenadas

ACP1 genotípicas y sus efectos principales fue alta, $r > 0.95$) y, por tanto, las diferencias entre genotipos vendrán dadas en proporción a sus efectos principales en cada ambiente. Por el contrario, ambientes como el **A4** no son **ni discriminantes** (pequeña coordenada ACP1) **ni representativos** (gran coordenada ACP2) y, por consiguiente, los genotipos que tuvieran altos rendimientos en A4 no necesariamente dan altos rendimientos medios en el conjunto de los ambientes ensayados.

REFERENCIAS CITADAS

BRADU, D. y K.R. GABRIEL. 1978. The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 20:47-68.

COOPER, M., R.E. STUCKER, I.H. DELACY y B.D. HARCH. 1997. Wheat breeding nurseries, target environments, and indirect selection for grain yield. *Crop Sci.* 37:1168-1176.

GAUCH, H.G. y R.W. ZOBEL. 1997. Identifying mega-environments and targeting genotypes. *Crop Sci.* 37:311-326.

YAN, W., HUNT, L.A., SHENG, Q. y Z. SZLAVNICS. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* 40:597-605.

YAN, W., y M.S. KANG. 2002. GGE biplot analysis. A graphical tool for breeders, geneticists, and agronomists. CRC PRESS.

ZOBEL, R.W., M.J. WRIGHT y H.G. GAUCH. 1988. Statistical analysis of a yield trial. *Agron. J.* 80:388-393.

15

**COMENTARIO Y
CONSEJOS FINALES**

15. COMENTARIOS Y CONSEJOS FINALES

Como se ha comentado a lo largo de este Manual, el obtener datos por medio de ensayos en muchas localidades permite estimar rendimientos con más precisión y permite seleccionar los mejores genotipos. La mayoría de los mejoradores y de los extensionistas poseen un buen conocimiento de los diferentes ambientes en que realizan sus ensayos, así como, en general, de la adaptación del material vegetal que manejan. Generalmente este conocimiento es empírico, habiéndose adquirido a partir de las observaciones de campo, incluso de simples comparaciones visuales entre líneas en selección y los testigos adecuados. Sin embargo, parece necesaria la utilización de métodos estadísticos para el cálculo de la estabilidad genotípica, complementando y haciendo objetivas las impresiones realizadas en campo por el mejorador. Dado el gran número de métodos existentes ha resultado necesaria una revisión de los más utilizados. Así, discutamos brevemente lo que en este Manual se ha comentado sobre los mismos.

15.1. Métodos Univariantes paramétricos

Los métodos univariantes para calcular estabilidad tienen la ventaja de su simplicidad en los cálculos.

En la Introducción ya se comentó que los parámetros S^2_{xi} y CV_i pertenecen al tipo 1 de estabilidad análogo al concepto de homeostasis. Una razón por la cual los mejoradores no han utilizado mucho estos parámetros es por la necesidad de encontrar genotipos no sólo con buena estabilidad de tipo 1 sino también con alto rendimiento. La estabilidad de tipo 1 está asociada generalmente con respuestas pobres y bajos rendimientos en ambientes en los que otros cultivares han tenido altos rendimientos. En nuestro caso práctico, los genotipos que han mostrado mayor estabilidad medida por medio del parámetro S^2_{xi} han resultado ser también los de menor rendimiento medio. En cambio para el CV_i los genotipos de mayor estabilidad también eran los de mayor rendimiento; sin embargo, el estadístico CV_i no resulta "a priori" muy apropiado como medida de estabilidad porque tiene una limitación intrínseca: bajos rendimientos implican altos CV_i , y altos rendimientos, bajos CV_i .

Para genetistas, mejoradores y agrónomos, los estadísticos paramétricos de estabilidad obtenidos por **regresión lineal** son matemáticamente muy simples y fácilmente interpretables desde un punto de vista biológico. Sin embargo, como se ha expuesto en este Manual, este método tiene graves inconvenientes, que podemos resumir fundamentalmente en tres:

- No suministra información cuando falta la linealidad,
- Es muy dependiente del grupo de genotipos y ambientes incluidos en el ensayo, y
- Tiende a simplificar sobremanera los diferentes patrones de respuesta para explicar la interacción **GxE** en una sola dimensión (coeficiente de regresión), cuando en realidad ésta puede ser muy compleja. **Evidentemente resulta peligroso tratar de sacrificar información relevante con objeto de facilitar la interpretación biológica y la estadística.**

Dado que para nuestro caso práctico la heterogeneidad de las regresiones ha resultado no significativa, la utilidad del método de análisis de regresión para la interpretación de los datos resulta, por lo menos, dudosa. En línea con nuestro resultado, Westcott llegó a afirmar que no deben recomendarse ni el método de regresión ni los parámetros de estabilidad deducidos de él, ni se pueden superar sus defectos con la utilización conjunta del análisis de grupos o el ACP (Westcott, 1986).

A pesar de todo lo expuesto anteriormente, conviene decir que **cuando la heterogeneidad de las regresiones resulte altamente significativa** el método de regresión puede ser utilizado. Se aconseja su utilización también cuando en el ensayo experimental intervienen un gran número de genotipos y se desea realizar un barrido previo de los mismos.

Los parámetros W_i , $W_i\text{-CM}$ y σ^2_i han producido resultados similares entre sí; de esta forma podemos usar indistintamente uno u otro, siendo preferible utilizar σ^2_i , ya que como se indica en el Capítulo 3, una exhaustiva investigación de los mismos reveló que la ecovalencia (W_i) era un paso intermedio para el cálculo de σ^2_i y que no es necesario el uso de W_i como una medida individual de estabilidad dado que σ^2_i es un parámetro insesgado cuya significación estadística



puede calcularse. El estadístico **W_i-CM** también puede ser probado estadísticamente, pero suponiendo que se distribuyera como una **F** de Snedecor, hipótesis que necesita ser probada.

El ordenamiento que produce **P_i** es el mismo que el originado por el rendimiento medio. Por tanto, la selección basada sobre este índice podría favorecer el seleccionar fundamentalmente por rendimiento, por lo que **P_i** no ofrece ninguna ventaja respecto de los métodos anteriores.

15.2. Métodos univariantes no-paramétricos

Si se está interesado en considerar simultáneamente estabilidad y rendimiento, los estadísticos no paramétricos a aplicar son **S⁽³⁾_i** y **S⁽⁶⁾_i** de Hühn, los cuales miden la estabilidad en unidades de rendimiento. Pero deben utilizarse los datos originales porque la transformación $Y_{ij} - Y_{ij}^*$ elimina el efecto genotípico de los datos.

El método de **Ketata** tiene la ventaja de su sencillez en los cálculos y su poder resolutivo en las gráficas. En éstas ya podemos estimar rápidamente qué genotipos son más estables y de mayor rendimiento. Nos combina, pues, en una gráfica, estabilidad y rendimiento.

El método de **Kang** ha resultado ser también un método útil para seleccionar simultáneamente para rendimiento y estabilidad. Además, si estuviéramos interesados en dar más peso a rendimiento o a estabilidad también se podría utilizar.

El método de rangos estratificados de **Fox** resulta simple y sencillo y, como todos los métodos no paramétricos, es independiente de la escala, además de dar igual peso a cada ambiente.

15.3. Otros métodos univariantes

El método de **Eskridge** tiene como principal preocupación eliminar todo tipo de riesgos para la elección de aquel cultivar que tenga poca probabilidad en dar rendimientos bajos. Ahora bien, al estar los índices basados en varianzas estimadas, las cuales pueden mostrar por sí mismas una gran error estándar, las estimas imprecisas de varianzas pueden producir ordenaciones de genotipos de dudoso valor.

El método **Estrella** posee la ventaja de su sencillez y rapidez de resultados que nos determina con una simple visualización aquellos genotipos más estables y de mayor rendimiento, pero no suministra ningún índice cualitativamente diferente a los ya comentados.

15.4. Métodos Multivariantes

Se ha utilizado en el presente Manual un amplio número de métodos multivariantes para analizar los ensayos multilocales y estimar la estabilidad del rendimiento. En algunos de ellos se superan las limitaciones de la regresión lineal, pero los resultados son, a menudo, difíciles de interpretar en relación a la interacción **GxE** (como es el caso del **Análisis de Componentes Principales** y el **Análisis de Grupos**).

Para nuestro ejemplo práctico, el **Análisis de Componentes Principales**, por sí sólo, no nos ha suministrado unos resultados que hayamos podido relacionar con el comportamiento de los genotipos en los distintos ambientes: en otros casos reales puede ocurrir lo contrario, pero ha de considerarse siempre que no es un método diseñado para el estudio de la interacción GxE.

El **Análisis de Grupos** de Lin, para nuestro caso práctico, ha mostrado ser más eficaz que el de Fox y Rosielle para el estudio de la interacción GxE. Pero recordemos que en los últimos años el más utilizado viene siendo el segundo y que ha dado, en general, excelentes resultados.

Ciertas técnicas multivariantes, o combinaciones de las mismas, ofrecen una información biológica relevante y son estadísticamente muy simples como, por ejemplo, el **AMMI**. La combinación del Análisis de Varianza y Componentes Principales en el modelo **AMMI**, junto con las estimas predictivas, hacen que éste sea un método muy válido para explicar la estructura de la interacción y estimar con más precisión los rendimientos, pero se necesita una investigación más profunda para cuantificar la probabilidad de acertar en la selección de un genotipo utilizando el valor **predictivo** del **AMMI**, comparado con los valores **predictivos** de otros modelos. Es, ciertamente, el método más completo en cuanto a información suministrada, aunque también el de análisis más complejo.



El **Análisis de Coordenadas Principales** (CCPP), con la medida de similaridad de Westcott, se ha mostrado como una potente herramienta para la determinación de aquellos genotipos más ampliamente adaptados y de alto rendimiento.

El **Análisis Factorial de Correspondencias** (AFC) se muestra potencialmente muy adecuado para el estudio de la interacción **GxE** pues permite visualizar de forma gráfica la presencia o ausencia de interacción **GxE**. Por otro lado, el **AFC** permite efectuar el estudio particular de cada uno de los genotipos, pudiendo extraer más información que con el **Análisis de Grupos** o en los casos en que el **Análisis de CCPP** no ofrece conclusiones claras.

El **Análisis Canónico** es un método muy adecuado para **confirmar** que la clasificación obtenida por otros métodos tiene un mínimo de arbitrariedad.

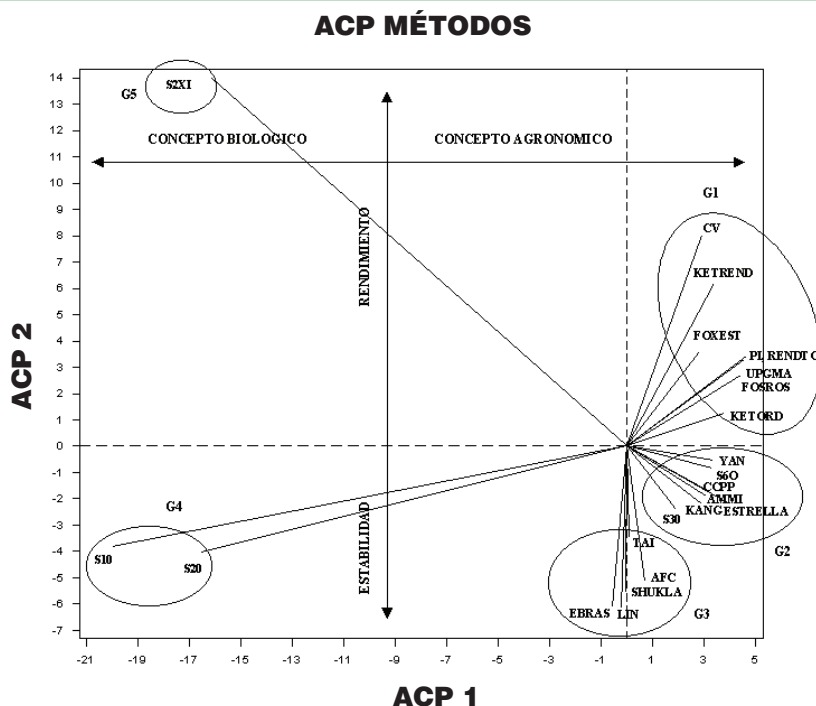
El modelo SREG del **GGE biplot** es otra técnica multivariante, híbrida como el AMMI, en la

cual combinamos un análisis de regresión y un análisis de componentes principales. Resulta, ciertamente, más complejo interpretar sus resultados que los obtenidos por el AMMI, pero renombrados biométricos han demostrado que con este método de estudio de la **GxE** conseguimos un nivel mayor de información y más minucioso que el aportado por el modelo AMMI. Con ello no queremos indicar que sean métodos excluyentes, sino todo lo contrario, pensemos que cuando encontremos una correlación baja entre los efectos aditivos genotípicos y las coordenadas ACP1 los efectos primarios o principales no quedan correctamente reflejados en el biplot, por lo que sería más razonable combinar ambos modelos.

15.5. Comparación de métodos univariantes y multivariantes

Lo que en este apartado vamos a comentar es un extracto de lo expuesto en el artículo Flores et al. (1998), ampliado con los resultados obtenidos al aplicar el método de Yan *et al.* (2000) al mismo conjunto de datos.

Figura 48 S2XI: Varianza ambiental, EBRAS: regresión según Eberhart y Russell, TAI: método de Tai, SHUKLA: varianza de estabilidad de Shukla, CV: Coeficiente de variación de Francis y Kannenberg, PI: parámetro de Lin y Binns, S10, S20, S30 y S60: medidas de estabilidad no-paramétrica de Hühn, KANG: método de sumas-rangos de Kang, KETRANG: método de Ketata con rangos, KETREND: método de Ketata con rendimiento, FOXRANG: método estratificado de rangos de Fox, ESTRELLA: método estrella de Flores, UPGMA: análisis de grupos usando el algoritmo UPGMA, LIN: análisis de grupos según Lin, FOXROS: análisis de grupos según Fox y Rosielle, AMMI: análisis AMMI según Zobel, CCPP: coordenadas principales según Westcott, AFC: análisis factorial de correspondencias según Hill, RENDTO: orden genotípico según su rendimiento, YAN: análisis según GGE biplot de Yan.



Se compararon veintitrés métodos diferentes (paramétricos, no-paramétricos y multivariantes) utilizados en este Manual para el estudio de la interacción GxE utilizando dos grupos de datos experimentales: un ensayo de 15 variedades de habas en 12 ambientes y un ensayo de 11 variedades de guisante en 16 ambientes. Se formaron las matrices de rangos de cada variedad en cada ambiente; se aplicó a cada matriz de rangos un análisis de componentes principales y se compararon los resultados mediante un análisis de correlación canónica. Los resultados se muestran en la Figura 48. La inclusión del método de Yan *et al.* (2000) no modificó las conclusiones de Flores *et al.* (1998).

Se identificaron cinco grupos de métodos muy diferentes en base a distintos conceptos de estabilidad. El primer eje del análisis de componentes principales separó los métodos que definen la **estabilidad homeostática** de un genotipo (S2XI, S1O y S2O) del resto de métodos que se basan en un **concepto agrónomico de estabilidad**.

El segundo componente principal nos ordena los métodos en tres grupos fundamentales, de tal forma que si un mejorador está interesado principalmente en aquellos métodos que le dan más importancia al *rendimiento genotípico* tendría que elegir uno de los métodos del grupo G1 (CV, KETREND, PI, UPGMA, FOXROS, FOXEST y KETORD); si la selección de genotipos la basamos dándole más importancia a su *estabilidad* que al rendimiento, entonces los métodos a elegir son los agrupados en G3 (TAI, AFC, SHUKLA, LIN y EBRAS) y, finalmente, si estamos interesados en seleccionar bajo un criterio que le de igual importancia a la estabilidad de los genotipos que a su nivel de rendimiento, entonces escogeríamos los métodos incluidos en el grupo G2 (YAN, CCPP, S6O, AMMI, ESTRELLA, KANG y S3O).

Por otro lado, para genéticos, mejoradores y agrónomos, los métodos univariantes son matemáticamente más simples y de más fácil interpretación. Estos métodos, sin embargo, tienen una gran desventaja: tienden a sobresimplificar el modelo de respuesta para explicar en una

sola dimensión la variación debida a la interacción, cuando, en realidad, la respuesta normalmente es multidimensional. Existe, pues, el peligro de sacrificar información relevante con el objeto de simplificar la interpretación estadística. Estos métodos tan simples podrían ser útiles cuando el mejorador tiene un gran número de genotipos para evaluar, pero cuando sea necesario un estudio más preciso (por ejemplo, en los análisis de los datos procedentes de ensayos llevados a cabo por Centros Internacionales de Investigación) los métodos multivariantes son los idóneos.

15.6. Plan de Trabajo

Así a la vista de este resumen, un **plan de trabajo** que recomendamos para un buen estudio de la interacción GxE es la aplicación sucesiva de los siguientes métodos estadísticos; tal plan es similar al de la primera edición con la inclusión del método del "GGE biplot":

- 1º. El método del GGE biplot.
- 2º. El análisis AMMI.
- 3º. El análisis canónico.
- 4º. Un método univariante (Kang).

- 1º. El **Método del "GGE biplot"** que es actualmente el método más potente para el estudio de la interacción GxE.
- 2º- El **Análisis AMMI**, que nos suministraría información adicional, sobre todo en aquellos casos de baja correlación entre efecto principales y las coordenadas ACP1 del modelo anterior.
- 3º. El **Análisis Canónico**, para rechazar toda arbitrariedad en la formación de los grupos por los métodos anteriores.
- 4º- Un **Método Univariante**, por ejemplo el método de Kang (1988), que nos ayudaría a perfilar la ordenación dentro de grupos.
- 5º. Recordemos que cuando en el ensayo intervienen un gran número de genotipos es muy conveniente hacer barridos previos con métodos univariantes antes de aplicar los multivariantes.

REFERENCIAS CITADAS

- FLORES, F, M.T. MORENO Y J.I. CUBERO. 1998. A comparison of univariate and multivariate methods to analyze GxE interaction. *Field Crops Res.* 56: 271-286.
- KANG, M.S. 1988: A Rank-sum method for selectig high-yielding, stable corn genotypes. *Cereal Res. Comm.* 16, 113-115.
- WESTCOTT, B. 1986: Some methods of analysing genotype-environment interaction. *Heredity* 56, 243-253.
- YAN, W., HUNT, L.A., SHENG, Q. Y Z. SZLAVNICS. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* 40:597-605.

 Σ α Π Ω \times \pm ϑ



16

Σ

α

Π

Ω

APÉNDICE

ϑ

16. APÉNDICE

Nota Previa

Se advierte que los programas en lenguaje BASIC han sido realizados específicamente para el caso práctico que se ha resuelto en este Manual, aunque si se poseen ciertos conocimientos del lenguaje BASIC pueden ser adaptados fácilmente para la resolución de otros casos.

16.1. Matrices de entrada para los distintos paquetes estadísticos utilizados

16.1.1. Matriz de datos denominada "DAT-COMP1.TXT" utilizada para las corrientes de control diseñadas para el paquete estadístico SAS.

AMBIENTE	GENOTIPO	REPET.	RENDIMIENTO
1.0	1.0	1.0	4208.3
1.0	1.0	2.0	3933.3
1.0	1.0	3.0	3408.3
1.0	2.0	1.0	3458.3
1.0	2.0	2.0	3679.2
1.0	2.0	3.0	3520.8
1.0	3.0	1.0	3475.0
1.0	3.0	2.0	3308.3
1.0	3.0	3.0	3275.0
1.0	4.0	1.0	3475.0
1.0	4.0	2.0	3216.7
1.0	4.0	3.0	3312.5
1.0	5.0	1.0	3820.8
1.0	5.0	2.0	3395.8
1.0	5.0	3.0	4070.8
1.0	6.0	1.0	4000.0
1.0	6.0	2.0	4195.8
1.0	6.0	3.0	4245.8
1.0	7.0	1.0	4166.7
1.0	7.0	2.0	3720.8
1.0	7.0	3.0	4225.0
1.0	8.0	1.0	3016.7
1.0	8.0	2.0	2716.7
1.0	8.0	3.0	2437.5
1.0	9.0	1.0	3504.2
1.0	9.0	2.0	3775.0
1.0	9.0	3.0	3637.5
1.0	10.0	1.0	3800.0
1.0	10.0	2.0	3804.2
1.0	10.0	3.0	3395.8
1.0	11.0	1.0	3504.2
1.0	11.0	2.0	3870.8
1.0	11.0	3.0	3470.8
2.0	1.0	1.0	4166.7
2.0	1.0	2.0	3125.0
2.0	1.0	3.0	3333.3
2.0	2.0	1.0	3958.3
2.0	2.0	2.0	3541.7
2.0	2.0	3.0	3750.0
2.0	3.0	1.0	3958.3
2.0	3.0	2.0	3958.3
2.0	3.0	3.0	3541.7
2.0	4.0	1.0	2500.0
2.0	4.0	2.0	2916.7
2.0	4.0	3.0	3125.0
2.0	5.0	1.0	4375.0
2.0	5.0	2.0	4375.0
2.0	5.0	3.0	4375.0
2.0	6.0	1.0	3750.0
2.0	6.0	2.0	3541.7
2.0	6.0	3.0	4166.7
2.0	7.0	1.0	3541.7
2.0	7.0	2.0	4375.0
2.0	7.0	3.0	4166.7
2.0	8.0	1.0	3541.7
2.0	8.0	2.0	3125.0
2.0	8.0	3.0	3125.0
2.0	9.0	1.0	2916.7

(Continúa)



Σ

α

Π

Ω

\times

\pm

ϑ

AMBIENTE	GENOTIPO	REPET.	RENDIMIENTO
2.0	9.0	2.0	3750.0
2.0	9.0	3.0	3541.7
2.0	10.0	1.0	3541.7
2.0	10.0	2.0	3750.0
2.0	10.0	3.0	3750.0
2.0	11.0	1.0	4166.7
2.0	11.0	2.0	3541.7
2.0	11.0	3.0	4375.0
3.0	1.0	1.0	3454.2
3.0	1.0	2.0	3087.5
3.0	1.0	3.0	3383.3
3.0	2.0	1.0	4000.0
3.0	2.0	2.0	3695.8
3.0	2.0	3.0	3679.2
3.0	3.0	1.0	2637.5
3.0	3.0	2.0	2575.0
3.0	3.0	3.0	1770.8
3.0	4.0	1.0	2920.8
3.0	4.0	2.0	2437.5
3.0	4.0	3.0	2804.2
3.0	5.0	1.0	3800.0
3.0	5.0	2.0	3766.7
3.0	5.0	3.0	3808.3
3.0	6.0	1.0	3950.0
3.0	6.0	2.0	3758.3
3.0	6.0	3.0	3529.2
3.0	7.0	1.0	3975.0
3.0	7.0	2.0	4054.2
3.0	7.0	3.0	4083.3
3.0	8.0	1.0	3437.5
3.0	8.0	2.0	3375.0
3.0	8.0	3.0	2720.8
3.0	9.0	1.0	3083.3
3.0	9.0	2.0	3304.2
3.0	9.0	3.0	2695.8
3.0	10.0	1.0	3645.8
3.0	10.0	2.0	3662.5
3.0	10.0	3.0	3341.7
3.0	11.0	1.0	3358.3
3.0	11.0	2.0	2162.5
3.0	11.0	3.0	3075.0
4.0	1.0	1.0	3825.0
4.0	1.0	2.0	3550.0
4.0	1.0	3.0	4229.2
4.0	2.0	1.0	3425.0
4.0	2.0	2.0	2945.8
4.0	2.0	3.0	3391.7
4.0	3.0	1.0	3475.0
4.0	3.0	2.0	2925.0
4.0	3.0	3.0	4241.7
4.0	4.0	1.0	3229.2
4.0	4.0	2.0	3062.5
4.0	4.0	3.0	3237.5
4.0	5.0	1.0	3141.7
4.0	5.0	2.0	3287.5
4.0	5.0	3.0	3500.0
4.0	6.0	1.0	3116.7
4.0	6.0	2.0	3450.0
4.0	6.0	3.0	3225.0
4.0	7.0	1.0	3654.2
4.0	7.0	2.0	4225.0
4.0	7.0	3.0	4512.5
4.0	8.0	1.0	2329.2
4.0	8.0	2.0	3004.2
4.0	8.0	3.0	2650.0
4.0	9.0	1.0	3054.2
4.0	9.0	2.0	3350.0
4.0	9.0	3.0	3058.3
4.0	10.0	1.0	3433.3
4.0	10.0	2.0	3904.2
4.0	10.0	3.0	3737.5
4.0	11.0	1.0	3458.3
4.0	11.0	2.0	3662.5
4.0	11.0	3.0	4008.3
5.0	1.0	1.0	2856.3
5.0	1.0	2.0	2562.5
5.0	1.0	3.0	3875.0
5.0	2.0	1.0	3162.5
5.0	2.0	2.0	3245.8

(Continúa)

AMBIENTE	GENOTIPO	REPET.	RENDIMIENTO
5.0	2.0	3.0	3389.6
5.0	3.0	1.0	2904.2
5.0	3.0	2.0	1904.2
5.0	3.0	3.0	1675.0
5.0	4.0	1.0	2220.8
5.0	4.0	2.0	1989.6
5.0	4.0	3.0	1354.2
5.0	5.0	1.0	3504.2
5.0	5.0	2.0	3100.0
5.0	5.0	3.0	2375.0
5.0	6.0	1.0	3462.5
5.0	6.0	2.0	3293.8
5.0	6.0	3.0	2629.2
5.0	7.0	1.0	3735.4
5.0	7.0	2.0	3675.0
5.0	7.0	3.0	4014.6
5.0	8.0	1.0	1310.4
5.0	8.0	2.0	3237.5
5.0	8.0	3.0	1608.3
5.0	9.0	1.0	2212.5
5.0	9.0	2.0	2683.3
5.0	9.0	3.0	2358.3
5.0	10.0	1.0	2672.9
5.0	10.0	2.0	2625.0
5.0	10.0	3.0	2775.0
5.0	11.0	1.0	2731.3
5.0	11.0	2.0	2904.2
5.0	11.0	3.0	3579.2
6.0	1.0	1.0	3958.3
6.0	1.0	2.0	3875.0
6.0	1.0	3.0	4083.3
6.0	2.0	1.0	3250.0
6.0	2.0	2.0	3916.7
6.0	2.0	3.0	2958.3
6.0	3.0	1.0	2666.7
6.0	3.0	2.0	3125.0
6.0	3.0	3.0	3208.3
6.0	4.0	1.0	3958.3
6.0	4.0	2.0	3125.0
6.0	4.0	3.0	3208.3
6.0	5.0	1.0	3541.7
6.0	5.0	2.0	3583.3
6.0	5.0	3.0	3166.7
6.0	6.0	1.0	3958.3
6.0	6.0	2.0	4000.0
6.0	6.0	3.0	3750.0
6.0	7.0	1.0	4375.0
6.0	7.0	2.0	4083.3
6.0	7.0	3.0	4000.0
6.0	8.0	1.0	3208.3
6.0	8.0	2.0	3708.3
6.0	8.0	3.0	3041.7
6.0	9.0	1.0	4166.7
6.0	9.0	2.0	3041.7
6.0	9.0	3.0	3041.7
6.0	10.0	1.0	3333.3
6.0	10.0	2.0	3458.3
6.0	10.0	3.0	3333.3
6.0	11.0	1.0	4166.7
6.0	11.0	2.0	3958.3
6.0	11.0	3.0	3291.7
7.0	1.0	1.0	3791.7
7.0	1.0	2.0	3854.2
7.0	1.0	3.0	4208.3
7.0	2.0	1.0	3958.3
7.0	2.0	2.0	4229.2
7.0	2.0	3.0	4416.7
7.0	3.0	1.0	3541.7
7.0	3.0	2.0	3041.7
7.0	3.0	3.0	3708.3
7.0	4.0	1.0	3125.0
7.0	4.0	2.0	3291.7
7.0	4.0	3.0	2875.0
7.0	5.0	1.0	4687.5
7.0	5.0	2.0	3541.7
7.0	5.0	3.0	3437.5
7.0	6.0	1.0	3958.3
7.0	6.0	2.0	3875.0
7.0	6.0	3.0	3854.2

(Continúa)



Σ

α

Π

Ω

\times

\pm

ϑ

AMBIENTE	GENOTIPO	REPET.	RENDIMIENTO
7.0	7.0	1.0	4625.0
7.0	7.0	2.0	3895.8
7.0	7.0	3.0	4166.7
7.0	8.0	1.0	4125.0
7.0	8.0	2.0	3708.3
7.0	8.0	3.0	3333.3
7.0	9.0	1.0	4166.7
7.0	9.0	2.0	3479.2
7.0	9.0	3.0	2958.3
7.0	10.0	1.0	3479.2
7.0	10.0	2.0	3291.7
7.0	10.0	3.0	4166.7
7.0	11.0	1.0	4062.5
7.0	11.0	2.0	3895.8
7.0	11.0	3.0	4125.0
8.0	1.0	1.0	5104.2
8.0	1.0	2.0	3958.3
8.0	1.0	3.0	3958.3
8.0	2.0	1.0	5729.2
8.0	2.0	2.0	4895.8
8.0	2.0	3.0	5000.0
8.0	3.0	1.0	3333.3
8.0	3.0	2.0	2916.7
8.0	3.0	3.0	2916.7
8.0	4.0	1.0	4375.0
8.0	4.0	2.0	3541.7
8.0	4.0	3.0	3229.2
8.0	5.0	1.0	5520.8
8.0	5.0	2.0	5104.2
8.0	5.0	3.0	4895.8
8.0	6.0	1.0	5625.0
8.0	6.0	2.0	5000.0
8.0	6.0	3.0	5208.3
8.0	7.0	1.0	6145.8
8.0	7.0	2.0	5416.7
8.0	7.0	3.0	4687.5
8.0	8.0	1.0	3750.0
8.0	8.0	2.0	4687.5
8.0	8.0	3.0	3854.2
8.0	9.0	1.0	4270.8
8.0	9.0	2.0	4479.2
8.0	9.0	3.0	3125.0
8.0	10.0	1.0	4895.8
8.0	10.0	2.0	3854.2
8.0	10.0	3.0	3958.3
8.0	11.0	1.0	4375.0
8.0	11.0	2.0	4062.5
8.0	11.0	3.0	4062.5
9.0	1.0	1.0	4704.2
9.0	1.0	2.0	4287.5
9.0	1.0	3.0	4491.7
9.0	2.0	1.0	6016.7
9.0	2.0	2.0	4925.0
9.0	2.0	3.0	5216.7
9.0	3.0	1.0	1816.7
9.0	3.0	2.0	2470.8
9.0	3.0	3.0	1600.0
9.0	4.0	1.0	2975.0
9.0	4.0	2.0	2341.7
9.0	4.0	3.0	3720.8
9.0	5.0	1.0	5645.8
9.0	5.0	2.0	4287.5
9.0	5.0	3.0	4883.3
9.0	6.0	1.0	5366.7
9.0	6.0	2.0	4841.7
9.0	6.0	3.0	4816.7
9.0	7.0	1.0	6016.7
9.0	7.0	2.0	5250.0
9.0	7.0	3.0	5070.8
9.0	8.0	1.0	4404.2
9.0	8.0	2.0	4508.3
9.0	8.0	3.0	4545.8
9.0	9.0	1.0	4000.0
9.0	9.0	2.0	4525.0
9.0	9.0	3.0	2854.2
9.0	10.0	1.0	4650.0
9.0	10.0	2.0	3325.0
9.0	10.0	3.0	4141.7
9.0	11.0	1.0	4675.0

(Continúa)

AMBIENTE	GENOTIPO	REPET.	RENDIMIENTO
9.0	11.0	2.0	4100.0
9.0	11.0	3.0	4075.0
10.0	1.0	1.0	3541.7
10.0	1.0	2.0	2187.5
10.0	1.0	3.0	1979.2
10.0	2.0	1.0	3125.0
10.0	2.0	2.0	2187.5
10.0	2.0	3.0	2083.3
10.0	3.0	1.0	2395.8
10.0	3.0	2.0	1979.2
10.0	3.0	3.0	1979.2
10.0	4.0	1.0	2708.3
10.0	4.0	2.0	1666.7
10.0	4.0	3.0	1666.7
10.0	5.0	1.0	3750.0
10.0	5.0	2.0	3125.0
10.0	5.0	3.0	3020.8
10.0	6.0	1.0	2812.5
10.0	6.0	2.0	3020.8
10.0	6.0	3.0	2812.5
10.0	7.0	1.0	3645.8
10.0	7.0	2.0	3125.0
10.0	7.0	3.0	2604.2
10.0	8.0	1.0	2916.7
10.0	8.0	2.0	2291.7
10.0	8.0	3.0	2083.3
10.0	9.0	1.0	3645.8
10.0	9.0	2.0	2812.5
10.0	9.0	3.0	2125.0
10.0	10.0	1.0	2500.0
10.0	10.0	2.0	2500.0
10.0	10.0	3.0	2812.5
10.0	11.0	1.0	2708.3
10.0	11.0	2.0	2812.5
10.0	11.0	3.0	2291.7

(Conclusión)

16.1.2. Matriz de entrada para el paquete estadístico NTSYS. Se trata de una matriz de tipo rectangular (1, para el Ntsys), con 10 filas y 11 columnas etiquetadas (11L, para el Ntsys) con

los nombres de G1, G2, ..., G11 (representando los 11 genotipos). Los datos son las medias de las tres repeticiones.

1 G1	10 G2	11L G3	0 G4	G5	G6	G7	G8	G9	G10	G11
3849.97	3552.77	3352.77	3334.73	3762.47	4147.20	4037.50	2723.63	3638.90	3666.67	3615.27
3541.67	3750.00	3819.43	2847.23	4375.00	3819.47	4027.80	3263.90	3402.80	3680.57	4027.80
3308.33	3791.67	2327.77	2720.83	3791.67	3745.83	4037.50	3177.77	3027.77	3550.00	2865.27
3868.07	3254.17	3547.23	3176.40	3309.73	3263.90	4130.57	2661.13	3154.17	3691.67	3709.70
3097.93	3265.97	2161.13	1854.87	2993.07	3128.50	3808.33	2052.07	2418.03	2690.97	3071.57
3972.20	3375.00	3000.00	3430.53	3430.57	3902.77	4152.77	3319.43	3416.70	3374.97	3805.57
3951.40	4201.40	3430.57	3097.23	3888.90	3895.83	4229.17	3722.20	3534.73	3645.87	4027.77
4340.27	5208.33	3055.57	3715.30	5173.60	5277.77	5416.67	4097.23	3958.33	4236.10	4166.67
4494.47	5386.13	1962.50	3012.50	4938.87	5008.37	5445.83	4486.10	3793.07	4038.90	4283.33
2569.47	2465.27	2118.07	2013.90	3298.60	2881.93	3125.00	2430.57	2861.10	2604.17	2604.17



16.1.3. Matriz de entrada para el NTSYS. Se trata de una matriz de tipo rectangular (1) de 11

filas por 10 columnas etiquetadas por A1, A2,, A10 (representando los 10 ambientes).

1	11	10L	0							
A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	
3849.97	3541.67	3308.33	3868.07	3097.93	3972.20	3951.40	4340.27	4494.47	2569.47	
3552.77	3750.00	3791.67	3254.17	3265.97	3375.00	4201.40	5208.33	5386.13	2465.27	
3352.77	3819.43	2327.77	3547.23	2161.13	3000.00	3430.57	3055.57	1962.50	2118.07	
3334.73	2847.23	2720.83	3176.40	1854.87	3430.53	3097.23	3715.30	3012.50	2013.90	
3762.47	4375.00	3791.67	3309.73	2993.07	3430.57	3888.90	5173.60	4938.87	3298.60	
4147.20	3819.47	3745.83	3263.90	3128.50	3902.77	3895.83	5277.77	5008.37	2881.93	
4037.50	4027.80	4037.50	4130.57	3808.33	4152.77	4229.17	5416.67	5445.83	3125.00	
2723.63	3263.90	3177.77	2661.13	2052.07	3319.43	3722.20	4097.23	4486.10	2430.57	
3638.90	3402.80	3027.77	3154.17	2418.03	3416.70	3534.73	3958.33	3793.07	2861.10	
3666.67	3680.57	3550.00	3691.67	2690.97	3374.97	3645.87	4236.10	4038.90	2604.17	
3615.27	4027.80	2865.27	3709.70	3071.57	3805.57	4027.77	4166.67	4283.33	2604.17	



16.1.4. Matriz **INT** con los valores (ge)_{ij} de la Interacción.

1	10	11L	0							
G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
90.100	-332.79	414.77	353.89	-194.27	179.55	-264.10	-530.26	257.85	88.190	-62.933
-297.64	-215.00	802.00	-213.05	338.83	-227.62	-353.24	-69.430	-57.686	22.654	270.16
-148.13	209.51	-306.82	43.397	138.34	81.594	39.304	227.28	-49.876	274.93	-509.53
282.30	-457.29	783.34	369.66	-472.90	-529.64	3.0674	-418.66	-52.779	287.29	205.60
168.92	211.27	53.999	-295.12	-132.81	-8.2881	337.59	-370.97	-132.16	-56.652	224.23
257.91	-464.98	107.59	495.27	-480.59	-19.301	-103.26	111.12	81.230	-157.93	172.95
14.878	139.19	315.92	-60.264	-244.49	-248.47	-249.09	291.65	-22.970	-109.26	172.91
-234.51	507.87	-697.33	-80.449	401.96	495.21	300.16	28.434	-237.62	-157.28	-326.44
82.945	848.92	-1627.1	-620.00	330.48	389.07	492.58	580.55	-239.64	-191.23	-46.521
-216.80	-446.69	153.68	6.6602	315.47	-112.11	-203.00	150.28	453.65	-0.7063	-100.43



16.1.5. Matriz **INTAMB** de 10x12, donde la primera columna es la media total de cada ambiente.

3607.4	90.100	-332.79	414.77	353.89	-194.27	179.55	-264.10	-530.26	257.85	88.190	-62.933
3686.9	-297.64	-215.00	802.00	-213.05	338.83	-227.62	-353.24	-69.430	-57.686	22.654	270.16
3304.0	-148.13	209.51	-306.82	43.397	138.34	81.594	39.304	227.28	-49.876	274.93	-509.53
3433.3	282.30	-457.29	783.34	369.66	-472.90	-529.64	3.0674	-418.66	-52.779	287.29	205.60
2776.6	168.92	211.27	53.999	-295.12	-132.81	-8.2881	337.59	-370.97	-132.16	-56.652	224.23
3561.9	257.91	-464.98	107.59	495.27	-480.59	-19.301	-103.26	111.12	81.230	-157.93	172.95
3784.1	14.878	139.19	315.92	-60.264	-244.49	-248.47	-249.09	291.65	-22.970	-109.26	172.91
4422.3	-234.51	507.87	-697.33	-80.449	401.96	495.21	300.16	28.434	-237.62	-157.28	-326.44
4259.1	82.945	848.92	-1627.1	-620.00	330.48	389.07	492.58	580.55	-239.64	-191.23	-46.521
2633.8	-216.80	-446.69	153.68	6.6602	315.47	-112.11	-203.00	150.28	453.65	-0.7063	-100.43



16.1.6. Matriz de datos **Medias.txt** para la corriente de control yan2000.sas.

AMBIENTE	GENOTIPO	RENDIMIENTO
1	1	3850
1	2	3552.8
1	3	3352.8
1	4	3334.7
1	5	3762.5
1	6	4147.2
1	7	4037.5
1	8	2723.6
1	9	3638.9
1	10	3666.7
1	11	3615.3
2	1	3541.7
2	2	3750
2	3	3819.4
2	4	2847.2
2	5	4375
2	6	3819.5
2	7	4027.8
2	8	3263.9
2	9	3402.8
2	10	3680.6
2	11	4027.8
3	1	3308.3
3	2	3791.7
3	3	2327.8
3	4	2720.8
3	5	3791.7
3	6	3745.8
3	7	4037.5
3	8	3177.8
3	9	3027.8
3	10	3550
3	11	2865.3
4	1	3868.1
4	2	3254.2
4	3	3547.2
4	4	3176.4
4	5	3309.7
4	6	3263.9
4	7	4130.6
4	8	2661.1
4	9	3154.2
4	10	3691.7
4	11	3709.7
5	1	3097.9
5	2	3266
5	3	2161.1
5	4	1854.9
5	5	2993.1
5	6	3128.5
5	7	3808.3
5	8	2052.1
5	9	2418
5	10	2691
5	11	3071.6
6	1	3972.2
6	2	3375
6	3	3000
6	4	3430.5
6	5	3430.6
6	6	3902.8
6	7	4152.8
6	8	3319.4
6	9	3416.7
6	10	3375
6	11	3805.6
7	1	3951.4
7	2	4201.4
7	3	3430.6
7	4	3097.2
7	5	3888.9
7	6	3895.8



Σ

α

Π

Ω

\times

\pm

ϑ

AMBIENTE	GENOTIPO	RENDIMIENTO
7	7	4229.2
7	8	3722.2
7	9	3534.7
7	10	3645.9
7	11	4027.8
8	1	4340.3
8	2	5208.3
8	3	3055.6
8	4	3715.3
8	5	5173.6
8	6	5277.8
8	7	5416.7
8	8	4097.2
8	9	3958.3
8	10	4236.1
8	11	4166.7
9	1	4494.5
9	2	5386.1
9	3	1962.5
9	4	3012.5
9	5	4938.9
9	6	5008.4
9	7	5445.8
9	8	4486.1
9	9	3793.1
9	10	4038.9
9	11	4283.3
10	1	2569.5
10	2	2465.3
10	3	2118.1
10	4	2013.9
10	5	3298.6
10	6	2881.9
10	7	3125
10	8	2430.6
10	9	2861.1
10	10	2604.2
10	11	2604.2

16.2. Anova y Regresión Conjunta

16.2.1. Corriente de control realizada por el Dr. Ignacio Romagosa para la realización del Análisis de la Varianza mixto para dos factores: Genotipo como factor fijo y Ambiente como factor aleatorio.

```
* PROGRAMA ANOVA.SAS;
options ps=80;
DATA RAW;
  INFILE 'datcomp1.txt';
  INPUT site $ genotype $ blq yld;
```

```
if genotype='local' then delete;
RUN;
PROC anova ;
  CLASS BLQ SITE GENOTYPE;
  MODEL YLD = SITE SITE*BLQ GENOTYPE
  SITE*GENOTYPE;
  TEST H=SITE          E=SITE*BLQ;
  TEST H=GENOTYPE      E=SITE*GENOTYPE;
  MEANS GENOTYPE /LSD E=SITE*GENOTYPE;
QUIT;
```

16.2.2. Salida del programa ANOVA.SAS.

Analysis of Variance Procedure / Class Level Information									
Class	Levels	Values							
BLQ	3	1	2	3					
SITE	10	1.000	2.000	3.000	4.000	5.000	6.000	7.000	8.000
		9.000	10.000						
GENOTYPE	11	1.000	2.000	3.000	4.000	5.000	6.000	7.000	8.000
		9.000	10.000	11.000					

Number of observations in data set = 330

Analysis of Variance Procedure					
Dependent Variable: YLD					
Source	DF	Sum of Squares	Mean Square	F. Value	Pr > F
Model	129	200111704.6	1551253.5	11.32	0.0001
Error	200	27405078.2	137025.4		
Corrected Total	329	227516782.8			
	R-Square	C.V.	Root MSE	YLD Mean	
	0.879547	10.43627	370.1694	3546.95273	
Source	DF	Anova SS	Mean Square	F. Value	Pr > F
Site	9	94126697.43	10458521.94	76.33	0.0001
Blq*Site	20	10737724.06	536886.20	3.92	0.0001
Genotype	10	55712403.38	5571240.34	40.66	0.0001
Site*Genotype	90	39534879.71	439276.44	3.21	0.0001

Tests of Hypotheses using the Anova MS for BLQ*SITE as an error term

Source	DF	Anova SS	Mean Square	F. Value	Pr > F
Site	9	94126697.43	10458521.94	19.48	0.0001

Tests of Hypotheses using the Anova MS for SITE*GENOTYPE as an error term

Source	DF	Anova SS	Mean Square	F. Value	Pr > F
Genotype	10	55712403.38	5571240.34	12.68	0.0001

Analysis of Variance Procedure

T tests (LSD) for variable: YLD

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.05 df= 90 MSE= 439276.4 / Critical Value of T= 1.99 / Least Significant Difference= 339.98

Means with the same letter are not significantly different.





T	Grouping	Mean	N	GENOTYPE
	A	4241.1	30	7.000
B	A	3907.2	30	6.000
B	A	3896.2	30	5.000
B		3825.1	30	2.000
B	C	3699.4	30	1.000
B	C	3617.7	30	11.000
B	C	3518.0	30	10.000
B	C	3320.6	30	9.000
B	C	3193.4	30	8.000
B	C	2920.4	30	4.000
B	C	2877.5	30	3.000

16.2.3.- Corriente de control realizada por el Dr. Ignacio Romagosa para la realización del Análisis conjunto de la Varianza y de la regresión.

```
* PROGRAMA JRA.SAS;
Options ps=66 noNUMBER;
DATA RAW0;
  INFILE 'A:DATCOMP1.txt';
  INPUT SITE $ GENOTYPE $ BLQ YLD;
RUN;
PROC SORT ;
  BY SITE ;
RUN;
PROC MEANS NOPRINT;
  BY SITE;
  VAR YLD;
  OUTPUT OUT=SITEMEAN MEAN=MSYLD;
RUN;
DATA RAW;
  SET RAW0;
  IF GENOTYPE ='local' THEN DELETE;
RUN;
DATA JOINTREG;
  MERGE RAW SITEMEAN;
  BY SITE;
  KEEP SITE GENOTYPE BLQ YLD MSYLD;
RUN;
PROC GLM;
  CLASS GENOTYPE BLQ;
  MODEL YLD = MSYLD MSYLD*BLQ
```

```
GENOTYPE MSYLD*GENOTYPE /SS1;
RUN;
PROC SORT;
  BY GENOTYPE;
RUN;
PROC REG NOPRINT OUTEST=JRA;
  BY GENOTYPE;
  MODEL YLD = MSYLD;
RUN ;
PROC SORT DATA=RAW;
  BY GENOTYPE;
RUN;
PROC MEANS NOPRINT ;
  BY GENOTYPE ;
  VAR YLD;
  OUTPUT OUT=GENOMEAN MEAN=GENOYLD;
RUN;
DATA CONVERGE;
  MERGE JRA (DROP=_TYPE_ _MODEL_)
  GENOMEAN;
  BY GENOTYPE ;
  KEEP GENOTYPE GENOYLD INTERCEP
  MSYLD _RMSE_;
RUN;
PROC PRINT;
  VAR GENOTYPE GENOYLD INTERCEP
  MSYLD _RMSE_;
RUN;
PROC CORR;
  var GENOYLD MSYLD;
QUIT;
```

16.2.4. Salida del programa JRA.SAS.

General Linear Models Procedure / Class Level Information

Class	Levels	Values
GENOTYPE	11	1.000 2.000 3.000 4.000 5.000 6.000 7.000 8.000 9.000 10.000 11.000
BLQ	3	1 2 3

Number of observations in data set = 330

General Linear Models Procedure

Dependent Variable: YLD

Source	DF	Sum of Squares	Mean Square	F. Value	Pr > F
Model	23	161227644.9	7009897.6	32.36	0.0001
Error	306	66289137.9	216631.2		
Corrected Total	329	227516782.8			

	R-Square	C.V.	Root MSE	YLD Mean
	0.708641	13.133125	465.4365	3546.95273

Source	DF	Type I SS	Mean Square	F. Value	Pr > F
MSYLD	1	94126697.43	94126697.43	434.50	0.0001
MSYLD*BLQ	2	4302861.89	2151430.95	9.93	0.0001
GENOTYPE	10	55712403.38	5571240.34	25.72	0.0001
MSYLD*GENOTYPE	10	7085682.23	708568.22	3.27	0.0005

OBS	GENOTYPE	GENOYLD	INTERCEP	MSYLD	_RMSE_
1	1.000	3699.38	280.89	0.96378	459.700
2	10.000	3517.99	404.74	0.87773	330.441
3	11.000	3617.71	346.74	0.92219	399.837
4	2.000	3825.07	-1302.64	1.44567	483.717
5	3.000	2877.50	1508.92	0.38585	720.744
6	4.000	2920.35	-180.75	0.87430	487.208
7	5.000	3896.25	-138.31	1.13747	482.624
8	6.000	3907.16	-609.12	1.27328	334.135
9	7.000	4241.11	193.11	1.14126	444.859
10	8.000	3193.40	-1109.96	1.21326	513.545
11	9.000	3320.56	606.39	0.76521	493.691

**CORRELATION ANALYSIS
2 'VAR' Variables: GENOYLD MSYLD**

Simple Statistics				
Variable	N	Mean	Std Dev	Sum
GENOYLD	11	3547	430.93852	39016
MSYLD	11	1.00000	0.28776	11.00000

Simple Statistics		
Variable	Minimum	Maximum
GENOYLD	2878	4241
MSYLD	0.38585	1.44567

Pearson Correlation Coefficients / Prob > IRI under Ho: Rho=0 / N=11

	GENOYLD	MSYLD
GENOYLD	1.00000	0.66822
	0.0	0.0246
MSYLD	0.66822	1.00000
	0.0246	0.0

16.3. Estadísticos Paramétricos Univariantes

15.3.1. PROGRAMA REALIZADO EN BASIC PARA EL CALCULO DE LOS SIGUIENTES PARAMETROS: Cuadrados Medios de la Ecovalencia (CM-W), Ecovalencia (Wi), Coeficiente de Regresión (Bi), Cuadrado Medio Residual de la Regresión (SD2), Varianza ambiental (SX2), Varianza de Shukla (1972) (SIGMA(2)), Coeficiente de determinación (R2), Parámetros de Eskridge (1991) (EV, FW, SH, ER) y el Coeficiente de Variación.

LA MATRIZ DE ENTRADA ES UNA MATRIZ COLUMNA, ES DECIR, LA COLUMNA DE LOS RENDIMIENTOS MEDIOS DE LA MATRIZ DAT-COMP1.TXT.

```

10 HA=10:K=11:R=3:REM ***HA=Núm. de
ambientes; K=Núm. genotipos
R=Núm. de repeticiones
20 DIM ENV (30,20,10), AMB (HA), GEN (K),
SIGMA (K), YIM (K), R2 (K), S2X (K), SD2 (K),
SEG (K), CONJ (HA,K), NIV (HA), ECOV (K), EV
(K), FW (K), SH (K), V1I (K), VI (K), ER (K)
30 DIM Z (HA), B1 (K), B (K), SAF (K), CV (K),
CV1 (K), CMWI (K)
40 OPEN "A:MATCOLUM.10" AS#1 LEN=12
50 FIELD#1,10 AS V1$
60 FOR H=1 TO HA:FOR H1=1 TO K:FOR H2=1
TO R:I=I+1
70 Y=EOF(1):IF Y=-1 THEN CLOSE#1:GOTO
130
80 GET#1,I
90 V1=VAL(V1$)
100 ENV(H,H1,H2)=V1:S1=S1+V1:NEXT H2:
NEXT H1:NEXT H
110 REM ** a(amb,gen,rep)=yld***
120 MEDIA=S1/(HA*K*R)
130 CLS
140 LOCATE 20,15:PRINT "***** ESTOY PRO-
CESANDO *****"
150 REM ***** ECOVALENCIAS *****
160 FOR I=1 TO HA:FOR I1=1 TO K:FOR I2=1
TO R
170 SUM1=SUM1+ENV(I,I1,I2)
180 NEXT I2:NEXT I1
190 AMB(I)=SUM1/(K*R):SUM1=0
200 NEXT I
210 FOR I=1 TO K:FOR I1=1 TO HA:FOR I2=1
TO R
220 SUM2=SUM2+ENV(I1,I,I2)
230 NEXT I2:NEXT I1
240 GEN(I)=SUM2/(HA*R):SUM2=0

```

```

250 NEXT I
260 FOR I=1 TO HA:FOR I1=1 TO K:FOR I2=1
TO R
270 SUM3=SUM3+ENV(I,I1,I2)
280 NEXT I2
290 CONJ(I,I1)=SUM3/R:SUM3=0
300 NEXT I1:NEXT I
310 FOR I=1 TO K:FOR I1=1 TO HA
320 NIV(I1)=CONJ(I1,I)-GEN(I)-AMB(I1)+MEDIA
330 SUM4=SUM4+(NIV(I1)*NIV(I1))
340 NEXT I1
350 ECOV (I) = SUM4: SUM4 = 0: SUME =
SUME + ECOV (I)
355 CMWI(I)=ECOV(I)/17.3
360 NEXT I
365 LPRINT "***** CUADRADOS MEDIOS DE
LAS ECOVALENCIAS DE WRICKE *****"
366 LPRINT:LPRINT:FOR I=1 TO K
367 LPRINT "CM-W(“;I;”) = “;CMWI(I):NEXT
I:GOSUB 1160
370 REM ***** COEF. REGRESION *****
380 FOR I1=1 TO K:SUM7=0:FOR I=1 TO HA
390 Z(I)=AMB(I)-MEDIA
400 SUM6=SUM6+((CONJ(I,I1)-GEN(I1)-
AMB(I)+MEDIA)*Z(I))
410 SUM7=SUM7+(Z(I)*Z(I))
420 NEXT I
430 B1(I1)=SUM6/SUM7:B(I1)=1+B1(I1)
440 SUM6=0:NEXT I1
450 REM ***DESVIACION MEDIA CUADRATICA
*****
460 FOR I=1 TO K
470 SEG(I)=B1(I)*B1(I):SEG(I)=SEG(I)*SUM7
480 SD2(I)=ECOV(I)-SEG(I):SD2(I)=SD2(I)/(HA-
2)
490 NEXT I
500 REM *****VARIANZA AMBIENTAL *****
510 FOR I=1 TO K:FOR I1=1 TO HA
520 SUM8 = (CONJ(I1,I)- GEN (I)): SUM9 =
SUM9 + (SUM8*SUM8)
530 NEXT I1
540 S2X(I)=((SUM9)/(HA-1)):SUM9=0
550 NEXT I
560 REM ***VARIANZA MEDIA ***
570 FOR I=1 TO HA
580 SUM18=(AMB(I)-MEDIA): SUM19 = SUM19
+ (SUM18*SUM18)
590 NEXT I:S2YMED = SUM19/(HA-1): MSE =
K*S2YMED
600 MSGE=3*SUME/((K-1)*(HA-1)): SIG2E =
(MSE-MSGE)/K
610 REM ***** COEF. DETERMINACION PINT-
HUS *****
620 FOR I=1 TO K
630 R2(I)=SD2(I)/S2X(I)
640 R2(I)=1-R2(I)

```



```

650 NEXT I
660 REM *** VARIANZA DE ESTABILIDAD ***
670 PA=(HA-1)*(K-2):PA=K/PA:PE=(K-2)*(HA-1)*(K-1)
680 FOR I=1 TO K
690 SIGMA(I)=PA*ECOV(I)-(SUME/PE):NEXT I
700 REM ***** COEFICIENTE DE VARIACION *****
710 FOR I=1 TO K
720 CV1(I)=S2X(I)^.5
730 CV(I)=(CV1(I)/GEN(I))*100
740 NEXT I
750 REM ***** DETERMINACION ESTADISTICOS "SAFETY-FIRTS"
760 REM *****VARIANZA AMBIENTAL*****
770 FOR I=1 TO K
780 YIM(I)=GEN(I):VI(I)=S2X(I)
790 NEXT I
800 GOSUB 1170:FOR I=1 TO K: EV(I) = SAF (I):
NEXT I
810 REM *****COEF. REGRESION *****
820 FOR I=1 TO K
830 CSTE = (1-(1/HA)): V1I(I) = CSTE*S2YMED
*B1(I)^2:VI(I)=V1I(I):NEXT I
840 GOSUB 1170:FOR I=1 TO K: FW(I) = SAF(I):
NEXT I
850 REM *****VAR. SHUKLA *****
860 FOR I=1 TO K:VI(I)=SIG2E+SIGMA(I):NEXT I
870 GOSUB 1170:FOR I=1 TO K: SH(I) = SAF(I):
NEXT I
880 REM *** DESV. CUADRATICA *****
890 FOR I=1 TO K
900 VI(I)=V1I(I)+SD2(I):NEXT I
910 GOSUB 1170:FOR I=1 TO
K:ER(I)=SAF(I):NEXT I
920 REM ***** SALIDAS POR PANTALLA O
IMPRESORA *****
930 CLS
940 PRINT "SS(GE) = ",3*SUME:PRINT "SY =";
S2YMED
950 PRINT "SIGMA}e = ",SIG2E
960 GOSUB 1160
970 LPRINT "GEN. Wi Bi SD2 SX2 SIGMA(2)
MED R2"
980 LPRINT "=====
=====
"
990 LPRINT:LPRINT
1000 FOR I=1 TO K
1010 LPRINT I; ECOV(I); B(I);SD2(I); S2X(I);
SIGMA(I); GEN(I); R2(I)
1020 NEXT I
1030 GOSUB 1160
1040 LPRINT:LPRINT:LPRINT "GEN. EV FW
SH ER MED "
1050 LPRINT "=====
=====
"

```

```

1060 LPRINT:LPRINT
1070 FOR I=1 TO K
1080 LPRINT I;EV(I);FW(I);SH(I);ER(I);GEN(I)
1090 NEXT I
1100 GOSUB 1160
1110 REM *****PRINTEO DEL CV.
*****
1120 LPRINT:LPRINT:FOR I=1 TO K
1130 LPRINT "COEF. DE VARIAC. GENOTIPO
(“;I;”) =”;CV(I)
1140 NEXT I
1150 END
1160 PRINT "PULSE TECLA":WHILE
INKEY$="" :WEND:RETURN
1170 REM *****RUTINA SAFETY *****
1180 FOR I=1 TO K
1190 ZA=1.645
1200 SAF(I)=YIM(I)-(ZA*VI(I)^.5)
1210 NEXT I:RETURN

```

16.3.2. Corriente de control, para el paquete estadístico SAS, realizada por el Dr. Ignacio Romagosa para el cálculo del índice de Superioridad de Lin y Binns.

```

* PROGRAMA SUPINDEX.SAS;
options ps=80;
DATA RAW;
  INFILE 'DATCOMP1.TXT';
  INPUT SITE $ GENOTYPE $ BLQ YLD;
RUN;
PROC SORT;
  BY SITE GENOTYPE;
RUN;
PROC MEANS NOPRINT;
  BY SITE GENOTYPE;
  VAR YLD;
  OUTPUT OUT=SITEGENO MEAN = YLD;
RUN;
PROC MEANS NOPRINT DATA=SITEGENO;
  BY SITE ;
  VAR YLD;
  OUTPUT OUT=MAXYIELD MAX = MAXYLD;
RUN;
DATA RAW2;
  MERGE SITEGENO MAXYIELD;
  BY SITE;
  P1= MAXYLD - YLD;
RUN;
PROC SORT;
  BY GENOTYPE;
RUN;
PROC MEANS NOPRINT DATA=RAW2 MEAN
STD uss;
  BY GENOTYPE;

```



```
VAR YLD P1;
OUTPUT OUT=SUPINDEX MEAN = YLD X STD
= STDYLD Y USS = Z P1;
RUN;
proc corr;
var yld stdyld p1;
run;
proc sort;
```

```
by p1;
run;
PROC PRINT DATA=SUPINDEX;
VAR GENOTYPE YLD STDYLD P1;
QUIT;
```

16.3.2.1. Salida del programa SUPINDEX.SAS.



Análisis del Superíndice de LIN y BINNS para los 11 genotipos

**CORRELATION ANALYSIS
3 'VAR' Variables: YLD STDYLD P1**

Simple Statistics				
Variable	N	Mean	Std Dev	Sum
YLD	11	3547	430.93852	39016
STDYLD	11	659.98864	126.09271	7260
P1	11	9125902	832201	100384922

Simple Statistics		
Variable	Minimum	Maximum
YLD	2878	4241
STDYLD	463.35813	898.33278
P1	162719	27985209

Pearson Correlation Coefficients / Prob > IRI under Ho: Rho=0 / N=11

	YLD	STDYLD	P1
YLD	1.00000 0.0	0.33002 0.3216	-0.94874 0.0001
STDYLD	0.33002 0.3216	1.00000 0.0	-0.22359 0.5087
P1	-0.94874 0.0001	-0.22359 0.5087	1.00000 0.0

OBS	GENOTYPE	YLD	STDYLD	P1
1	7.000	4241.11	700.386	162718.89
2	6.000	3907.16	764.045	2164863.72
3	5.000	3896.25	723.460	2500331.71
4	2.000	3825.07	898.333	3213751.86
5	1.000	3699.38	581.971	4593180.09
6	11.000	3617.71	578.907	6054839.48
7	10.000	3517.99	518.824	7192596.11
8	9.000	3320.56	463.358	11183205.99
9	8.000	3193.40	758.453	13610867.50
10	4.000	2920.35	593.609	21723358.02
11	3.000	2877.50	678.528	27985208.55

16.4. Método de Tai

16.4.1. Programa en BASIC para el cálculo de los parámetros α y λ del método de Tai. La matriz MATCOLUM.10 es la matriz columna de rendimientos.

```
10 HA=10:K=11:R=3
20 DIM ENV(30,20,10), AMB(HA), GEN(K),
S2GL(K), SLGL(K), GL(K,HA), EXIL(HA),
ALFA(K), LAMBDA(K), Y(K,HA)
25 REM ***** LA MATRIZ DE ENTRADA ES UNA
```

```
MATRIZ COLUMNA ****
30 OPEN "A:MATCOLUM.10" AS#1 LEN=12
40 FIELD#1,10 AS V1$
50 FOR H=1 TO HA:FOR H1=1 TO K:FOR H2=1
TO R:I=I+1
60 Y=EOF(1):IF Y=-1 THEN CLOSE#1:GOTO
120
70 GET#1,I
80 V1=VAL(V1$)
90 ENV(H,H1,H2)=V1:S1=S1+V1:NEXT H2:
NEXT H1:NEXT H
100 REM ** a(amb,gen,rep)=yld***
```



```

110 MEDIA=S1/(HA*K*R)
120 CLS
130 FOR I=1 TO K
140 FOR I2=1 TO HA:FOR I3=1 TO R
150 SUM1=SUM1+ENV(I2,I,I3)
160 NEXT I3:NEXT I2
170 GEN(I) = SUM1: GEN(I) = (GEN(I))/(HA*R):
SUM1=0
180 NEXT I
190 FOR H=1 TO HA
200 FOR H2=1 TO K:FOR H3=1 TO R
210 SUM2 =SUM2+ENV(H,H2,H3)
220 NEXT H3:NEXT H2
230 AMB(H)=(SUM2/(K*R)):EXIL(H)=AMB(H)-
MEDIA:SUM2=0
240 NEXT H
250 FOR I=1 TO K
260 FOR I1=1 TO HA
270 FOR I3=1 TO R
280 SUM3=SUM3+ENV(I1,I,I3)
290 NEXT I3
300 Y(I,I1)=SUM3/(R):SUM3=0
310 NEXT I1:NEXT I
320 FOR I=1 TO HA:FOR H=1 TO K
330 GL(H,I)=Y(H,I)-GEN(H)-AMB(I) + MEDIA:
NEXT H:NEXT I
340 MSL=10458521.9#
350 MSB=536886.2
360 MSE=137025.4
370 FOR I=1 TO K:FOR H=1 TO HA
380 SLGL(I)=SLGL(I)+EXIL(H)*GL(I,H):
S2GL(I)=S2GL(I)+GL(I,H)*GL(I,H):NEXT H
390 SLGL(I) = SLGL(I) / 19: DENO = MSL-MSB:
DENO = DENO/(K*R)
400 ALFA(I)=SLGL(I)/DENO
410 NUM1 = ALFA(I) * SLGL(I): NUM1 = (S2GL
(I)/19)- NUM1
420 DENO2 = 14*MSE/(K*R): LAMBDA(I) =
NUM1/DENO2:NEXT I
430 LPRINT "RESULTADOS DEL PROGRAMA
DE TAI (1971) "
440 LPRINT " "
450 LPRINT " GENOTIPO MEDIA
ALFA LAMBDA "
460 FOR I=1 TO K
470 LPRINT USING "#####.#####"; I,
GEN(I), ALFA(I), LAMBDA(I): NEXT I
480 END
490 PRINT "PULSE TECLA":WHILE INKEY$ =
"":WEND:RETURN

```

16.4.2. Programa en BASIC para generar las curvas de confianza para el parámetro α .

```

10 REM ***** PROGRAMA PARA CURVA ALFA-

```

```

LAMBDA *****
20 DIM AL FA(200),LAMBDA(200)
25 REM ***** MSL, MSB, MSE SON RECOGIDOS
DEL ANOVA *****
30 MSL=10458521.9#
40 MSB=536886.2
50 MSE=137025.4#
60 T80=1.397:T90=1.86:T95=2.306: REM ***
VALOR DE t SEGUN EL NIVEL
DE PROBABILIDAD *****
70 M1=10:N2=8
80 NUM1=M1*MSE*MSL:K=1!:H=2
90 DEN1=MSL-MSB: DEN2 = N2*MSL: DEN31 =
((T80*T80)+N2)*MSB:
DEN32=((T90*T90)+N2)*MSB
100 DEN33=((T95*T95)+N2)*MSB
105 OPEN "A:CURVA95.TXT" FOR OUTPUT
AS#1
110 FOR I=0! TO 7 STEP .2
120 NUM2=I*NUM1:DENOT=(DEN1*(DEN2-
DEN33)):ALF=(NUM2/DENOT)^.5
130 ALFA(I)=K*T95*ALF
132 PRINT#1, USING "#####.#####";I,ALFA(I)
140 NEXT I
145 IF H=2 THEN K=(-1):H=3:GOTO 110
147 CLOSE#1
200 END

```

El fichero CURVA95.TXT generado puede ser introducido en un programa de diseño de gráficos como el Harvard Graphics para la representación de las curvas de confianza del parámetro α .

16.5. Estadísticos univariantes No-paramétricos

16.5.1. Programa en lenguaje SAS para la obtención del fichero texto que contiene las ordenaciones de los genotipos en los distintos ambientes.

```

* PROGRAMA RANG.SAS;
OPTIONS PS=80 NONUMBER;
DATA RAW;
INFILE 'DATMED.TXT';
* DATMED.TXT ES UN FICHERO DE MEDIAS,
SIN REPETICIONES;
INPUT SITE $ GENOTYPE $ YLD;
RUN;
PROC SORT;
BY SITE GENOTYPE;
RUN;
PROC RANK DESCENDING OUT = RGE TIES =
MEAN;

```





BY SITE;
RANKS RYLD;
VAR YLD;
RUN;
PROC SORT;
BY GENOTYPE;
RUN;

PROC PRINT DATA=RANG.TXT;
VAR SITE GENOTYPE RYLD;
RUN;

16.5.2. Salida del programa RANG.SAS.



AMB.	GEN.	RANGO
1.0000	1.0000	3.0000
1.0000	2.0000	8.0000
1.0000	3.0000	9.0000
1.0000	4.0000	10.0000
1.0000	5.0000	4.0000
1.0000	6.0000	1.0000
1.0000	7.0000	2.0000
1.0000	8.0000	11.0000
1.0000	9.0000	6.0000
1.0000	10.0000	5.0000
1.0000	11.0000	7.0000
2.0000	1.0000	8.0000
2.0000	2.0000	6.0000
2.0000	3.0000	5.0000
2.0000	4.0000	11.0000
2.0000	5.0000	1.0000
2.0000	6.0000	4.0000
2.0000	7.0000	2.5000
2.0000	8.0000	10.0000
2.0000	9.0000	9.0000
2.0000	10.0000	7.0000
2.0000	11.0000	2.5000
3.0000	1.0000	6.0000
3.0000	2.0000	2.5000
3.0000	3.0000	11.0000
3.0000	4.0000	10.0000
3.0000	5.0000	2.5000
3.0000	6.0000	4.0000
3.0000	7.0000	1.0000
3.0000	8.0000	7.0000
3.0000	9.0000	8.0000
3.0000	10.0000	5.0000
3.0000	11.0000	9.0000
4.0000	1.0000	2.0000
4.0000	2.0000	8.0000
4.0000	3.0000	5.0000
4.0000	4.0000	9.0000
4.0000	5.0000	6.0000
4.0000	6.0000	7.0000
4.0000	7.0000	1.0000
4.0000	8.0000	11.0000
4.0000	9.0000	10.0000
4.0000	10.0000	4.0000
4.0000	11.0000	3.0000
5.0000	1.0000	4.0000
5.0000	2.0000	2.0000
5.0000	3.0000	9.0000
5.0000	4.0000	11.0000
5.0000	5.0000	6.0000
5.0000	6.0000	3.0000
5.0000	7.0000	1.0000
5.0000	8.0000	10.0000
5.0000	9.0000	8.0000
5.0000	10.0000	7.0000
5.0000	11.0000	5.0000
6.0000	1.0000	2.0000
6.0000	2.0000	8.0000
6.0000	3.0000	11.0000
6.0000	4.0000	6.0000
6.0000	5.0000	5.0000
6.0000	6.0000	3.0000
6.0000	7.0000	1.0000
6.0000	8.0000	10.0000
6.0000	9.0000	7.0000
6.0000	10.0000	9.0000

(Continúa)

AMB.	GEN.	RANGO
6.0000	11.0000	4.0000
7.0000	1.0000	4.0000
7.0000	2.0000	2.0000
7.0000	3.0000	10.0000
7.0000	4.0000	11.0000
7.0000	5.0000	6.0000
7.0000	6.0000	5.0000
7.0000	7.0000	1.0000
7.0000	8.0000	7.0000
7.0000	9.0000	9.0000
7.0000	10.0000	8.0000
7.0000	11.0000	3.0000
8.0000	1.0000	5.0000
8.0000	2.0000	3.0000
8.0000	3.0000	11.0000
8.0000	4.0000	10.0000
8.0000	5.0000	4.0000
8.0000	6.0000	2.0000
8.0000	7.0000	1.0000
8.0000	8.0000	8.0000
8.0000	9.0000	9.0000
8.0000	10.0000	6.0000
8.0000	11.0000	7.0000
9.0000	1.0000	5.0000
9.0000	2.0000	2.0000
9.0000	3.0000	11.0000
9.0000	4.0000	10.0000
9.0000	5.0000	4.0000
9.0000	6.0000	3.0000
9.0000	7.0000	1.0000
9.0000	8.0000	6.0000
9.0000	9.0000	9.0000
9.0000	10.0000	8.0000
9.0000	11.0000	7.0000
10.0000	1.0000	7.0000
10.0000	2.0000	8.0000
10.0000	3.0000	10.0000
10.0000	4.0000	11.0000
10.0000	5.0000	1.0000
10.0000	6.0000	3.0000
10.0000	7.0000	2.0000
10.0000	8.0000	9.0000
10.0000	9.0000	4.0000
10.0000	10.0000	5.5000
10.0000	11.0000	5.5000

(Conclusión)

16.5.3. Programa en BASIC para el cálculo de los estadísticos no- paramétricos de Hühn.

```

10 REM **** PROGRAMA DE ESTADISTICOS
NO-PARAMETRICOS ***
20 K=11:HA=10
30 REM ***** K=NUM. DE GENOTIPOS; HA=
NUM. DE AMBIENTES ****
35 DIM R(K,HA), S1(K), S2(K), S3(K), INRME(K),
RME(K), Z1(K), Z2(K), S6(K)
37 OPEN "A:RANG.TXT" AS#1 LEN=12: REM ***
RANG.TXT= FICHERO TEXTO
    QUE CONTIENE LAS ORDENACIONES DE
CADA GENOTIPO EN CADA AMBIENTE
40 FIELD#1,10 AS V1$
50 FOR I=1 TO HA
60 FOR H=1 TO K
70 Z=Z+1
80 Y=EOF(1):IF Y=-1 THEN CLOSE#1:GOTO 140

```

```

90 GET#1,Z
100 V1=VAL(V1$)
110 R(H,I)=V1
120 NEXT H
130 NEXT I
140 REM ***** ESTADISTICOS NO-PARAMETRI-
COS *****
150 REM ***** S(1) *****
160 AMB1=HA*(HA-1):AMB=2/AMB1
170 FOR I=1 TO K
180 FOR H=1 TO HA-1
190 FOR J=H+1 TO HA
200 RG=R(I,H)-R(I,J)
210 RG1=ABS(RG)
220 SUM=SUM+RG1
230 NEXT J
240 SUM1=SUM1+SUM:SUM=0
250 NEXT H
260 S1(I)=SUM1*AMB:SUM=0:SUM1=0
270 NEXT I

```

```

280 REM ***** S(2) Y S(3) *****
290 FOR I=1 TO K:FOR H=1 TO HA
300 SUM2=SUM2+R(I,H)
310 NEXT H
320 RME(I)=SUM2/(HA): INRME(I)=15-RME(I):
SUM2=0
330 NEXT I
340 FOR I=1 TO K
350 FOR H=1 TO HA
360 CUA=R(I,H)-RME(I)
370 CUA2=CUA*CUA
380 SUM3=SUM3+CUA2
390 NEXT H
400 S2(I)=SUM3/(HA-1)
405 S3(I)=SUM3/INRME(I):SUM3=0
410 NEXT I
420 REM ***** S(6) *****
430 FOR I=1 TO K
440 FOR J=1 TO HA
450 DQ=R(I,J)-RME(I)
460 DQ1=ABS(DQ)
480 SUM4=SUM4+DQ1
490 NEXT J
500 S6(I)=SUM4/INRME(I):SUM4=0
510 NEXT I
550 REM ***** ESTADISTICOS *****
560 REM ***** ZM(I) *****
570 K2=K*K:ES1=(K2-1)/(3*K)
580 ES2=(K2-1)/12
590 VS1=K2-1:VSS1=((K2-4)*(HA+3))+30
600 VSS2=K2*45*HA*(HA-1)
610 VS1=VS1*VSS1/VSS2
620 V21S=K2-1: V22S=(K2-4)*(HA-1)*2: V23S=
(K2-1)*5
630 V24S=360*HA*(HA-1)
640 VS2=V21S*(V22S+V23S)/V24S
650 FOR I=1 TO K
660 Z11=S1(I)-ES1
670 Z12=Z11*Z11
680 Z1(I)=Z12/VS1
690 Z22=S2(I)-ES2
700 Z23=Z22*Z22
710 Z2(I)=Z23/VS2
720 SUM7=SUM7+Z1(I)
730 SUM8=SUM8+Z2(I)
740 NEXT I
750 LPRINT " GEN. S1(I) S2(I) Z1(I) Z2(I)
S3(I) S6(I) "
760 FOR I=1 TO K
770 LPRINT USING
"#####.#####";I;S1(I);S2(I);Z1(I);Z2(I);S3(I);S6(I)
780 NEXT I

```

```

790 LPRINT "S1 = ";SUM7,"S2 = ";SUM8
800 END

```

16.5.4. Corriente de control realizada por el Dr. Ignacio Romagosa para la obtención de las gráficas de Ketata.

```

* PROGRAMA KETATA.SAS;
options ps=80 NONUMBER;
DATA RAW;
  INFILE 'DATCOMP1.TXT';
  INPUT SITE $ GENOTYPE $ BLQ YLD;
RUN;
PROC SORT ;
  BY SITE GENOTYPE;
RUN;
PROC MEANS NOPRINT;
  BY SITE;
  VAR YLD;
  OUTPUT OUT=SITEMEAN MEAN =MYLDG;
RUN;
DATA RAW2;
  MERGE RAW SITEMEAN;
  BY SITE;
RUN;
PROC MEANS NOPRINT;
  BY SITE GENOTYPE;
  VAR YLD;
  OUTPUT OUT=MGE MEAN =MYLD;
RUN;
proc rank descending out=rge ties=mean;
  by site ;
  ranks ryld;
  var myld;
run;
proc sort;
  by genotype;
run;
proc means noprint;
  by genotype ;
  var myld ryld;
  output out=pge mean=myld mryld std=syld
sryld;
run;
proc print data=pge ;
  var genotype myld syld mryld sryld;
run;
proc plot;
  plot syld*myld = genotype;
  plot sryld*mryld = genotype;
quit;

```

16.5.4.1. Salida del programa KETATA.SAS.

ANALISIS DE KETATA PARA LOS 11 GENOTIPOS EN LOS 10 AMBIENTES

OBS	GENOTYPE	MYLD	SYLD	MRYLD	SRYLD
1	8.000	3193.40	758.453	8.90	1.79196
2	9.000	3320.56	463.358	7.90	1.79196
3	10.000	3517.99	518.824	6.45	1.60641
4	11.000	3617.71	578.907	5.30	2.17562
5	1.000	3699.38	581.971	4.60	2.01108
6	2.000	3825.07	898.333	4.95	2.87180
7	3.000	2877.50	678.528	9.20	2.34758
8	4.000	2920.35	593.609	9.90	1.52388
9	5.000	3896.25	723.460	3.95	1.92137
10	6.000	3907.16	764.045	3.50	1.64992
11	7.000	4241.11	700.386	1.35	0.57975

16.6. Análisis AMMI

16.6.1. Corriente de control, para el paquete estadístico SAS, realizada por el Dr. Ignacio Romagosa para el cálculo del ANOVA y las coordenadas genotípicas y ambientales para los tres primeros ejes del análisis AMMI.

```
* PROGRAMA AMMI.SAS;
options ps=80;
DATA RAW;
  INFILE 'A:DATCOMP1.TXT';
  INPUT SITE $ GENOTYPE $ BLQ YLD;
RUN;
PROC SORT ;
  BY SITE;
RUN;
PROC MEANS NOPRINT;
  BY SITE;
  VAR YLD;
  OUTPUT OUT=SITEMEAN MEAN=MYLDS;
RUN;
data sitename;
  l=1;
  DO WHILE(LAST=0);
    set sitemean END=LAST;
    numsite = l;
    keep numsite site mylds;
    output;
    l=l+1;
  END;
run;
data recoding;
  merge raw sitename;
  by site;
  keep numsite genotype yld;
```

```
run;
PROC SORT ;
  BY GENOTYPE;
RUN;
PROC MEANS NOPRINT ;
  BY GENOTYPE;
  VAR YLD;
  OUTPUT OUT=GENOMEAN MEAN=MYLDG;
RUN;
data varname;
  l=1;
  DO WHILE(LAST=0);
    set genomean END=LAST;
    numvar=l;
    if i<10 then varn='var'||l||'0'III;
    if i> 9 then varn='var'III;
    VARNUM =COMPRESS(VARN);
    keep numvar varnum genotype myldg;
    output;
    l=l+1;
  END;
run;
data rawrecod;
  merge recoding varname;
  by genotype;
  keep numsite numvar varnum yld ;
run;
PROC anova OUTSTAT=OUTSS;
  CLASS NUMSITE NUMVAR ;
  MODEL YLD = NUMSITE NUMVAR NUMSI-
TE*NUMVAR;
RUN;
* PARA OBTENER LA SUMA DE CUADRADOS
DE GxE;
DATA SSGxE;
  SET OUTSS;
```

Σ
 α
 Π
 Ω
 \times
 \pm
 ϑ

```

IF _SOURCE_ = 'NUMSITE*NUMVAR';
SSGE = SS;
KEEP SSGE;
RUN;
PROC SORT data=rawrecod;
  BY NUMSITE VARNUM;
RUN;
PROC MEANS NOPRINT;
  BY NUMSITE VARNUM;
  VAR YLD;
  OUTPUT OUT=SITEGENO MEAN=YLD;
RUN;
PROC GLM DATA=SITEGENO NOPRINT;
  CLASS NUMSITE VARNUM ;
  MODEL YLD = NUMSITE VARNUM;
  OUTPUT OUT=OUTRES R=RESID;
RUN;
DATA AMMI01;
  SET OUTRES;
  KEEP NUMSITE VARNUM RESID;
RUN;
PROC TRANSPOSE;
  BY NUMSITE;
  ID VARNUM;
RUN;
PROC PRINCOMP COV PREFIX=PCA
OUT=OUT1 OUTSTAT=OUT2 NOPRINT;
  VAR VAR01-VAR11; * num variedades
RUN;
PROC TRANSPOSE DATA=OUT2 OUT=OUT3;
RUN;
*PARA OBTENER LOS EIGENVELUES;
DATA EIGEN;
  SET OUT2;
  IF _TYPE_ = 'EIGENVAL';
  TOTAL = SUM(OF VAR01-VAR11); * 11 ES EL
NUMERO DE VARIETADES;
  EIGEN1 = VAR01/TOTAL;
  EIGEN2 = VAR02/TOTAL;
  EIGEN3 = VAR03/TOTAL;
  KEEP EIGEN1-EIGEN3;
RUN;
PROC PRINT;
RUN;
DATA PCAGENO;
  MERGE VARNAME OUT3;
  YLD = MYLDG;
  ID = 'GENOTYPE';
  KEEP _NAME_ ID numvar YLD PCA1-PCA3;
RUN;
DATA PCASITE;
  MERGE SITENAME OUT1 ;
  YLD = MYLDS;
  ID = 'SITE';
  KEEP _NAME_ ID numsite YLD PCA1-PCA3;
RUN;

```

```

* CALCULO DE LA SUMA DE CUADRADOS DE
AMMI1 ;
DATA AMMISS;
  b = 3; * num reps;
  n =11; * NUM VAR;
  m =10; * NUM LOC;
  MERGE EIGEN SSGxE;
  DO I=1 TO n+m;
    SSAMMI1 = EIGEN1*SSGE/b;
    SSAMMI2 = EIGEN2*SSGE/b;
    SSAMMI3 = EIGEN3*SSGE/b;
    KEEP SSAMMI1-SSAMMI3;
  OUTPUT;
  END;
RUN;
DATA PCAGS;
  SET PCAGENO PCASITE;
RUN;
DATA PCA13;
  MERGE PCAGS AMMISS;
  IF ID= 'GENOTYPE' THEN DO;
    num = numvar;
    PCA1 = -PCA1*(SSAMMI1**.25);
    PCA2 = -PCA2*(SSAMMI2**.25);
    PCA3 = -PCA3*(SSAMMI3**.25);
  OUTPUT;
  END;
  IF ID= 'SITE' THEN DO;
    num = numsite;
    PCA1 = -PCA1/(SSAMMI1**.25);
    PCA2 = -PCA2/(SSAMMI2**.25);
    PCA3 = -PCA3/(SSAMMI3**.25);
  OUTPUT;
  END;
  KEEP ID num numvar numsite YLD PCA1-
PCA3;
RUN;
proc sort;
  by numvar;
run;
data PCAGNAME;
  merge pca13 varname;
  by numvar;
run;
proc sort;
  by numsite;
run;
data PCASNAME;
  merge PCAGNAME sitename;
  by numsite;
  keep site genotype yld pca1-pca3;
run;
proc print;
  var site genotype yld pca1-pca3;
run;

```

16.6.1.1. Salida del programa AMMI.SAS.

Analysis of Variance Procedure / Class Level Information

Class	Levels	Values
NUMSITE	10	1 2 3 4 5 6 7 8 9 10
NUMVAR	11	1 2 3 4 5 6 7 8 9 10 11

Number of observations in data set = 330

Analysis of Variance Procedure

Dependent Variable: YLD

Source	DF	Sum of Squares	Mean Square	F. Value	Pr > F
Model	109	189373980.5	1737376.0	10.02	0.0001
Error	220	38142802.3	173376.4		
Corrected Total	329	227516782.8			

R-Square	C.V.	Root MSE	YLD Mean
0.832352	11.73923	416.3849	3546.95273

Source	DF	Anova SS	Mean Square	F. Value	Pr > F
NUMSITE	9	94126697.43	10458521.94	60.32	0.0001
NUMVAR	10	55712403.38	5571240.34	32.13	0.0001
NUMSITE*NUMVAR	90	39534879.71	439276.44	2.53	0.0001

OBS	EIGEN1	EIGEN2	EIGEN3
1	0.68240	0.10733	0.082337

OBS	SITE	GENOTYPE	YLD	PCA1	PCA2	PCA3
1	•	1	3699.38	-2.5573	17.3034	1.5354
2	•	2	3825.07	23.0687	1.5062	11.6628
3	•	3	2877.50	-38.9961	-8.9041	9.5565
4	•	4	2920.35	-13.5479	9.1164	-17.9384
5	•	5	3896.25	11.6675	-22.3115	0.3092
6	•	6	3907.16	13.4225	0.1297	-9.8802
7	•	7	4241.11	12.4567	10.9238	4.6409
8	•	8	3193.40	12.9454	-7.0165	-2.4739
9	•	9	3320.56	-6.4823	-5.2852	-10.7955
10	•	10	3517.99	-4.7490	-0.9414	-2.2163
11	•	11	3617.71	-7.2285	5.4793	15.5994
12	1	•	3607.44	-13.5129	3.3579	-11.3628
13	2	•	3686.88	-12.8966	-19.9734	13.6269
14	3	•	3304.04	8.4802	-5.1697	-10.2307
15	4	•	3433.34	-22.4565	12.7635	5.0229
16	5	•	2776.58	1.1942	9.5041	15.4644
17	6	•	3561.86	-9.6812	14.1253	-10.9673
18	7	•	3784.10	-4.8110	-0.8982	9.0582
19	8	•	4422.35	20.2378	-3.3857	-5.0709
20	9	•	4259.10	38.9322	5.6555	4.7934
21	10	•	2633.84	-5.4862	-15.9793	-10.3340

16.6.2. Programa en BASIC para la obtención de los modelos con dos (A) o tres (B) repeticiones totalmente al azar para el cálculo del estadístico RMS PD.

PROGRAMA A:

```

10 REM ****PROGRAMA PARA LA OBTENCION
DE LOS *****
20 REM ***** MODELOS DE DOS REPETICIO-
NES *****
30 REM ***** TOTALMENTE AL AZAR PARA
EL *****
40 REM ***** CALCULO DEL ESTADISTICO
RMS PD *****
45 REM ** LA MATRIZ DE ENTRADA ES UNA
MATRIZ COLUMNA *
50 CLS
60 RANDOMIZE TIMER
70 FOR W=1 TO 5
80 FG=INT((3000*RND(1))+1000):SOUND
FG,.5:NEXT W
90 LOCATE 15,15:PRINT "***** BIENVENIDO
AL PROGRAMA RMS PD *****"
100 PRINT
110 LOCATE 17,22:PRINT "Autor, 1991: FER-
NANDO FLORES GIL "
120 FOR O=1 TO 7:PRINT:NEXT
O:PRINT:GOSUB 450
130 CLS
140 HA=10:K=11:R=3
145 DIM
ENV(HA,K,R),ENVREP(HA,K,R),VALI(HA,K)
150 OPEN "A:MATCOLUM.10" AS#1 LEN=12
160 FIELD#1,10 AS V1$
170 FOR H=1 TO HA:FOR H1=1 TO K:FOR H2=1
TO R:I=I+1
180 Y=EOF(1):IF Y=-1 THEN CLOSE#1:GOTO
220
190 GET#1,I
200 V1=VAL(V1$)
210 ENV(H,H1,H2)=V1:S1=S1+V1:NEXT
H2:NEXT H1:NEXT H
220 REM *****A(AMB,GEN,REP)=YLD *****
230 MEDIA=S1/(HA*K*R):DIM REP(HA*K*R)
240 REP(0)=0
250 FOR I=1 TO HA:FOR H=1 TO K
260 V=V+1
270 J=J+1:G=G+1
280 IF G=3 THEN 320
290 REP(J)=INT(3*RND(1))+1
300 IF REP(J)=REP(J-1) THEN 290
310 ENVREP(I,H,G)=ENV(I,H,REP(J)):GOTO 270
320 FOR M=1 TO 3
330 IF M=REP(J-1) OR M=REP(J-2) THEN 340
ELSE 350

```

```

340 NEXT M
350 VALI(I,H)=ENV(I,H,M)
360 G=0:NEXT H:NEXT I
370 OPEN "A:MODEL1.TXT" FOR OUTPUT AS#2
380 FOR I=1 TO HA:FOR H=1 TO K:FOR R=1 TO
2
390 PRINT#2, USING "#####.###";I,H,R,ENV-
REP(I,H,R)
400 NEXT R:NEXT H:NEXT I
410 CLOSE#2
420 OPEN "A:VAL1.TXT" FOR OUTPUT AS#3
430 FOR I=1 TO HA:FOR H=1 TO K:PRINT#3,
VALI(I,H):NEXT H:NEXT I
440 END
450 PRINT "PULSA UNA TECLA *****":WHILE
INKEY$="":WEND:RETURN

```

PROGRAMA B:

```

10 REM *****PROGRAMA PARA LA
OBTENCION DE *****
20 REM ***** LOS 10 MODELOS DE TRES
REPETICIONES *****
25 REM ***** TOTALMENTE AL AZAR
*****
30 REM ***** NECESARIOS PARA EL CAL-
CULO DEL *****
40 REM ***** ESTADISTICO RMS PD
*****
50 REM ***** LA MATRIZ DE ENTRADA ES UNA
MATRIZ COLUMNA ****
60 CLS
70 RANDOMIZE TIMER
80 FOR W=1 TO 5
90 FG=INT ((3000*RND(1)) + 1000): SOUND
FG,.5: NEXT W
100 LOCATE 15,15:PRINT "***** BIENVENIDO
AL PROGRAMA RMS PD *****"
110 PRINT
120 LOCATE 17,22:PRINT "Autor, 1991: FER-
NANDO FLORES GIL "
130 FOR O=1 TO 7:PRINT:NEXT
O:PRINT:GOSUB 470
140 CLS
150 HA=17:K=11:R=4
160 DIM ENV (HA,K,R), ENVREP (HA,K,R), VALI
(HA,K)
170 OPEN "a:columna.10" AS#1 LEN=12
180 FIELD#1,10 AS V1$
190 FOR H=1 TO HA:FOR H1=1 TO K:FOR H2=1
TO R:I=I+1
200 Y=EOF(1):IF Y=-1 THEN CLOSE#1:GOTO
240
210 GET#1,I
220 V1=VAL(V1$)

```

```

230 ENV (H,H1,H2) = V1: S1 = S1+V1: NEXT H2:
NEXT H1:NEXT H
240 REM *****A(AMB,GEN,REP)=YLD *****
250 MEDIA=S1/(HA*K*R):DIM REP((20)*K*R)
260 FOR I=1 TO HA:FOR H=1 TO K
270 GOSUB 480
280 ENVREP(I,H,J)=ENV(I,H,REP(J))
290 GOSUB 480:IF J=5 THEN 340
300 FOR M=1 TO J-1
310 IF REP(M)=REP(J) THEN J=J-1:GOTO 290
320 NEXT M
330 ENVREP(I,H,J)=ENV(I,H,REP(J)):GOTO 290
340 NEXT H:NEXT I
350 LOCATE 17,22:PRINT " ESTOY GRA-
BANDO "
360 OPEN "A:MODEL1.TXT" FOR OUTPUT AS#2
370 FOR I=1 TO HA:FOR H=1 TO K:FOR R=1 TO
2
380 PRINT#2, USING "#####.####";I,H,R,ENV-
REP(I,H,R)
390 NEXT R:NEXT H:NEXT I
400 CLOSE#2
410 FOR I=1 TO HA:FOR H=1 TO K
420 VALI (I,H) = (ENVREP (I,H,3) + ENVREP
(I,H,4))/2
430 NEXT H:NEXT I
440 OPEN "A:VAL1.TXT" FOR OUTPUT AS#3
450 FOR I=1 TO HA:FOR H=1 TO K:PRINT#3,
VALI(I,H):NEXT H:NEXT I
460 CLS:END
470 PRINT "PULSA UNA TECLA *****":WHILE
INKEY$="" :WEND:RETURN
480 IF J=5 THEN FOR J=1 TO 4:REP(J)=0:NEXT
J:J=1:REP(J)=INT(4*RND(1))+1 :RETURN
490 J=J+1:REP(J)=INT(4*RND(1))+1:RETURN
    
```

16.6.3.- Programa en BASIC con el cual obtene-
mos el análisis de la varianza y partición de la
suma de cuadrados de la interacción GxE en las
correspondientes a los diez ejes del ACP.
También se obtienen las coordenadas genotípi-
cas y ambientales del análisis AMMI previa intro-
ducción de los ficheros texto que contienen los
vectores propios unitarios genotípicos y ambien-
tales resultantes del ACP realizado con el
NTSYS.

```

10 REM ****PROGRAMA DEL ANALISIS DE LA
VARIANZA *****
20 REM ***** TRIPLE FACTORIAL
*****
30 REM ***** Y CALCULO DE LAS COORDE-
NADAS *****
40 REM ***** PARA EL MODELO AMMI
*****
    
```

```

50 CLS
60 RANDOMIZE TIMER
70 FOR W=1 TO 5
80 FG = INT ((3000*RND(1)) + 1000): SOUND
FG,.5: NEXT W
90 FOR I=1 TO 10:PRINT:NEXT I:I=0:PRINT "
***** BIENVENIDO AL PROGRAMA A M M I
*****"
100 PRINT
110 PRINT " Autor, 1991: FER-
NANDO FLORES GIL "
120 FOR O=1 TO 7:PRINT:NEXT
O:PRINT:GOSUB 1630
130 CLS
140 INPUT "NUMERO DEL MODELO ";NUM
150 LPRINT:LPRINT:LPRINT:LPRINT "*****
MODELO A M M I Num. ";NUM
160 HA=17:K=11:R=2:N=HA*K*R
170 DIM SU(K): DIM SU2 (HA), LAMBDA (12),
VG (K,12)
180 DIM INT1 (HA,K), ENV (HA,K,R), VA (HA,12),
COOGEN (K,12)
190 DIM L (HA), L2 (HA), COOAM (HA,12),
RENDG (K), RENDAM (HA)
200 INPUT "NOMBRE DEL FICHERO DE DATOS
";FIC$
210 OPEN FIC$ AS#1 LEN=46
220 FIELD#1,11 AS V4$,11 AS V3$,11 AS V2$,11
AS V1$
230 FOR H1=1 TO HA:FOR H2=1 TO K
240 FOR H3=1 TO R
250 I=I+1
260 GET#1,I
270 Y=EOF(1):IF Y=-1 THEN CLOSE#1:GOTO
310
280 V1=VAL(V1$)
290 ENV (H1,H2,H3) = V1: S1 = S1 + V1: S2 = S2
+ (V1*V1)
300 NEXT H3:NEXT H2:NEXT H1
310 FC=S1*S1:FC=FC/N:MEDIA=S1/N
320 REM ***GENOTIPOS ***
330 FOR I=1 TO K
340 SUM=0
350 FOR I2=1 TO HA:FOR I3=1 TO R
360 SUM=SUM+ENV(I2,I,I3)
370 NEXT I3:NEXT I2
380 RENDG(I)=SUM/(HA*R)
390 SU(I)=SUM*SUM
400 ZU=ZU+SU(I)
410 NEXT I
420 GEN1=HA*R:GEN=ZU/GEN1:GENO=GEN-
FC:REM ***S.C. DE GENOTIPOS
430 REM **** MACROAMBIENTES ****
440 FOR I=1 TO HA
450 SUM2=0
460 FOR I2=1 TO K:FOR I3=1 TO R
    
```



470 SUM2=SUM2+ENV(I,I2,I3)
 480 NEXT I3:NEXT I2
 490 RENDAM(I)=SUM2/(K*R)
 500 SU2(I)=SUM2*SUM2
 510 ZU2=ZU2+SU2(I)
 520 NEXT I
 530 LOC1 = K*R: LOC2 = ZU2/LOC1: LOCAL =
 LOC2-FC: REM **SC LOCAL
 540 REM *** INTERACCION MACRO*GEN ****
 550 FOR I=1 TO HA:FOR I1=1 TO K
 560 SUM4=0
 570 FOR I3=1 TO R
 580 SUM4=SUM4+ENV(I,I1,I3)
 590 NEXT I3
 600 INT1(I,I1)=SUM4*SUM4
 610 ZU4=ZU4+INT1(I,I1)
 620 NEXT I1:NEXT I
 630 LV1 = R: LV = ZU4/LV1: INLG = LV-GENO-
 LOCAL-FC: REM ** S.C. LO*GEN **
 640 ERROR1 = S2-LV: ER = K*HA*(R-1):
 ERROR2 = ERROR1/ER
 650 REM *** PARTICION DE LA INTER. GxE ***
 660 PRINT "NOMBRE DEL FICHERO VECT.
 PROPIOS GENOTIPIICOS ":"INPUT G\$
 670 PRINT "NOMBRE DEL FICHERO VECT.
 PROPIOS AMBIENTALES ":"INPUT AM\$
 680 PRINT "PORCENTAJES DE VARIANZA DE
 LOS AUTOVALORES 1-10":"INPUT
 AUT1,AUT2,AUT3,AUT4,AUT5,AUT6,AUT7,AUT
 8,AUT9,AUT10
 690 SSPCA1 = INLG*AUT1: SSPCA2 =
 INLG*AUT2: SSPCA3 = INLG*AUT3: SSPCA4 =
 INLG*AUT4: SSPCA5 = INLG*AUT5: SSPCA6 =
 INLG*AUT6
 700 SSPCA7 = INLG*AUT7: SSPCA8 =
 INLG*AUT8: SSPCA9 = INLG*AUT9: SSPCA10
 = INLG*AUT10
 710 LAMBDA(1)=(SSPCA1/R)^.25:LAMBDA(2)
 = (SSPCA2/R)^.25:LAMBDA(3) =
 (SSPCA3/R)^.25:LAMBDA(4) =
 (SSPCA4/R)^.25:LAMBDA(5) =
 (SSPCA5/R)^.25:LAMBDA(6) = (SSPCA6/R)^.25
 720 LAMBDA(7)=(SSPCA7/R)^.25: LAMBDA(8)
 = (SSPCA8/R)^.25:LAMBDA(9) = (SSPCA9/R)
 ^.25: LAMBDA(10) = (SSPCA10/R)^.25
 730 OPEN G\$ AS#2 LEN=132
 740 FIELD#2,13 AS VG1\$,13 AS VG2\$,13 AS
 VG3\$,13 AS VG4\$,13 AS VG5\$,13 AS VG6\$,13
 AS VG7\$,13 AS VG8\$,13 AS VG9\$,13 AS VG10\$
 750 FOR I=1 TO K
 760 GET#2,I
 770 Y2=EOF(2):IF Y2=-1 THEN CLOSE#2:GOTO
 870
 780 VG1 = VAL(VG1\$): VG2 = VAL(VG2\$): VG3
 = VAL(VG3\$)
 790 VG4 = VAL(VG4\$): VG5 = VAL(VG5\$): VG6

= VAL(VG6\$)
 800 VG7 = VAL(VG7\$): VG8 = VAL(VG8\$): VG9
 = VAL(VG9\$)
 810 VG10=VAL(VG10\$)
 820 VG(I,1)=VG1:VG(I,2)=VG2:VG(I,3)=VG3
 830 VG(I,4)=VG4:VG(I,5)=VG5:VG(I,6)=VG6
 840 VG(I,7)=VG7:VG(I,8)=VG8:VG(I,9)=VG9
 850 VG(I,10)=VG10
 860 NEXT I
 870 OPEN AM\$ AS#3 LEN=132
 880 FIELD#3,13 AS VA1\$,13 AS VA2\$,13 AS
 VA3\$,13 AS VA4\$,13 AS VA5\$,13 AS VA6\$,13
 AS VA7\$,13 AS VA8\$,13 AS VA9\$,13 AS VA10\$
 890 FOR I=1 TO HA
 900 GET#3,I
 910 Y3=EOF(3):IF Y3=-1 THEN CLOSE#3:GOTO
 1010
 920 VA1 = VAL(VA1\$): VA2 = VAL(VA2\$): VA3 =
 VAL(VA3\$)
 930 VA4 = VAL(VA4\$): VA5 = VAL(VA5\$): VA6 =
 VAL(VA6\$)
 940 VA7 = VAL(VA7\$): VA8 = VAL(VA8\$): VA9 =
 VAL(VA9\$)
 950 VA10=VAL(VA10\$)
 960 VA(I,1)=VA1:VA(I,2)=VA2:VA(I,3)=VA3
 970 VA(I,4)=VA4:VA(I,5)=VA5:VA(I,6)=VA6
 980 VA(I,7)=VA7:VA(I,8)=VA8:VA(I,9)=VA9
 990 VA(I,10)=VA10
 1000 NEXT I
 1010 REM *****SALIDAS POR IMPRESORA

 1020 GLGL=(K-1)*(HA-1):QW=K-1
 1030 GLDGL=(K-1)*(HA-2)
 1040 GENO1=GENO/(K-1)
 1050 LOCAL1=LOCAL/(HA-1)
 1060 INLG1 = INLG/GLGL:GLPC1 = K+HA-3:
 GLPC2 = GLPC1-2: GLPC3 = GLPC2-2:
 MSPCA1 = SSPCA1/GLPC1: MSPCA2 =
 SSPCA2/GLPC2: MSPCA3 = SSPCA3/GLPC3
 1070 GLPC4 = GLPC3-2: GLPC5 = GLPC4-2:
 GLPC6 = GLPC5-2: MSPCA4 = SSPCA4/ GLPC4:
 MSPCA5 = SSPCA5/GLPC5: MSPCA6 =
 SSPCA6/GLPC6
 1080 GLRES = GLGL-GLPC1-GLPC2-GLPC3-
 GLPC4-GLPC5-GLPC6:SSRES = INLG-
 SSPCA1- SSPCA2-SSPCA3-SSPCA4-SSPCA5-
 SSPCA6:MSRES = SSRES/GLRES
 1090 REM ***** CALCULO DE LAS COORDE-
 NADAS *****
 1100 FOR I=1 TO K
 1110 FOR H=1 TO 10
 1120 COOGEN(I,H)=VG(I,H)*LAMBDA(H)
 1130 NEXT H:NEXT I
 1140 FOR J=1 TO HA:FOR E=1 TO 10
 1150 COOAM(J,E)=VA(J,E)*(LAMBDA(E))*(-1)
 1160 NEXT E:NEXT J

```

1170 REM ***** DISPLAY DE LAS COORDENA-
DAS*****
1180 LPRINT "***** SALIDA DEL PROGRAMA
AMMI *****"
1190 LPRINT:LPRINT " A N O V A DEL ANALI-
S A M M I "
1200 LPRINT
"#####"
1210 LPRINT
1220 LPRINT "F. DE V. G.L. S.C.
M.C. F "
1230 LPRINT "=====
=====
===== "
1240 LPRINT
1250 LPRINT "GENOTIPOS";LPRINT USING
"#####.#"; K-
1,GENO,GENO1,GENO1/INLG1
1260 LPRINT
1270 LPRINT "MACROAMB.";LPRINT USING
"#####.#"; HA-
1,LOCAL,LOCAL1,LOCAL1/ERROR2
1280 LPRINT
1290 LPRINT "GEN x LOC";LPRINT USING
" # # # # # # # # # # # # . # " ;
GLGL,INLG,INLG1,INLG1/ERROR2
1300 LPRINT " PCA1 ";LPRINT USING
" # # # # # # # # # # # # . # " ;
GLPC1,SSPCA1,MSPCA1,MSPCA1/ERROR2
1310 LPRINT " PCA2 ";LPRINT USING
" # # # # # # # # # # # # . # " ;
GLPC2,SSPCA2,MSPCA2,MSPCA2/ERROR2
1320 LPRINT " PCA3 ";LPRINT USING
" # # # # # # # # # # # # . # " ;
GLPC3,SSPCA3,MSPCA3,MSPCA3/ERROR2
1330 LPRINT " PCA4 ";LPRINT USING
" # # # # # # # # # # # # . # " ;
GLPC4,SSPCA4,MSPCA4,MSPCA4/ERROR2
1340 LPRINT " PCA5 ";LPRINT USING
" # # # # # # # # # # # # . # " ;
GLPC5,SSPCA5,MSPCA5,MSPCA5/ERROR2
1350 LPRINT " PCA6 ";LPRINT USING
" # # # # # # # # # # # # . # " ;
GLPC6,SSPCA6,MSPCA6,MSPCA6/ERROR2
1360 LPRINT "RESIDUAL ";LPRINT USING
" # # # # # # # # # # # # . # " ;
GLRES,SSRES,MSRES,MSRES/ERROR2
1370 LPRINT
1380 LPRINT "ERROR ";LPRINT USING
"#####.#";ER,ERROR1,ERROR2
1390 LPRINT " S.S. DE LOS PCA7 AL PCA10 ="
1400 LPRINT USING
"#####.#";SSPCA7,SSPCA8,SSPCA9,SSPC
A10
1410 LPRINT:LPRINT:LPRINT:LPRINT "*****
VALORES PROPIOS : ";LPRINT
USING"#####.#";LAMBDA(1);LAMBDA(2);LA

```

```

MBDA(3);LAMBDA(4);LAMBDA(5); LAMB-
DA(6):LPRINT:LPRINT
1420 LPRINT "***** VALORES PROPIOS :
":LPRINT USING "#####.#";
LAMBDA(7);LAMBDA(8);LAMBDA(9);LAMB-
DA(10):LPRINT:LPRINT
1430 LPRINT "COORDENADAS PARA LOS
DOCE EJES DEL AMMI"
1440 LPRINT
"#####
#####"
1450 LPRINT:LPRINT "GENOTIPOS PCA1
PCA2 PCA3 PCA4 PCA5 PCA6
PCA7 PCA8 PCA9 PCA10 "
1460 FOR I=1 TO K
1470 LPRINT USING
"#####.#";I;COOGEN(I,1);COOGEN(I,2);CO
OGEN(I,3); COOGEN(I,4);COOGEN(I,5);COO-
GEN(I,6);
1480 LPRINT USING
"#####.#";COOGEN(I,7);COOGEN(I,8);COO
GEN(I,9); COOGEN(I,10)
1490 NEXT I
1500 LPRINT:LPRINT:LPRINT "AMBIENTES
PCA1 PCA2 PCA3 PCA4 PCA5
PCA6 PCA7 PCA8 PCA9 PCA10 "
1510 FOR H=1 TO HA
1520 LPRINT USING
"#####.#";H;COOAM(H,1);COOAM(H,2);CO
O A M ( H , 3 ) ;
COOAM(H,4);COOAM(H,5);COOAM(H,6);
1530 LPRINT USING
"#####.#";COOAM(H,7);COOAM(H,8);COO
AM(H,9); COOAM(H,10)
1540 NEXT H
1550 INPUT "NOMBRE ARCHIVO COO-GEN
";COG$
1560 INPUT "NOMBRE ARCHIVO COO-AMB
";COA$
1570 OPEN COG$ FOR OUTPUT AS#4
1580 FOR I=1 TO K:PRINT#4, USING
"#####.#";COOGEN(I,1);COOGEN(I,2);
COOGEN(I,3);COOGEN(I,4);COOGEN(I,5);COO
GEN(I,6);COOGEN(I,7);COOGEN(I,8);COO-
GEN(I,9);COOGEN(I,10);RENDG(I):NEXT I
1590 CLOSE#4
1600 OPEN COA$ FOR OUTPUT AS#5
1610 FOR I=1 TO HA:PRINT#5, USING
"#####.#";COOAM(I,1);COOAM(I,2);
COOAM(I,3);COOAM(I,4);COOAM(I,5);COOAM(
I , 6 ) ; C O O A M ( I , 7 ) ; C O O A M ( I , 8 ) ;
COOAM(I,9);COOAM(I,10);RENDAM(I):NEXT
I:CLOSE#5
1620 END
1630 FOR W=1 TO
7:FINH=INT((3000*RND(1))+1000):SOUND

```



FINH,1:NEXT W:PRINT "PULSA UNA TECLA
...":WHILE INKEY\$="" :WEND:RETURN

16.6.4. Programa realizado en BASIC para el cálculo del estadístico RMS PD. Se obtienen los valores del RMS PD para los seis primeros ejes, para el modelo con sólo efectos aditivos y para el modelo completo.

```

10 CLS
15 REM PROGRAMA RMSPD PARA MODELO
COMPLETO
20 DIM A(15,13),B(12,13):K=11:HA=10
30 GOSUB 1200:INPUT "MODELO Num. ";NUM
40 INPUT " NOMBRE DEL FICHERO DE COOR-
DENADAS GENETICAS ";A$
50 INPUT " NOMBRE DEL FICHERO DE COOR-
DENADAS AMBIENTALES ";B$
60 OPEN A$ AS#1 LEN=123
70 FIELD#1,11 AS V1$,11 AS V2$,11 AS V3$,11
AS V4$,11 AS V5$,11 AS V6$,11 AS V7$,11 AS
V8$,11 AS V9$,11 AS V10$,11 AS V11$
80 FOR I=1 TO K
90 GET#1,I
100 Y=EOF(1):IF Y=-1 THEN CLOSE#1: GOTO
160
110 V1 = VAL(V1$): V2 = VAL(V2$): V3 =
VAL(V3$): V4 = VAL(V4$)
120 V5 = VAL(V5$): V6 = VAL(V6$): V7 =
VAL(V7$)
121 V8 = VAL(V8$): V9 = VAL(V9$): V10 =
VAL(V10$): V11 = VAL(V11$)
130 A(I,1)=V1:A(I,2)=V2:A(I,3)=V3:A(I,4)=V4
140 A(I,5)=V5:A(I,6)=V6:A(I,7)=V7
141 A(I,8) = V8: A(I,9) = V9: A(I,10) = V10: A(I,11)
= V11
150 NEXT I
160 REM
170 OPEN B$ AS#2 LEN=123
180 FIELD#2,11 AS B1$,11 AS B2$,11 AS
B3$,11 AS B4$,11 AS B5$,11 AS B6$,11 AS
B7$,11 AS B8$,11 AS B9$,11 AS B10$,11 AS
B11$
190 FOR I=1 TO HA
200 GET#2,I
210 Y1=EOF(2):IF Y1=-1 THEN CLOSE#2:
GOTO 270
220 B1 = VAL(B1$): B2 = VAL(B2$): B3 =
VAL(B3$): B4 = VAL(B4$)
230 B5=VAL(B5$):B6=VAL(B6$):B7=VAL(B7$)
231 B8 = VAL(B8$): B9 = VAL(B9$): B10 =
VAL(B10$): B11 = VAL(B11$)
240 B(I,1)=B1:B(I,2)=B2:B(I,3)=B3:B(I,4)=B4
250 B(I,5)=B5:B(I,6)=B6:B(I,7)=B7
251 B(I,8) = B8: B(I,9) = B9: B(I,10) = B10:

```

```

B(I,11) = B11
260 NEXT I
270 FOR I=1 TO K
280 SUM=SUM+A(I,11)
290 NEXT I
300 MEDIA=SUM/K
310 REM ** CALCULO DE LOS VALORES ADITI-
VOS
320 DIM ADIT(12,15)
330 FOR I=1 TO K:FOR H=1 TO HA
340 ADIT(H,I)=A(I,11)+B(H,11)-MEDIA
350 NEXT H:NEXT I
360 REM ***** CALCULO DE LAS COORDENA-
DAS
370 DIM COOT(12,12,15),ESTDO(12,180)
380 FOR T=1 TO HA:FOR I=1 TO K:FOR H=1 TO
HA
390 COOT(T,H,I)=A(I,T)*B(H,T)
400 NEXT H:NEXT I:NEXT T
410 REM ***ESTIMADOS MODELO AMMI
420 FOR S=1 TO HA:CONT1=0:FOR H=1 TO
HA:FOR I=1 TO K
430 CONT1=CONT1+1
440 ESTDO(S,CONT1)=COOT(S,H,I)
450 NEXT I:NEXT H:NEXT S
460 GOSUB 1200:INPUT "NOMBRE DEL MODE-
LO DE VALIDACION ";VL$
470 DIM VALI(180),AMMI(12,180)
480 OPEN VL$ FOR INPUT AS#3
490 IF EOF(3) THEN CLOSE#3:GOTO 530
500 VL=VL+1:INPUT#3,VL1$
510 VALI(VL)=VAL(VL1$)
520 GOTO 490
530 REM ***** RMSPD DE LA PARTE ADITIVA
540 DIM AMMI0(180):FOR I=1 TO HA:FOR T=1
TO K:CD=CD+1
550 AMMI0(CD)=ADIT(I,T)
560 DIF7=AMMI0(CD)-VALI(CD)
570 DIF72=DIF7*DIF7:SDIF7=SDIF7+DIF72
580 NEXT T:NEXT I
590 RMS7=SDIF7/(110):RMSPD(7)=SQR(RMS7)
600 REM **** CALCULO DE LOS RMSPD DEL
MODELO
610 REM RMSPD 1
620 FOR I=1 TO 110
630 AMMI(1,I)=ESTDO(1,I)
640 AMMI(2,I)=ESTDO(1,I)+ESTDO(2,I)
650 AMMI(3,I) = ESTDO(1,I) + ESTDO(2,I) +
ESTDO(3,I)
660 AMMI(4,I) = ESTDO(1,I) + ESTDO(2,I) +
ESTDO(3,I) + ESTDO(4,I)
670 AMMI(5,I) = ESTDO(1,I) + ESTDO(2,I) +
ESTDO(3,I) + ESTDO(4,I) + ESTDO(5,I)
680 AMMI(6,I) = ESTDO(1,I) + ESTDO(2,I) +
ESTDO(3,I) + ESTDO(4,I) + ESTDO(5,I) +
ESTDO(6,I)

```



```

6          8          1
AMMI(7,I)=ESTDO(7,I)+ESTDO(8,I)+ESTDO(9,I)
+ESTDO(10,I)
682 AMMI(8,I)=AMMI(6,I)+AMMI(7,I)
690 DIF1=(AMMI0(I)+AMMI(1,I))-VALI(I)
700 DIF2=(AMMI0(I)+AMMI(2,I))-VALI(I)
710 DIF3=(AMMI0(I)+AMMI(3,I))-VALI(I)
720 DIF4=(AMMI0(I)+AMMI(4,I))-VALI(I)
730 DIF5=(AMMI0(I)+AMMI(5,I))-VALI(I)
740 DIF6=(AMMI0(I)+AMMI(6,I))-VALI(I)
741 DIF9=(AMMI0(I)+AMMI(8,I))-VALI(I)
750 DIF12=DIF1*DIF1:SDIF1=SDIF1+DIF12
751 DIF92=DIF9*DIF9:SDIF9=SDIF9+DIF92
760 DIF22=DIF2*DIF2:SDIF2=SDIF2+DIF22
770 DIF32=DIF3*DIF3:SDIF3=SDIF3+DIF32
780 DIF42=DIF4*DIF4:SDIF4=SDIF4+DIF42
790 DIF52=DIF5*DIF5:SDIF5=SDIF5+DIF52
800 DIF62=DIF6*DIF6:SDIF6=SDIF6+DIF62
810                                     NEXT
I:RMS1=SDIF1/110:RMSPD(1)=SQR(RMS1):RE
M *** 110=NUM. DATOS VALIDACION *****
820 RMS2=SDIF2/(110):RMSPD(2)=SQR(RMS2)
830 RMS3=SDIF3/(110):RMSPD(3)=SQR(RMS3)
840 RMS4=SDIF4/(110):RMSPD(4)=SQR(RMS4)
850 RMS5=SDIF5/(110):RMSPD(5)=SQR(RMS5)
860 RMS6=SDIF6/(110):RMSPD(6)=SQR(RMS6)
870 RMS9=SDIF8/(110):RMSPD(8)=SQR(RMS8)
1040 LPRINT " RMS PD PARA EL MODELO Num.
";NUM
1050 LPRINT " DEL 1-6 SEGUN EJES ADICIO-
NADOS AL MODELO"
1060 LPRINT " 7 = SOLO FACTORES ADITI-
VOS."
1070 LPRINT " 8 = MODELO COMPLETADO
CON TODOS LOS EJES."
1080 LPRINT:LPRINT:LPRINT
1090 FOR I=1 TO 9
1100 LPRINT "RMS PD (";I;") = ";RMSPD(I)
1110 NEXT I
1120 LPRINT:LPRINT:LPRINT
1121 PRINT " RMS PD PARA EL MODELO Num.
";NUM
1122 PRINT " DEL 1-6 SEGUN EJES ADICIONA-
DOS AL MODELO"
1123 PRINT " 7 = SOLO FACTORES ADITIVOS."
1124 PRINT " 8= MODELO COMPLETADO CON
TODOS LOS EJES."
1125 PRINT:PRINT:PRINT
1126 FOR I=1 TO 8
1127 PRINT "RMS PD (";I;") = ";RMSPD(I)
1128 NEXT I
1130 END
1200          FOR          W=1          TO
2:FINH=INT((3000*RND(1))+1000):SOUND
FINH,1:NEXT W:RETURN
3000 PRINT "PULSA TECLA "":WHILE
    
```

```
INKEY$="":WEND:RETURN
```

16.7. Análisis de grupos según Fox y Rosielle

16.7.1. Corriente de control, para el paquete estadístico SAS, realizada por el Dr. Ignacio Romagosa para la obtención del Análisis de grupos de genotipos y ambientes según el método de Fox y Rosielle.

```

* PROGRAMA F&R.SAS;
options ps=66 noNUMBER;
DATA yldmeans;
  INFILE 'DATCOMP1.TXT';
  INPUT SITE $ GENOTYPE $ BLQ YLD;
  STDYLD=YLD;
  IF GENOTYPE ='local' THEN DELETE;
RUN;
PROC SORT;
  BY SITE GENOTYPE;
RUN;
PROC MEANS NOPRINT;
  BY SITE GENOTYPE;
  VAR YLD stdyld;
  OUTPUT OUT=SITEGENO MEAN = YLD
STDYLD;
RUN;
PROC STANDARD NOPRINT DATA=SITEGENO
MEAN=0 STD=1 OUT=STAND;
  BY SITE;
  VAR STDYLD;
RUN;
DATA STDMEANS;
  SET STAND;
  KEEP SITE GENOTYPE STDYLD;
RUN;
PROC SORT DATA=STDMEANS;
  BY SITE;
RUN;
PROC TRANSPOSE OUT=TRANSTD;
  BY SITE;
  ID GENOTYPE;
RUN;
PROC CLUSTER METHOD=WARD
DATA=TRANSTD;
  ID SITE;
PROC TREE HORIZONTAL SPACES=2;
  ID SITE;
RUN;
data rawmeans;
  set sitegeno;
  keep site genotype yld;
run;
PROC SORT;
    
```



```

BY GENOTYPE;
RUN;
PROC TRANSPOSE OUT=TRANSRAW;
  BY GENOTYPE;
  ID SITE;
RUN;
PROC CLUSTER METHOD = WARD DATA =
TRANSRAW;
    
```

```

ID GENOTYPE;
PROC TREE HORIZONTAL SPACES=2;
ID GENOTYPE;
RUN;
    
```

16.7.2. Salida del programa F&R.SAS.



Análisis de grupos para el conjunto de 10 ambientes y 11 genotipos
Ward's Minimum Variance Cluster Analysis
Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	1.69045	0.735780	0.429579	0.42958
2	0.95467	0.322249	0.242602	0.67218
3	0.63242	0.319706	0.160712	0.83289
4	0.31272	0.103467	0.079468	0.91236
5	0.20925	0.124859	0.053175	0.96554
6	0.08439	0.051290	0.021446	0.98698
7	0.03310	0.019322	0.008412	0.99539
8	0.01378	0.009438	0.003502	0.99890
9	0.00434	0.004343	0.001104	1.00000
10	-0.00000	0.000000	-0.000000	1.00000
11	-0.00000	•	-0.000000	1.00000

Root-Mean-Square Total-Sample Standard Deviation = 0.598114
Root-Mean-Square Distance Between Observations = 2.805403



Number of Clusters	Clusters	Joined	Frequency of New Cluster	Semipartial R-Squared	R-Squared	Tie
9	8.000	3.000	2	0.013263	0.986737	
8	CL9	9.000	3	0.023550	0.963187	
7	5.000	7.000	2	0.025943	0.937244	
6	10.000	2.000	2	0.087513	0.849731	
5	6.000	1.000	2	0.090093	0.759638	
4	CL5	4.000	3	0.114509	0.645129	
3	CL7	CL8	5	0.116282	0.528846	
2	CL3	CL6	7	0.191630	0.337216	
1	CL2	CL4	10	0.337216	0.000000	





16.9.2. Corriente de control diseñada para la realización del análisis canónico con el paquete estadístico SAS.

```
* PROGRAMA CANON.SAS;
OPTIONS PS=80 NONUMBER;
DATA FER;
TITLE 'ANALISIS CANONICO DE LOS 11 GENO-
TIPOS';
INFILE 'A:AMB10.DAT';
INPUT GRUP A1 A2 A3 A4 A5 A6 A7 A8 A9 A10;
RUN;
PROC CANDISC DATA=FER OUT=OUTCAN
DISTANCE ANOVA;
CLASS GRUP;
VAR A1 A2 A3 A4 A5 A6 A7 A8 A9 A10;
RUN;
PROC PLOT;
    PLOT CAN2*CAN1=GRUP;
format grup specchar.;
TITLE 'PLOT DE LAS VARIABLES CANONICAS';
RUN;
PROC PRINT DATA=OUTCAN;
RUN;
```



16.9.3. Salida del programa CANON.SAS.

**ANÁLISIS CANÓNICO DE LOS 11 GENOTIPOS
CANONICAL DISCRIMINANT ANALYSIS**

11 Observations 14 DF Total

10 Variables 12 DF Within Classes
3 Classes 2 DF Between Classes

Class Level Information

GRUP	Frequency	Weight	Proportion
1	5	5.0000	0.454545
2	5	5.0000	0.454545
3	1	1.0000	0.090909

**CANONICAL DISCRIMINANT ANALYSIS
PAIRWISE SQUARED DISTANCES
BETWEEN GROUPS**

$$D(ij) = (\bar{X}_i - \bar{X}_j)' \text{COV}^{-1} (\bar{X}_i - \bar{X}_j)$$

Squared Distance to GRUP

From GRUP	1	2	3
1	0	3732	14567
2	3732	0	3935
3	14567	3935	0

CANONICAL DISCRIMINANT ANALYSIS

Univariate Test Statistics

F Statistics, Num DF= 2 Den DF= 12

Variable	Total STD	Pooled STD	Between STD	R-Squared	RSQ/ (1-RSQ)
A1	346.5833	343.3330	163.4471	0.158859	0.1889
A2	57.1106	342.8615	193.5830	0.209894	0.2657
A3	633.0104	267.0736	689.4846	0.847421	5.5540
A4	348.0440	358.0039	125.6517	0.093098	0.1027
A5	577.0900	473.5956	443.9535	0.422727	0.7323
A6	372.4527	265.8305	330.7724	0.563363	1.2902
A7	325.8896	249.7967	271.6755	0.496400	0.9857
A8	893.9598	339.6482	990	0.876270	7.0821
A9	1383	399.8689	1576	0.928304	12.9477
A10	395.6833	287.8715	346.0454	0.546314	1.2042



Univariate Test Statistics



F Statistics, Num DF= 2 Den DF= 12

Variable	F	Pr > F
A1	1.1332	0.3542
A2	1.5939	0.2433
A3	33.3240	0.0001
A4	0.6159	0.5564
A5	4.3937	0.0370
A6	7.7414	0.0069
A7	5.9142	0.0163
A8	42.4926	0.0001
A9	77.6864	0.0001
A10	7.2250	0.0087



Average R-Squared: Unweighted = 0.514265
Weighted by Variance = 0.7660442

Multivariate Statistics and F Approximations



S=2 M=3.5 N=0.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.00000608	121.3262	20	6	0.0001
Pillai's Trace	1.98120027	42.1538	20	8	0.0001
Hotelling-Lawley Trace	3088.0329837	308.8033	20	4	0.0001
Roy's Greatest Root	3034.8921297	1213.957	10	4	0.0001



NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

CANONICAL DISCRIMINANT ANALYSIS



	Adjusted Canonical Correlation	Approx Canonical Correlation	Squared Standard Error	Canonical Correlation
1	0.999835	.	0.000088	0.999671
2	0.990722	.	0.004936	0.981530

Eigenvalues of INV(E)*H
= CanRsqr/(1-CanRsqr)

	Eigenvalue	Difference	Proportion	Cumulative
1	3034.892	2981.751	0.9828	0.9828
2	53.1409	.	0.0172	1.0000



Test of H0: The canonical correlations in the current row and all that follow are zero

	Likelihood Ratio	Approx F	Num DF	Den DF	Pr > F
1	0.00000608	121.3262	20	6	0.0001
2	0.01847034	23.6182	9	4	0.0040





Total Canonical Structure

	CAN1	CAN2
A1	0.360129	0.172505
A2	0.026176	-0.461678
A3	0.919386	0.049758
A4	-0.283411	0.114209
A5	0.650165	0.012442
A6	0.672403	0.336871
A7	0.696601	-0.107333
A8	0.936152	0.013490
A9	0.952928	0.144628
A10	0.738274	0.038377



Between Canonical Structure

	CAN1	CAN2
A1	0.903403	0.428793
A2	0.057126	-0.998367
A3	0.998565	0.053551
A4	-0.928698	0.370836
A5	0.999820	0.018958
A6	0.895703	0.444653
A7	0.988545	-0.150927
A8	0.999898	0.014277
A9	0.988880	0.148716
A10	0.998676	0.051441



Pooled Within Canonical Structure

	CAN1	CAN2
A1	0.007127	0.025563
A2	0.000534	-0.070588
A3	0.042718	0.017312
A4	-0.005401	0.016299
A5	0.015531	0.002225
A6	0.018468	0.069285
A7	0.017815	-0.020555
A8	0.048302	0.005212
A9	0.064591	0.073407
A10	0.019893	0.007743



CANONICAL DISCRIMINANT ANALYSIS

Standardized Canonical Coefficients

	CAN1	CAN2
A1	-22.98633981	8.86339199
A2	2.78650407	-1.25239488
A3	15.64427353	-17.23712010
A4	-16.19827011	10.11874777
A5	20.00525124	-11.33907624
A6	16.97664096	-8.68925309
A7	-13.84216015	-7.11100919
A8	37.98546656	-12.35761614
A9	-11.75550768	47.57845068
A10	1.49746051	-0.06180503



Raw Canonical Coefficients

	CAN1	CAN2
A1	-0.663226920	0.0255736242
A2	0.0078029158	-0.0035070222
A3	0.0247140880	-0.0272303922
A4	-0.0465408681	0.0290731852
A5	0.0346657374	-0.0196487130
A6	0.0455806622	-0.0233298160
A7	-0.0424749994	-0.0218203017
A8	0.0424912486	-0.0138234590
A9	-0.0085024725	0.0344123351
A10	0.0037844923	-0.0001561982

Class Means on Canonical Variables

GRUP	CAN1	CAN2
1	60.06071991	-4.67566515
2	0.57034263	9.22061118
3	-60.63106254	-4.54494603



GEN	GRUP	A1	A2	A3	A4	A5	A6	A7	A8	A9
1	2	3850.0	3541.7	3308.3	3868.1	3097.9	3972.2	3951.4	4340.3	4494.5
2	1	3552.8	3750.0	3791.7	3254.2	3266.0	3375.0	4201.4	5208.3	5386.1
3	3	3352.8	3819.4	2327.8	3547.2	2161.1	3000.0	3430.6	3055.6	1962.5
4	2	3334.7	2847.2	2720.8	3176.4	1854.9	3430.5	3097.2	3715.3	3012.5
5	1	3762.5	4375.0	3791.7	3309.7	2993.1	3430.6	3888.9	5173.6	4938.9
6	1	4147.2	3819.5	3745.8	3263.9	3128.5	3902.8	3895.8	5277.8	5008.4
7	1	4037.5	4027.8	4037.5	4130.6	3808.3	4152.8	4229.2	5416.7	5445.8
8	1	2723.6	3263.9	3177.8	2661.1	2052.1	3319.4	3722.2	4097.2	4486.1
9	2	3638.9	3402.8	3027.8	3154.2	2418.0	3416.7	3534.7	3958.3	3793.1
10	2	3666.7	3680.6	3550.0	3691.7	2691.0	3375.0	3645.9	4236.1	4038.9
11	2	3615.3	4027.8	2865.3	3709.7	3071.6	3805.6	4027.8	4166.7	4283.3



GEN	A10	CAN1	CAN2	CAN3	CAN4	CAN5	CAN6	CAN7	CAN8	CAN9	CAN10
1	2569.5	1.5779	10.7940
2	2465.3	60.3277	-4.6741
3	2118.1	-62.0127	-2.3757
4	2013.9	0.4503	9.3668
5	3298.6	60.5425	-4.0455
6	2881.9	59.8744	-5.1519
7	3125.0	59.5503	-4.8038
8	2430.6	60.0087	-4.7030
9	2861.1	0.4336	9.2749
10	2604.2	0.2292	8.2596
11	2604.2	0.1606	8.4078

16.10. Modelo GGE biplot

16.10.1.- Corriente de control Yan2000.sas, para el paquete estadístico SAS, realizada por los doctores Juan Burgueño, José Crossa y Mateo Vargas, de la Unidad de Biometría y Estadística del CIM-MYT.

OPTIONS PS = 5000 LS=78 NODATE;

FILENAME BIPLLOT 'c:\yan2000.CGM'; /*Nombre del archivo gráfico de salida para el biplot*/
 GOPTIONS DEVICE=CGMMWWC
 GSFNAME=BIPLLOT gsfmode=replace;
 DATA RAW;

INFILE 'c:\medias.txt'; /* Archivo que debemos centrar ambientalmente*/

```

INPUT ENV $ GEN $ YLD;
if env=1 THEN YLD=(YLD-3607.4)/73.51;
if env=2 THEN YLD=(YLD-3686.9)/85.28;
if env=3 THEN YLD=(YLD-3304)/100.77;
if env=4 THEN YLD=(YLD-3433.3)/82.98;
if env=5 THEN YLD=(YLD-2776.6)/125.7;
if env=6 THEN YLD=(YLD-3561.9)/77.89;
if env=7 THEN YLD=(YLD-3784.1)/80.07;
if env=8 THEN YLD=(YLD-4422.3)/147.53;
if env=9 THEN YLD=(YLD-4259.1)/192.23;
if env=10 THEN YLD=(YLD-2633.8)/97.98;
PROC GLM DATA=RAW OUTSTAT=STATS ;
CLASS ENV GEN;
MODEL YLD = ENV GEN ENV*GEN/SS4;
DATA STATS2;
SET STATS ;
DROP _NAME_ _TYPE_;
IF _SOURCE_ = 'ERROR' THEN DELETE;
MSE=137025 ; /* Cuadrados medios del error del ANOVA*/
DFE=200 ; /* Grados de libertad del MSE*/
NREP=3 ; /* Número de repeticiones del ensayo*/
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;
PROC GLM DATA=RAW NOPRINT;
CLASS ENV;
MODEL YLD = ENV/ 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;
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);
MINIMO=MIN(NGEN,NENV);
L=L[1:MINIMO,];
SS=(L##2)*NREP;
SUMA=SUM(SS);
PORCENT=((1/SUMA)#SS)*100;
MINIMO=MIN(NGEN,NENV);
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;
DFE=J(MINIMO,1,DFEM);
MSE=J(MINIMO,1,MSEM);
SSDF=SSIIIPORCENTIIPORCENACIIDFAIIDFEIIMSE;
L12=L##0.5;
SCOREG1=U[,1]#L12[1,];
SCOREG2=U[,2]#L12[2,];
SCOREG3=U[,3]#L12[3,];
SCOREE1=V[,1]#L12[1,];
SCOREE2=V[,2]#L12[2,];
SCOREE3=V[,3]#L12[3,];
SCOREG=SCOREG1||SCOREG2||SCOREG3;
SCOREE=SCOREE1||SCOREE2||SCOREE3;
SCORES=SCOREG//SCOREE;
CREATE SUMAS FROM SSDF;
APPEND FROM SSDF;
CLOSE SUMAS;
CREATE SCORES FROM SCORES;
APPEND FROM SCORES ;
CLOSE SCORES;

/*Obtencion del poligono y perpendiculares */
d1=scoreg[,1:2][cvexhull(scoreg[,1:2])[loc(cvexhull(scoreg[,1:2])>0),,]];
d=d1//d1[1,];
xxx=J(nrow(d)-1,1,0);
yyy=J(nrow(d)-1,1,0);
ppp={0 1,1 0};
do i=1 to nrow(d)-1 ;
  dd=d[i,i+1,];
  if dd[1,1]>dd[2,1] then ddd=ppp*dd;
  else ddd=dd;
  p=(ddd[2,2]-ddd[1,2])/(ddd[2,1]-ddd[1,1]) ;
  if p<0 then ss=1 ;
  else ss=-1 ;
  r=tan((180-90-abs(atan(p))*180/3.14156))*3.14156/180)*ss ;
  aa=(ddd[1,2]+ddd[2,2])/2-p*(ddd[1,1]+ddd[2,1])/2;
  xx=aa/(r-p) ;
  /*xx=(p*d[i,1]-d[i,2])/(p-r) ;*/
  if abs(r)<1 then xxx[i,]=3.5; /* máximo para ejes ACP*/

```



```

else xxx[i,]=3.5/abs(r); /* máximo para ejes ACP*/
    if xx<0 then xxx[i,]=-xxx[i,] ;
    else xxx[i,]=xxx[i,];
yyy[i,]=xxx[i,]*r;
end;
kk=xxxllyyy;
xx1={V1 V2};
create pol from d[colNAME=xx1];
append from d ;
close pol;
xx2={V3 V4};
create perp from kk[colNAME=xx2];
append from kk ;
close perp;
data pol; set pol; TYPE="pol";
data perp; set perp; TYPE="per";
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);
PROC PRINT DATA=SSAMMI NOOBS;
    VAR SSAMMI PORCENT PORCENAC DFAMMI MSAMMI F_AMMI PROBF;
PROC SORT DATA=RAW;
    BY GEN;
PROC MEANS DATA = RAW NOPRINT;
    BY GEN ;
    VAR YLD;
    OUTPUT OUT = MEDIAG MEAN=YLD;
DATA NAMEG;
    SET MEDIAG;
    TYPE = 'GEN';
    NAME = 'G'IIIGEN;
    KEEP TYPE NAME YLD;
PROC SORT DATA=RAW;
    BY ENV;
PROC MEANS DATA = RAW NOPRINT;
    BY ENV ;
    VAR YLD;
    OUTPUT OUT = MEDIAE MEAN=YLD;
DATA NAMEE;
    SET MEDIAE;
    TYPE = 'ENV';
    NAME = 'A'IIENV;
    KEEP TYPE NAME YLD;
DATA NAMETYPE;
    SET NAMEG NAMEE;
DATA BIPL0T0 ;
    MERGE NAMETYPE SCORES;

```

Σ

```

DIM1=COL1;
DIM2=COL2;
DIM3=COL3;
DROP COL1-COL3;

```

```

data biplot ;
set biplot0 pol perp;
PROC PRINT DATA=BIPLOT NOOBS;
  VAR TYPE NAME YLD DIM1 DIM2 DIM3;

```

 α

```

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

```

 Π

```

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

```

 Ω

```

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;

```

 \times

```

if type = "per" then do;
  color='green';
  line=2;
  size = 0.6;
  style = 'hwcgm001';
  x=0.0;

```

 \pm

```

  y=0.0;
  function='MOVE';
  output;
  x=v3;
  y=v4;
  function='DRAW';
  output;

```

 ϑ



```

end;
Proc gplot data=biplot;
Plot dim2*dim1 v2*v1 /overlay 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=blue i=j line=3 ;
axis2
  length = 6.0 in
  order = (-3.5 to 3.5 by 0.5) /* máximo y mínimo eje ACP2*/
  label=(f=hwcm001 h=1.2 a=90 r=0 'ACP 2')
  value=(h=0.8)
  minor=none;
axis1
  length = 6.0 in
  order = (-3.5 to 3.5 by 0.5) /* máximo y mínimo eje ACP1*/
  label=(f=hwcm001 h=1.2 'ACP 1')
  value=(h=0.8)
  minor=none;
Title1 f=hwcm001 h=1.0 'GGE biplot'; /* Título para el GGE biplot*/
run;

```

16.10.2. Salida generada por el SAS.

La salida gráfica de la corriente de control Yan2000.sas es la Figura 47. Y la salida de texto es lo siguiente:

GGE biplot 6

General Linear Models Procedure
Class Level Information

Class	Levels	Values
ENV	10	1 10 2 3 4 5 6 7 8 9
GEN	11	1 10 11 2 3 4 5 6 7 8 9

Number of observations in data set = 110

GGE biplot

7

General Linear Models Procedure

Dependent Variable: YLD

Source	Sum of DF	Mean Squares	Square	F Value	Pr > F
Model	109	2351.633548	21.574620	•	•
Error	0	•	•		
Corrected Total	109	2351.633548			
	R-Square	C.V.	Root MSE		YLD Mean
	1.000000	0	0		0.000203
Source	DF	Type IV SS	Mean Square	F Value	Pr > F
ENV	9	0.000014	0.000002	•	•
GEN	10	1505.985284	150.598528	•	•
ENV*GEN	90	845.648250	9.396092	•	•

GGE biplot

8

SOURCE	DF	SS	MS	F	PROB
ENV	9	0.00	0.000	.0000000	1
GEN	10	4517.96	451.796	.0032972	1
ENV*GEN	90	2536.94	28.188	.0002057	1

GGE biplot

9

SSAMMI	PORCENT	PORCENAC	DFAMMI	MSAMMI	F_AMMI	PROBF
4666.68	66.1481	66.148	18	259.260	.0018921	1.00000
1105.76	15.6736	81.822	16	69.110	.0005044	1.00000
564.83	8.0062	89.828	14	40.345	.0002944	1.00000
363.29	5.1494	94.977	12	30.274	.0002209	1.00000
176.55	2.5026	97.480	10	17.655	.0001288	1.00000
117.51	1.6656	99.146	8	14.688	.0001072	1.00000
41.01	0.5813	99.727	6	6.835	.0000499	1.00000
11.31	0.1603	99.887	4	2.828	.0000206	1.00000
7.69	0.1091	99.996	2	3.847	.0000281	0.99997
0.27	0.0038	100.000	0	•	•	•

GGE biplot

10

TYPE	NAME	YLD	DIM1	DIM2	DIM3
GEN	1	1.68058	0.74512	1.17922	-1.38505
GEN	10	-0.12282	-0.09904	0.45713	0.14234
GEN	11	0.96928	0.32028	1.24277	0.47948
GEN	2	1.88549	1.19598	-1.56955	0.33652
GEN	3	-5.32244	-2.94713	1.99535	1.69161
GEN	4	-5.76417	-2.89544	-0.05814	-1.88008
GEN	5	3.02883	1.58363	-0.74798	2.03543
GEN	6	3.20410	1.74135	-0.37674	-0.67397
GEN	7	6.48103	3.21257	0.87694	-0.64489
GEN	8	-4.02880	-1.83521	-2.84837	0.06129
GEN	9	-2.00884	-1.02212	-0.15062	-0.16270
ENV	1	0.00074	2.00365	1.76220	-0.63938
ENV	10	0.00056	1.62048	-0.24989	0.83812
ENV	2	-0.00021	1.66503	0.88299	2.81963
ENV	3	0.00045	2.36492	-1.34084	-0.22634
ENV	4	0.00055	1.23219	2.82905	-0.35704
ENV	5	-0.00007	2.29473	0.61807	0.05376
ENV	6	-0.00035	1.65775	0.72331	-1.97774
ENV	7	0.00000	1.86987	-0.26001	0.53438
ENV	8	0.00037	2.37776	-1.21291	-0.19106
ENV	9	-0.00000	2.39843	-1.73373	-0.49462



