

UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE CIENCIAS BIOLÓGICAS Y AGROPECUARIAS
DIVISIÓN DE CIENCIAS AGRONÓMICAS
COORDINACIÓN DE POSGRADO



INTRODUCCIÓN AL ANÁLISIS DE LOS PARÁMETROS DE ESTABILIDAD

MARCIANO EMMANUEL ARGOTE OLIVERA

T E S I S

Presentada como requisito parcial
para obtener el grado de:
**MAESTRO EN CIENCIAS EN
MANEJO DE ÁREAS DE TEMPORAL**

ZAPOPAN, JALISCO, JUNIO DE 1998

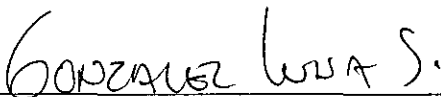
La presente tesis se desarrolló bajo la dirección del Consejo Particular indicado, ha sido aprobada y aceptada como requisito parcial para la obtención del grado de:

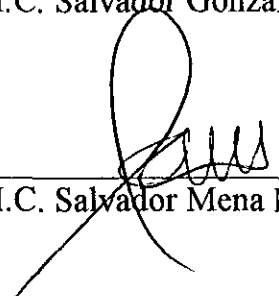
MAESTRO EN CIENCIAS

EN

MANEJO DE ÁREAS DE TEMPORAL.

DIRECTOR: 
Dr. Mario Abel García Vázquez

ASESOR: 
M.C. Salvador González Luna

ASESOR: 
M.C. Salvador Mena Munguia

Zapopan, Jalisco. Mayo de 1998

CONTENIDO

<i>ÍNDICE DE CUADROS</i>	<i>iii</i>
<i>ÍNDICE DE FIGURAS</i>	<i>iv</i>
<i>RESUMEN</i>	<i>v</i>
1. INTRODUCCIÓN	1
2. OBJETIVOS	5
3. ESTABILIDAD GENÉTICA	6
4. PARÁMETROS DE ESTABILIDAD.	8
5. METODOLOGÍA DEL ANÁLISIS	15
5.1 Análisis de Varianza	28
5.1.1 <i>Suma de Cuadrados Total del experimento.</i>	28
5.1.2 <i>Suma de Cuadrados de Variedades.</i>	28
5.1.3 <i>Suma de Cuadrados para Ambientes.</i>	29
5.1.4 <i>Suma de Cuadrados para la interacción Variedad por Ambiente.</i>	29
5.1.5 <i>La suma de cuadrados de las desviaciones ponderadas.</i>	31
5.1.6 <i>La suma de cuadrados del error ponderado.</i>	31
6. APLICACIÓN DEL MÉTODO	38
6.1 Suma de Cuadrados Total.	43
6.2 Suma de Cuadrados de Variedades	44
6.3 Suma de Cuadrados del Residual.	45
6.4 Suma de cuadrados de ambientes (lineal)	45
6.5 Suma de Cuadrados de Variedad por Ambiente (lineal)	48
6.6 Suma de Cuadrados de las Desviaciones Ponderadas	50

6.7	<i>Cuadrado Medio del Error Ponderado.</i>	52
6.8	<i>Cuadrados Medios y valores de "F".</i>	54
6.9	<i>Cálculo de los Coeficientes de Regresión (b_i).</i>	56
6.10	<i>Cálculo de las Desviaciones de Regresión ($S^2_{d_i}$).</i>	57
7.	UN MÉTODO SENCILLO.	59
7.1	<i>Cálculos para la regresión de variedades sobre índices ambientales.</i>	61
7.1.1	<i>Suma de Cuadrados Total para las regresiones.</i>	61
7.1.2	<i>Coeficiente de Determinación (r^2).</i>	62
7.1.3	<i>Coeficiente de Regresión (b_i).</i>	62
7.1.4	<i>Suma de Cuadrados de la Regresión.</i>	63
7.1.5	<i>Suma de Cuadrados de las Desviaciones de Regresión.</i>	63
7.2	<i>Análisis de varianza partiendo de las regresiones.</i>	64
7.2.1	<i>Suma de Cuadrados Total</i>	64
7.2.2	<i>Suma de Cuadrados de Variedades.</i>	64
7.2.3	<i>Suma de Cuadrados de Ambientes (lineal).</i>	65
7.2.4	<i>Suma de Cuadrados de Variedad por Ambiente (lineal).</i>	65
7.2.5	<i>Suma de Cuadrados de las Desviaciones Ponderadas. (SCDP)</i>	65
7.2.6	<i>Cuadrados medios y valores observados de "F".</i>	66
7.2.7	<i>Desviaciones de Regresión ($S^2_{d_i}$).</i>	68
8.	RESULTADOS Y DISCUSIÓN DEL ESQUEMA PRESENTADO.	73
9.	CONCLUSIONES.	77
	BIBLIOGRAFÍA	78

ÍNDICE DE CUADROS

Cuadro 1. <i>Notación estadística utilizada en el presente escrito.</i> _____	17
Cuadro 2. <i>Simbología utilizada.</i> _____	18
Cuadro 3. <i>Cinco variedades probadas en 6 ambientes. Rendimiento en kilogramos por parcela.</i> _____	20
Cuadro 4. <i>Hoja de cálculo del programa "EXCEL".</i> _____	21
Cuadro 5. <i>Análisis de varianza para el ejemplo.</i> _____	31
Cuadro 6. <i>Fórmulas para el Análisis de Varianza</i> _____	39
Cuadro 7. <i>Datos de rendimiento de maíz en kg /Ha. (media de cuatro repeticiones) por localidad, y cálculos para su análisis.</i> _____	40
Cuadro 8. <i>Cálculo del cuadrado medio para el error ponderado (CMEP).</i> _____	53
Cuadro 9. <i>Análisis de varianza del experimento.</i> _____	55
Cuadro 10. <i>Medias y coeficientes de regresión para las variedades</i> _____	58
Cuadro 11. <i>Análisis de regresión de cada variedad sobre los índices ambientales.</i> _____	60

ÍNDICE DE FIGURAS

Figura 1.	<i>Parámetros de la línea recta</i>	15
Figura 2.	<i>Variación total</i>	30
Figura 3.	<i>Variación explicada</i>	30
Figura 4.	<i>Variación no explicada</i>	30
Figura 5.	<i>Diagramas de dispersión y líneas de regresión para el ejemplo.</i>	32
Figura 6.	<i>Diagramas de dispersión y líneas de regresión para el experimento.</i>	69

CUCBA



BIBLIOTECA CENTRAL

RESUMEN

La diversidad genética es lo que permite a las especies adaptarse a las diferentes condiciones que se les presenten o perecer. Los mejoradores de plantas han utilizado esto en favor de la humanidad, consiguiendo cultivos cada vez más vigorosos y rentables. Los cultivos manifiestan su potencial genético a través diferentes ambientes, respondiendo a los factores que lo conforman. La producción es el resultado de los componentes genético y ambiental, pero existe también la interacción genético-ambiental que limita y retarda el trabajo de los mejoradores de cultivos. La detección y cuantificación de la interacción genético-ambiental se vuelve prioritario para la formación de materiales genéticos más estables.

Uno de los primeros métodos estadísticos que se presentó para facilitar lo anterior es el análisis de parámetros de estabilidad de Eberhart y Russell.

El modelo $Y_{ij} = \mu_i + \beta_i I_j + \delta_{ij}$ define parámetros de estabilidad que pueden ser usados para describir el rendimiento de una variedad a través de una serie de ambientes. Y_{ij} es la media varietal de la variedad "i" en el ambiente "j". μ_i es la media de la variedad "i" en todos los ambientes. β_i es el coeficiente de regresión, primer *parámetro de estabilidad*, que mide la respuesta de la variedad "i" a las variaciones de ambientes. δ_{ij} es la desviación de regresión de la variedad "i" en el ambiente "j". I_j es el índice ambiental.

Este método utiliza la regresión de los rendimientos de las variedades a través de ambientes sobre índices ambientales y establece como parámetros de estabilidad al *coeficiente de regresión* (b_i) y a las *desviaciones de regresión* ($S_{d_i}^2$). Eberhart y Russell consideran como variedad "estable" a aquellas que presenten un $b_i = 1$; y consideran como variedad "consistente", a aquellas que presenten $S_{d_i}^2 = 0$.

El índice ambiental ideal, sería aquel que dependiera de valores completamente ajenos al cultivo, esto es, de valores tales como la precipitación, temperatura,

condiciones del suelo, etcétera. Por diversas razones, esto no es posible con el estado actual de la ciencia, por lo que calculan los índices ambientales basados netamente en los rendimientos obtenidos. Esto trae dificultades estadísticas que pueden desaparecer en la medida en que se incremente el número de variedades y de ambientes probados.

Realizando el análisis a partir de los *análisis de regresión* de cada una de las variedades sobre los índices ambientales, se obtienen los mismos resultados que con el método comúnmente utilizado y, desde luego, se llega a las mismas conclusiones. El método es más sencillo por lo que puede ser utilizado sin mayor dificultad. Los cálculos se realizaron en un programa computacional de muy fácil acceso. Con ello, se disminuyen las probabilidades de un error.

1. INTRODUCCIÓN

La diversidad genética de las variedades ocasiona que éstas respondan de diversas maneras ante el medio ambiente y a los factores, en particular, que conforman el ambiente, pudiendo ser mayor o menor la influencia de los factores sobre algunas especies y variedades en particular. Sobre la producción incide el factor genético y el factor ambiental, pero también existe el efecto de la interacción genético-ambiental que es diferente a los dos primeros.

Al probar genotipos en varios ambientes, es necesario evaluar la interacción genotipo-ambiente (GA) con el objeto de tomar medidas en la elección del mejor genotipo o del mejor ambiente para experimentar. En una prueba de variedades a través de varios ambientes, cuando la media de una variedad no mantiene su comportamiento diferencial mas o menos igual que la media de las demás variedades, se dice que hay una interacción genotipo-ambiente. Por ejemplo, si la mayoría de un grupo de variedades presentaron un aumento aproximado de unos 500 Kg. por Ha. al pasar de un ambiente a otro y una de ellas presentó un aumento de solo 50 Kg. por Ha o incluso el rendimiento bajó, se puede decir que hay una interacción manifiesta. Si una variedad aparentemente rendidora está interaccionando en forma significativa con cierto ambiente, solamente se podrá recomendar para tal ambiente y con ciertas limitantes por la consabida limitación de que un ambiente también varía en el eje tiempo. De no interactuar con el ambiente, se podrá hacer la recomendación para condiciones ambientales similares potencialmente hablando. Sin embargo, sí es muy deseable que un cultivo responda a un ambiente favorable con un buen rendimiento.

Si se intentara elaborar la gráfica de la respuesta de una variedad a través de varios ambientes, se tendría que darle valor a ambas variables: como respuesta del cultivo, se puede tomar una característica fenotípica cuantificable como puede ser el rendimiento, altura de planta y días a floración, entre otros; como variable dependiente y para "cuantificar" al ambiente se utilizará un índice ambiental que

podría ser la simple numeración de los ambientes de menor media de producción a mayor media. A cada observación fenotípica le corresponderá un valor ambiental formando pares de datos, (X, Y) , se localizan en un sistema de ejes cartesianos formando, los puntos, lo que se conoce como un diagrama de dispersión. Ajustando una línea recta a los puntos trazados se obtiene una línea de regresión (*Ver Figura 5*).

El índice ambiental ideal sería uno completamente independiente de los rendimientos para que no hubiese problemas de *auto-regresión*¹ cuando se hiciera un análisis estadístico. Tal índice deberá de tomar en cuenta todos los factores que estén determinando al ambiente, como son: la cantidad y distribución de la precipitación, temperatura, condiciones generales del suelo, insolación, evaporación, eje tecnológico etc. Para el diseño de un índice ambiental de éste tipo no se puede pasar por alto la ley del mínimo, o sea que si se tienen muchos factores favorables y solo uno desfavorable, éste toma una gran dimensión, un suelo es tan fértil como su nutriente más pobre. También debe ser un índice para cada cultivo ya que un factor limitante para un cultivo puede no serlo para otro. Ante la enorme dificultad que esto significa, se opta normalmente por tomar un índice ambiental basado en los mismos rendimientos: se resta la media general del experimento a cada una de las medias ambientales, como se verá en forma detallada posteriormente. El error que el índice ambiental así calculado pueda traer, tiende a minimizarse en la medida en que se aumente el número de variedades y ambientes.

En un análisis estadístico de parámetros de estabilidad, se debe definir qué es lo que se debe de buscar, por ejemplo, como definición se podría decir que una variedad estable es la que presente poca o nula interacción con el medio ambiente y que rinda consistentemente en los ambientes probados, sin embargo el hecho de no detectar interacción genotipo-ambiente no significa necesariamente que no la haya, como podría ser el caso de variedades similares en su interacción con el ambiente o,

¹ Se dice que existe una *AUTO-REGRESIÓN* cuando las variables de una regresión son las mismas o dependen directamente una de la otra, como en éste caso el índice ambiental (variable independiente) se está calculando a partir de los rendimientos del conjunto de variedades en estudio (variable dependiente).

ambientes razonablemente homogéneos. Por otro lado, una variedad que responda razonablemente igual en buenos y malos ambientes no es deseable por que no reacciona favorablemente a un buen ambiente en términos de rendimiento. Se puede encontrar variedades que rindan razonablemente bien en todos los ambientes, otras que rindan muy bien en buenos ambientes y mal en ambientes deficientes y aquellas que responden razonablemente bien en malos ambientes pero no responden como se podría esperar en buenos ambientes. Agronómicamente un genotipo estable tiene una interacción mínima con el ambiente pero responde favorablemente en ambientes buenos.

Básicamente, los parámetros de estabilidad indicarán al investigador los genotipos cuya interacción con el ambiente sea mínima, para posteriormente trabajar con ellos y obtener variedades estables de alto rendimiento. Los parámetros de estabilidad auxiliarán, también, en la toma de la decisión de la variedad a recomendar en un ambiente determinado o la elección de un ambiente para evaluar genotipos con la consabida evaluación de riesgos.

Los parámetros de estabilidad de Eberhart y Russell (1966) adolecen de algunas contradicciones matemáticas y limitaciones de tipo estadístico, además resulta difícil para el investigador encontrar una guía mas o menos completa y sencilla para la utilización de los parámetros de estabilidad. Por otro lado, algunos de los experimentos que se analizan con los citados parámetros no fueron diseñados para tal propósito y se toman de diseños con diferentes propósitos.

Se revisaron artículos sobre el tema y presentan cierto grado de dificultad para que se entienda el aspecto matemático y metodológico, incluso para aquellas personas que tienen noción de la estadística, quedando en ocasiones fuera del alcance de los poco avezados en el tema siendo que puede ser una herramienta de provecho.

Se considera muy útil presentar una guía, para el investigador o para el estudiante, en la cual se pueda entender paso por paso el cómo y el porqué de los parámetros calculados al efectuar una prueba de ésta índole, pues comprendiendo los

fundamentos, alcances y limitaciones de la prueba, aunque los cálculos se realicen con una máquina, se tendrá mayor capacidad y autoridad matemática para la interpretación de los parámetros obtenidos e inferir conclusiones más acertadas. No hay que olvidar que lo más importante de la estadística es su interpretación, pues el conocimiento de la teoría de probabilidad y las trampas de la estadística son factores que no se deben dejar de tomar en cuenta. Si las estadísticas indican una cosa y el sentido común otra, el problema es grave. En ocasiones se escucha la opinión de personas, con conocimiento de esta rama de las matemáticas, sugiriendo que las estadísticas sirven para demostrar lo que el investigador quiere demostrar.

En la actualidad los análisis estadísticos son cada vez más complicados, por lo que concebir su cálculo de manera manual es prácticamente imposible, haciéndose obligado el uso de computadoras. Algo que no debe perderse de vista es la rapidez con que paquetes de computación se vuelven “obsoletos”, y las mismas máquinas lo hacen, sin embargo, hay estructuras que han perdurado y solo se van refinando, como programas procesadores de palabras (los de uso más común) y algunas hojas de cálculo que han ido evolucionando y que cada vez son más y más completas. Existen paquetes de programas especiales de estadísticas, pero en ocasiones son caros o de difícil acceso, por lo que resulta valioso aprender a utilizar los programas de uso común en las computadoras para aprovechar su potencial. En la mayoría de los casos las modestas “hojas de cálculo”, incluidas en paquetes de computación básicos, son a menudo menospreciadas por que la lógica indica que tales programas deben ser muy limitados para cálculos especializados como puede ser la estadística inferencial, sin embargo su potencial es amplísimo, y la cantidad de funciones que tienen es verdaderamente asombrosa. Es necesario, pues, introducir a los que trabajan con estadística inferencial en el uso de programas básicos de uso común, ya que están más al alcance y en la mayoría de las computadoras existentes.

2. OBJETIVOS

- I. Analizar bondades y limitaciones de los parámetros de estabilidad de Eberhart y Russell.
- II. Presentar un método sencillo, en comprensión y seguimiento, para el análisis de parámetros de estabilidad de Eberhart y Russell como originalmente lo concibieron los autores.
- III. Desarrollar una metodología alternativa rápida y accesible para los fitomejoradores, que les permita aplicar la teoría del análisis de parámetros de estabilidad para seleccionar genotipos que interactúen menos con el ambiente, o a los mejores genotipos para cada ambiente.

3. ESTABILIDAD GENÉTICA

Los conceptos de estabilidad genética pueden ser biológicos o agronómicos (Becker 1981). Biológicamente un genotipo con varianza total mínima bajo diferentes ambientes, es considerado estable. Agronómicamente un genotipo estable tiene una interacción mínima con el ambiente pero rinde bien en buenos ambientes (Eberhart y Russell, 1966). Lin et al (1986) clasificaron a la estabilidad en tres tipos: La estabilidad Tipo 1 sigue los conceptos biológicos y es medida por la varianza mínima a través de un rango de ambientes; la estabilidad Tipo 2 de un genotipo es cuando su respuesta ambiental es paralela a la respuesta media de todos los genotipos de la prueba y la estabilidad Tipo 3 para un genotipo se presenta cuando los cuadrados medios para la desviación de regresión son insignificantes. Lin y Binns (1986) propusieron a la varianza de medias genotípicas a través de ambientes impredecibles (años) promediados a través de ambientes predecibles (localidades) como un nuevo parámetro de estabilidad y designado como Tipo 4.

Márquez (1992) sintetiza el concepto escribiendo que una variedad estable responde a un ambiente pero no interactúa, también señala que los promedios de todas las variedades en cada uno de los ambientes sirven para estimar cuantitativamente a los ambientes, un refinamiento adicional es restar a dicha media la media general. Simmonds (1962), citado por Márquez, define a la adaptabilidad como la capacidad de responder a la selección. Una variedad genéticamente heterogénea es más adaptable.

Carballo y Márquez (1972) proponen el término *consistencia* para una variedad cuya suma de cuadrados de las desviaciones de regresión sea estadísticamente igual a cero. Eberhart y Russell (1966), en su artículo original, señalan la importancia de la interacción GA y la dificultad que éste hecho impone al tratar de demostrar la superioridad significativa de una variedad y que dicha interacción se presenta en cualquier material con el que se pueda trabajar. Citan a Comstock y Moll (1963) que

han demostrado estadísticamente que el efecto de las interacciones frena el proceso de selección. En su mismo artículo mencionan que, para reducir el efecto de la ~~interacción genotipo-ambiente, se ha utilizado la estratificación de ambientes con éxito.~~ La estratificación está usualmente basada en diferencias macroambientales como temperatura, distribución de lluvia y tipos de suelos. De cualquier manera, aún con estos lineamientos, la interacción de genotipos con localidades en una sub-región y con los ambientes presentes de la misma localidad en diferentes años persiste mucho. Allard y Bradshaw (1964) Clasificaron como impredecibles las variaciones ambientales por lo que la estratificación no es, del todo, efectiva.

Si la estratificación de ambientes no funcionó, Eberhart y Russell (1966) proponen la selección de genotipos estables que interactúen menos con los ambientes en los cuales van a crecer. Si la estabilidad del rendimiento o la habilidad mostrada a interactuar con el ambiente en forma mínima es una característica genética, entonces la evaluación preliminar puede ser planeada para identificar genotipos estables.

Jensen (1962) sugirió que una variedad multilínea, comparada con una línea pura, posee mayor estabilidad de producción. También es de esperarse que la interacción genético-ambiental en una población compuesta por varios genotipos sea menor que para un solo genotipo.

Allard y Bradshaw (1964) también sugirieron que una población heterocigota y heterogénea ofrecen la mejor oportunidad de producir variedades que presenten una interacción genotipo-ambiente pequeña.

En Márquez *et. al.* (1981), para el caso del maíz, se observó que las variedades de altos rendimientos tienen coeficientes de regresión mayores que la unidad, esto es, muy altos rendimientos en ambientes más favorables e inferiores, inclusive, a las de las variedades no mejoradas, en los ambientes desfavorables. Con las variedades no mejoradas sucede lo contrario, no responden bien en ambientes favorables pero su rendimiento no cae tan drásticamente en aquellos ambientes desfavorables.

4. PARÁMETROS DE ESTABILIDAD.

El comportamiento de un genotipo en varios ambientes, puede formularse como la regresión del valor fenotípico de interés sobre el efecto ambiental. La diferencia entre un método de otro, hablando de parámetros de estabilidad, puede ser el cómo se toma el efecto del ambiente y cual será la variable dependiente. Uno de los estudios más completos es el de Eberhart y Russell (1966). Ellos realizaron la regresión de los valores fenotípicos Y_{ij} sobre el efecto ambiental A_j al cual se le llamó índice ambiental I_j y lo expresaron en la siguiente ecuación:

$$Y_{ij} = \mu_i + \beta_i I_j + \delta_{ij} \quad i)$$

Donde Y_{ij} es la media de la i -ésima variedad en el j -ésimo ambiente; μ_i es la media de la i -ésima variedad sobre todos los ambientes; β_i es el coeficiente de regresión que mide la respuesta de la i -ésima variedad al cambio de ambientes; δ_{ij} es la desviación de la regresión de la i -ésima variedad al j -ésimo ambiente e I_j es el índice ambiental obtenido de la diferencia de la media del j -ésimo ambiente menos la media general.

$$I_j = \sum_i Y_{ij} / v - \sum_i \sum_j Y_{ij} / av = \bar{Y}_{.j} - \bar{Y}_{..} ; \quad \sum_j I_j = 0 \quad ii)$$

Lo óptimo, sería un índice ambiental independiente de las variedades obtenido de factores ambientales como la precipitación, temperatura y fertilidad del suelo. Eberhart y Russell (1966) señalan que los actuales conocimientos de esos factores y del rendimiento no permiten el cálculo de dichos índices. De cualquier forma, las variedades deben ser cultivadas en un número adecuado de ambientes cubriendo el mayor rango de posibles condiciones ambientales si se quiere que los parámetros de estabilidad proporcionen información útil. Con la ecuación "i", es posible calcular los *Parámetros de Estabilidad* β_i y $\sigma_{d_i}^2$, en donde β_i es una medida de la relación de cambio entre las variables; cuánto cambia el valor fenotípico medido (Y)

al cambiar el efecto ambiental (X) y $\sigma_{d_i}^2$ expresa las desviaciones de la regresión.

Un estimador para β_i se obtiene mediante la fórmula:

$$\hat{\beta}_i = b_i = \frac{\sum_j Y_{ij} I_j}{\sum_j I_j^2} \quad \text{iii)}$$

y un estimador para $\sigma_{d_i}^2$ se obtiene mediante la fórmula:

$$\hat{\sigma}_{d_i}^2 = S_{d_i}^2 = \left[\sum_j \hat{\delta}_{ij}^2 / (n-2) \right] - S_e^2 / r \quad \text{iv)}$$

donde:

$\hat{\delta}_{ij} = Y_{ij} - \hat{Y}_{ij}$ Es la desviación de regresión de la i -ésima variedad en el j -ésimo ambiente

$$\sum_j \hat{\delta}_{ij}^2 = \left[\sum_j Y_{ij}^2 - (Y_{i.})^2 / a \right] - \left[\left(\sum_j Y_{ij} I_j \right)^2 / \sum_j I_j^2 \right]$$

S_e^2 / r es el cuadrado medio del error ponderado.

Los valores esperados \hat{Y}_{ij} se calculan a partir de la ecuación de regresión:

$$\hat{Y}_{ij} = \bar{Y}_{i.} + b_i I_j \quad \text{Donde:}$$

\hat{Y}_{ij} es el valor esperado de Y_{ij} estimado a partir de la ecuación.

$\bar{Y}_{i.} = \mu_i$ = media de la i -ésima variedad en todos los ambientes

Las situaciones que se pueden presentar con los parámetros son:

$$b_i < 1; \quad b_i = 1; \quad b_i > 1 \quad \text{y} \quad S_{d_i}^2 = 0; \quad S_{d_i}^2 > 0$$

De las condiciones anteriores, se pueden formar 6 combinaciones:

- | | | |
|-----------------------------|-----------------------------|-----------------------------|
| 1) $b_i < 1; S_{d_i}^2 = 0$ | 3) $b_i = 1; S_{d_i}^2 = 0$ | 5) $b_i > 1; S_{d_i}^2 = 0$ |
| 2) $b_i < 1; S_{d_i}^2 > 0$ | 4) $b_i = 1; S_{d_i}^2 > 0$ | 6) $b_i > 1; S_{d_i}^2 > 0$ |

Cuando se presenta un valor combinado de $b_i = 1$ y $S_{d_i}^2 = 0$ para una variedad, se dirá que es una variedad estable. Si se presentan dos o más variedades con éstas condiciones, se elige a la de mayor media ambiental.

Anteriormente, por variedad estable se entendía como a aquella que rendía razonablemente igual en un rango amplio de ambientes, lo que sugería que una variedad estable rendía relativamente bien en ambientes adversos y no muy bien en condiciones favorables, es decir, con $b_i < 1$. Eberhart y Russell (1966) establecen que en pruebas de rendimiento de maíz en la Universidad de Iowa han indicado que los híbridos con coeficientes de regresión menor de 1 ($b_i < 1$) usualmente tienen medias de rendimiento por abajo de la media general del experimento. En condiciones de cultivo adversas o para garantizar un mínimo de cosecha, es posible que una variedad con $b_i < 1$ sea aún recomendada; pero en buenas condiciones de cultivo, el mejorador quiere una variedad que rinda por encima del promedio en todos los ambientes, de aquí que se desee una variedad con alto promedio de rendimiento, coeficiente de regresión de 1 ($b_i = 1$) y las desviaciones de regresión tan pequeñas como sea posible ($S_{d_i}^2 = 0$).

La significancia de la diferencia de medias entre variedades ($H_0: \mu_1 = \mu_2 = \mu_3 \dots \dots = \mu_v$) puede ser hecha con una prueba de "F" ($F = CM_1/CM_3$) (ver Cuadro 6) con desviaciones de los cuadrados medios homogéneas dado que CM_3 es la desviación ponderada. Si no hay diferencia entre los coeficientes de regresión ($\beta_1 = \beta_2 = \beta_3 \dots \dots \dots \beta_v$) la prueba de "F" será aproximadamente la misma que $F = CM_1/CM_2$ en el análisis estándar de varianza. Cuando las desviaciones de las variedades no son homogéneas, aparecen dificultades para obtener pruebas de significancia. En estos casos se puede utilizar la prueba aproximada de "t" entre pares de efectos con

diferentes varianzas. [Ver Steel and Torrie, 1960]. De cualquier manera, la prueba de “F” para diferencias entre variedades, CM_1/CM_3 no puede ser menos válida que la usual prueba de “F” [$CM_1/CM(\text{variedad} \times \text{ambiente})$].

La hipótesis de que no hay diferencias genéticas entre variedades para su regresión sobre su índice ambiental, $H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_v$, puede ser también probada aproximadamente por una prueba de “F”

$$F \text{ (aprox.)} = CM_2/CM_3$$

La hipótesis de que cualquier coeficiente de regresión no difiere de la unidad puede ser probada con una prueba de “t” apropiada. Se puede obtener una prueba aproximada para las desviaciones de regresión de cada variedad:

$$F \text{ (aprox.)} = [\sum_j \hat{\delta}_y^2 / (a-2)] / \text{Error ponderado}$$

Se han encontrado experimentos en los cuales la respuesta no es lineal. Molina (1990) señala que entre las respuestas no lineales, el modelo de más fácil interpretación agronómica es la “lineal-cuadrática”:

$$\hat{Y}_i = b_0 + b_{i1}X_i + b_{i2}X_i^2$$

$$\hat{Y}_i = b_0 + b_{i1}X_i - b_{i2}X_i^2$$

También en éstas ecuaciones se analizan los conceptos de b_i y $S_{d_i}^2$, pero los criterios a tomar en cuenta son el coeficiente de determinación r^2 del modelo y el nivel de significancia de las desviaciones de regresión del modelo utilizado. El mejor modelo es el que explique más del 90% de la variación ($r^2 > 0.9$) y las desviaciones de regresión estadísticamente iguales a cero. Molina (1992) establece para el análisis de parámetros de estabilidad de Eberhart y Russell (1966), que las siguientes tres situaciones:

$$1) b_i < 1; S_{d_i}^2 > 0$$

$$2) b_i = 1; S_{d_i}^2 > 0$$

$$3) b_i > 1; S_{d_i}^2 > 0$$

Corresponden a casos en que el ajuste no es lineal sino lineal cuadrático, cuadrático, cúbico o de cualquier otro tipo.

Eberhart y Russell (1966) marcan básicamente dos parámetros de estabilidad: el primero es el coeficiente de regresión b_i que estima la relación de cambio de cada variedad con respecto de un índice ambiental asumiendo que la relación es lineal, en otras palabras cuantifica la respuesta de una variedad a los cambios en el factor ambiente de las localidades en las que se desarrolló. Si se hace una gráfica los puntos observados tomando como variable independiente a los índices ambientales y el rendimiento como variable dependiente y se ajusta una línea recta a dichos puntos, la pendiente de la recta es b_i . El segundo parámetro es $S_{d_i}^2$ que representa a las desviaciones al cuadrado y evalúa la consistencia de cada variedad, es decir, las diferencias entre los rendimientos esperados por la regresión y los rendimientos reales, si $S_{d_i}^2$ tiene un valor estadísticamente igual a cero, se dice que la variedad es consistente dado que los valores esperados son muy similares o estadísticamente iguales a los reales, dicho de otra manera, su comportamiento es predecible o pronosticable. En el caso del segundo parámetro, una variedad que presente una $S_{d_i}^2$ elevada no es pronosticable y no se sabe cómo se vaya a comportar. Puede ser que lo anterior no necesariamente indicará que la variedad es inconsistente sino que la relación entre el ambiente y el rendimiento puede ser "no-lineal" en cuyo caso habrá de buscarse otro tipo de modelo que se ajuste más al estudio que el lineal.

Ante la pregunta: *¿Por qué el valor observado es diferente del valor estimado?* (Y_{ij} diferente de \hat{Y}_{ij}). La respuesta que se considera, es que existen factores no contemplados ni precisados en el modelo, lo que se conoce como error. De ésta manera, el mejor modelo para la estimación de parámetros de estabilidad es aquel que detecte al factor interacción, puesto que si existe interacción, el cálculo de parámetros de estabilidad ni es confiable ni tiene caso, debido a que las interacciones pueden resultar imprevisibles y, precisamente, lo que se está tratando de seleccionar

son variedades que interactúen de manera mínima con el ambiente. Sin embargo lo que sí se podría cuestionar sería la necesaria *linearidad* de la respuesta fenotípica a un índice ambiental, es decir, que exista otro tipo de función (modelo) que se adapte mejor a la descripción del fenómeno observado que es la respuesta de la variedad al medio ambiente. Si se considera lo anterior, habría que establecer los nuevos parámetros de estabilidad y los criterios de interpretación, de los cuales ya se trató.

En un análisis de regresión de variedades en diversos ambientes se presupone que, de no haber interacción GA, las rectas trazadas para cada variedad serían razonablemente paralelas, pero no hay que perder de vista que la respuesta bien podría no ser lineal y que una recta forzada en el ajuste daría un cruce de rectas en la gráfica indicando la existencia de una interacción que bien puede no ser real y conduciendo lo anterior a una inferencia equivocada.

Molina (1990) cita que para un grupo de “v” variedades que se prueban en “a” ambientes, la respuesta en producción o de alguna característica del cultivo se puede establecer como:

$$Y_{ij} = \mu + G_i + A_j + (GA)_{ij} \quad v)$$

Donde:

$\mu = \sum_i \sum_j Y_{ij} / va$: Es la media poblacional o general del experimento (sumatoria general entre el número de observaciones total)

$G_i = \sum_j Y_{ij} / a - \mu$: Efecto genético de la variedad “i” (media de la variedad “i” menos la media general)

$(GA)_{ij} = Y_{ij} - (\mu + G_i + A_j)$: Efecto de la interacción de variedad “i” con ambiente “j”.

$A_j = \sum_i Y_{ij} / v - \mu$: Efecto del ambiente “j”. (media ambiental menos la media general)

Jalaluddin y Harrison (1993), en su artículo sobre repetibilidad de parámetros, establecen que para que una característica o parámetro sea utilizado en la selección de genotipos superiores, éstos deben ser controlados genéticamente y deben ser

repetibles a través de muestras de los ambientes para los cuales las selecciones fueron hechas. Mencionan también que para disminuir las contradicciones estadísticas de los modelos, se hace necesario incluir la mayor cantidad de genotipos en el mayor número de ambientes posibles.

Se han hecho experimentos variados con diferentes cultivos para analizar la repetibilidad de parámetros con el fin de establecer su origen. Lin y Binns (1991) estudiaron las propiedades genéticas de los cuatro tipos de estabilidad, anteriormente mencionados, y concluyeron que los Tipos 1 y 4 sí eran heredables y los Tipos 2 y 3 no lo eran. Lo anterior es importante para los mejoradores de plantas para identificar un estimador de estabilidad que sea repetible y que sea fácil de utilizar para que la selección para estabilidad de una característica sea efectiva para avanzar en el desarrollo de genotipos altamente productivos y estables. (Jalaluddin y Harrison, 1993). Jalaluddin y Harrison (1993) En su estudio para determinar la repetibilidad de estimadores de parámetros de estabilidad derivados de datos de rendimiento de experimentos de trigo de invierno desarrollados en 36 ambientes, probaron los parámetros siguientes: b_i , $|1 - b_i|$, σ_i^2 de Shukla (1972), S_i^2 y Cv_i de Francis y Kannenberg (1978), el coeficiente de determinación r^2 de Pinthus (1973) y la media de producción y encontraron que los estimadores, $S_{d_i}^2$ y $|1 - b_i|$, no fueron repetibles. Lin y Binns (1988) también encontraron que $S_{d_i}^2$ no era genético. b_i fue repetible en 4 casos y r^2 en dos casos de 7 comparaciones. La alta heredabilidad de la respuesta de regresión fue encontrada por varios autores, entre ellos Finaly (1961), Eberhart y Russell (1966), Bains (1976) y Lin y Binns (1991).

En el cálculo de b_i su valor no se altera si al efectuar la regresión de las Y_{ij} se hace sobre las medias ambientales ($\bar{Y}_{.j}$) en lugar de los índices ambientales debido a que la media general es una constante.

5. METODOLOGÍA DEL ANÁLISIS

En el presente trabajo se hará el desarrollo detallado, tal como Eberhart y Russell (1966) lo propusieron, explicando detenidamente los cálculos y las posibles alternativas que se tienen cuando se maneja un programa computacional de fácil acceso como una hoja de cálculo, en éste caso EXCEL, y se propondrá una alternativa que matemáticamente es equivalente pero la manera de presentarlo es más sencilla en su comprensión y aplicación mediante análisis de varianza de las regresiones de las variedades sobre los índices ambientales de manera independiente.

Los conocimientos que se requiere como antecedentes son: la ecuación de la línea recta, y conocimientos básicos de los principales parámetros estadísticos.

Todos los puntos que forman una línea recta trazada sobre un plano cartesiano, satisfacen una ecuación llamada "ecuación de la recta" y es de la forma:

$$Y = b_0 + b_1X$$

X e Y son las coordenadas de un punto cualquiera perteneciente a la recta; b_1 es la pendiente de la recta y corresponde a la tangente del ángulo que forma la recta con el eje de las abscisas ("X"); y b_0 es el punto donde la recta toca al eje de las ordenadas ("Y") (ver Figura 1).

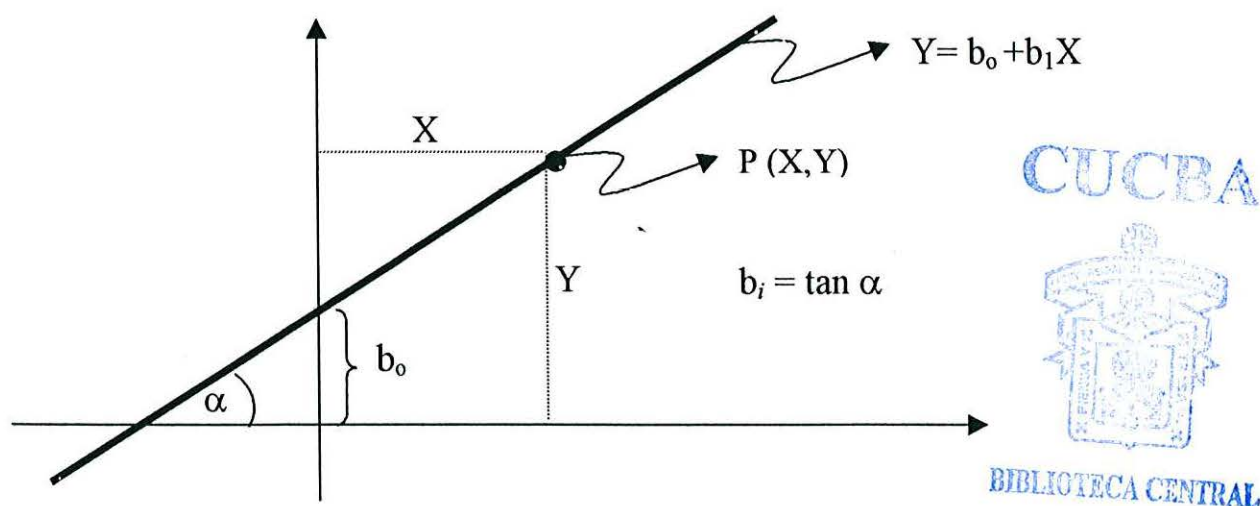


Figura 1. Parámetros de la línea recta

En el Cuadro 1 se presentan los rendimientos de “v” variedades probadas en “a” ambientes. Y_{ij} es el valor promedio del rendimiento de la i -ésima variedad en el j -ésimo ambiente, “i” toma valores desde 1 hasta “v” y “j” toma valores desde 1 hasta a.

La sumatoria “j” de las Y_{ij} ($\sum_j Y_{ij} = Y_{i.}$) es la suma de los rendimientos de la variedad “i” a través de todos los ambientes: $\sum_j Y_{ij} = Y_{11} + Y_{12} + Y_{13} + Y_{14} + \dots + Y_{1a} = Y_{1.}$

La sumatoria “i” de las Y_{ij} ($\sum_i Y_{ij} = Y_{.j}$) es la suma de todas las variedades en el ambiente “j”

$$\sum_i Y_{i1} = Y_{11} + Y_{21} + Y_{31} + Y_{41} + \dots + Y_{v1} = Y_{.1}$$

La sumatoria general de todas las variables se indica con la fórmula $\sum_i \sum_j Y_{ij} = Y_{..}$

$$\sum_i \sum_j Y_{ij} = Y_{11} + Y_{21} + \dots + Y_{12} + Y_{22} + \dots + Y_{av} = Y_{..}$$

La media aritmética de la variedad “i”:

$$\bar{Y}_{i.} = \frac{Y_{i.}}{a}$$

La media aritmética del ambiente “j”:

$$\bar{Y}_{.j} = \frac{Y_{.j}}{v}$$

La media general del experimento:

$$\bar{Y}_{..} = \frac{Y_{..}}{av}$$

La media es un parámetro que indica las tendencias centrales de la población, es decir, la ubicación o localización de la población; mientras que la desviación estándar y la varianza indican qué tan heterogénea es una población, o sea, qué tan alejada está la población, en promedio, de su media. Si se tiene una varianza relativamente alta, ello indica que los elementos pertenecientes a la población se desvían mucho de la media, que son muy diferentes entre ellos, mientras que una varianza de cero, por ejemplo, indicaría que todos los elementos de una población son iguales. En una distribución normal, la media determina la *posición* de la “campana” en un sistema cartesiano, y la varianza la *forma* o amplitud de la misma. En el presente análisis se asume que los datos presentan una distribución normal.

En el Cuadro 2, se presentan los símbolos que se utilizan en el presente escrito.

Cuadro 1. *Notación estadística utilizada en el presente escrito.*

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5VAR v	$\Sigma_i Y_{ij} = Y_{.j}$
AMB 1	Y_{11}	Y_{21}	Y_{31}	Y_{41}	Y_{51} Y_{v1}	$Y_{.1}$
AMB 2	Y_{12}	Y_{22}	Y_{32}	Y_{42}	Y_{52} Y_{v2}	$Y_{.2}$
AMB 3	Y_{13}	Y_{23}	Y_{33}	Y_{43}	Y_{53} Y_{v3}	$Y_{.3}$
AMB 4	Y_{14}	Y_{24}	Y_{34}	Y_{44}	Y_{54} Y_{v4}	$Y_{.4}$
:	:	:	:	:	:	:	:
AMB a	Y_{1a}	Y_{2a}	Y_{3a}	Y_{4a}	Y_{5a} Y_{va}	$Y_{.a}$
$\Sigma_j Y_{ij} = Y_{i.}$	$Y_{1.}$	$Y_{2.}$	$Y_{3.}$	$Y_{4.}$	$Y_{5.}$ $Y_{v.}$	$Y_{..}$

$$\sum_j Y_{ij} = Y_{i.}$$

$$\sum_j Y_{.j} = Y_{11} + Y_{12} + Y_{13} + Y_{14} + \dots + Y_{1a} = Y_{1.}$$

$$\sum_i Y_{ij} = Y_{.j}$$

$$\sum_i Y_{i1} = Y_{11} + Y_{21} + Y_{31} + Y_{41} + \dots + Y_{v1} = Y_{.1}$$

$$\sum_i \sum_j Y_{ij} = Y_{..}$$

$$\sum_i \sum_j Y_{ij} = Y_{11} + Y_{21} + \dots + Y_{12} + Y_{22} + \dots + Y_{av} = Y_{..}$$

$$\bar{Y}_{i.} = \frac{Y_{i.}}{a}$$

$$\bar{Y}_{.j} = \frac{Y_{.j}}{v}$$

$$\bar{Y}_{..} = \frac{Y_{..}}{av}$$

$$S^2 = \frac{\Sigma(Y_{ij} - \bar{Y}_{ij})^2}{n-1}$$

Cuadro 2. Simbología utilizada.

Σ	Sumatoria.
$>$	Mayor que.
$>>$	Mucho mayor que.
$<$	Menor que.
$<<$	Mucho menor que.
\approx	Aproximado
Y_{ij}	Variable dependiente (valor observado de la variedad “i” en el ambiente “j”).
$\bar{Y}_{.j}$	Media aritmética de todas las variedades en el ambiente “j”.
\bar{Y}_i	Media aritmética de la variedad “i” a través de todos los ambientes.
$\bar{Y}_{..}$	Media aritmética de todo el experimento.
\wedge	
Y_{ij}	Valor esperado para la variedad “i” en el ambiente “j”.
μ	Media poblacional.
β_i	Coefficiente de regresión para la i -ésima variedad. (Parámetro de estabilidad)
b_i	Valor estimado de β_i .
$S^2_{d_i}$	Desviaciones de regresión (Parámetro de estabilidad)
I_j	Índice ambiental par el j -ésimo ambiente.
R^2_i	coeficiente de determinación para la i -ésima variedad.
R_i	coeficiente de correlación para la i -ésima variedad.

La fórmula general, por definición, para calcular la varianza en un experimento es la

siguiente:

$$S^2 = \frac{\sum(Y_{ij} - \bar{Y}_{ij})^2}{n-1}$$

Realizando un simple desarrollo matemático que consiste en desarrollar el binomio al cuadrado y simplificar, se obtiene la siguiente fórmula:

$$S^2 = \frac{\sum Y_{ij}^2 - \frac{(\sum Y_{ij})^2}{n}}{n-1}$$

Que para fines de cómputo es más útil y matemáticamente se equivalen.

En ambas fórmulas anteriores, la parte superior es la *suma de cuadrados*, y la parte inferior los *grados de libertad*, y se van modificando dependiendo del caso y la utilización que se les vaya a dar.

Considerando que el objetivo del trabajo es presentar al método de una manera más comprensible, se juzgó conveniente empezar con un ejemplo pequeño para ilustrar los conceptos básicos más comúnmente utilizados. El fin es familiarizarse con los estadísticos y sus cálculos, los que se explicarán detalladamente.

Suponiendo que se probaron 5 variedades de maíz en 6 ambientes diferentes que se acomodaron de menor a mayor media ambiental y se quiere revisar cómo están relacionados el ambiente y las variedades. Los datos se presentan en el Cuadro 3.

En las hojas de cálculo de programas computacionales, normalmente se nombran a las columnas con letras y a las filas con números, por lo que cada celda tiene una letra y un número indicando su posición. En el Cuadro 4 se encuentran los mismos datos del Cuadro 3, acomodados en una hoja de cálculo de EXCEL, las variedades en columnas y los ambientes en filas. Cada celda puede contener: un dato numérico con el que se pueden hacer cálculos matemáticos, una función matemática que trabaje con los datos de otras celdas, o bien, letras de referencia con las que *no* se puede hacer cálculos, ya que las operaciones con éstas celdas conducen a un error.

Cuadro 3. Cinco variedades probadas en 6 ambientes. Rendimiento en kilogramos por parcela.

	V1	V2	V3	V4	V5	$\bar{Y}_{.j}$	$I_j = \bar{Y}_{.j} - \bar{Y}_{..}$	
AMB1	20	20	22	23	24	21.8	-1.52	=21.8-23.323
AMB2	18.8	19	24	24	25	22.16	-1.16	=22.16-23.32
AMB3	21.8	22	25	23	24	23.16	-0.16	=23.16-23.32
AMB4	20	24	23	25	26	23.6	0.28	=23.6-23.323
AMB5	21.3	23.5	24	26	26	24.16	0.84	=24.16-23.32
AMB6	19.3	25	26	27	28	25.06	1.74	=25.06-23.32
\bar{Y}_i	20.2	22.25	24	24.667	25.5	23.3233		
								$\sum_j I_j^2 = 7.49313$
$\sum_j y_{ij}^2 =$	6.620	27.875	10.000	13.333	11.500	69.328	SCA=	37.465667
$\sum_j Y_{ij} I_j =$	0.974	13.555	6.080	8.847	8.010		SCV=	107.44533
$b_i =$	0.130	1.809	0.811	1.181	1.069		SCT=	176.77367
$r^2 =$	0.019	0.880	0.493	0.783	0.745			
SCDR =	6.493	3.354	5.067	2.889	2.937	20.740	SCGA=	11.122667
SCR =	0.12661	24.52085	4.933370	10.4447	8.56252	48.588		
	Y_{1j}	\hat{Y}_{1j}	$\hat{Y}_{1j} - \bar{Y}_1$	$\hat{Y}_{1j} - Y_{1j}$				
	20	20.002	-0.198	0.002				
	18.8	20.049	-0.151	1.249				
	21.8	20.179	-0.021	-1.621				
	20	20.236	0.036	0.236				
	21.3	20.309	0.109	-0.991				
	19.3	20.426	0.226	1.126				
$\Sigma(\text{Desv.})^2$	6.62	0.12661	0.12661	6.4934				

$\sum_j (Y_{1j} - \bar{Y}_1)^2 = 6.62$ representa la variación total de la variedad 1 a través de ambientes.

$\sum_j (\hat{Y}_{1j} - \bar{Y}_1)^2 = 0.12661$ es la porción de la variación total de la variedad 1 explicada por el modelo de regresión.

$\sum_j (\hat{Y}_{1j} - Y_{1j})^2 = 6.4934$ es la porción de la variación total de la variedad 1 que no es explicada por el modelo de regresión


$${}^2 \sum_j y_{ij}^2 = \sum_j (Y_{ij} - \bar{Y}_i)^2$$

Cuadro 4. Hoja de cálculo del programa "EXCEL".

Microsoft Excel - TESIS MARCIANO ARGOTE								
Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ?								
PRONOSTICO X ✓ = =PRONOSTICO(H2;B2:B7;H2:H7)								
	A	B	C	D	E	F	G	H
1		V1	V2	V3	V4	V5	\bar{Y}_j	I_j
2	AMB1	20	20	22	23	24	21.8	-1.52
3	AMB2	18.8	19	24	24	25	22.16	-1.16
4	AMB3	21.8	22	25	23	24	23.16	-0.16
5	AMB4	20	24	23	25	26	23.6	0.28
6	AMB5	21.3	23.5	24	26	26	24.16	0.84
7	AMB6	19.3	25	26	27	28	25.06	1.74
8	\bar{Y}_i	20.2	22.25	24	24.66667	25.5	23.32333	
9							$\Sigma I_j^2 = 7.493133$	
11	$\Sigma_j y_{ij}^2 =$	6.620	27.875	10.000	13.333	11.500	69.328	
12	$\Sigma_j Y_{ij} * I_j =$	0.974	13.555	6.080	8.847	8.010		
13	$b_i =$	0.130	1.809	0.811	1.181	1.069		
14	$r^2 =$	0.019	0.880	0.493	0.783	0.745		
15	SCDR =	6.493	3.354	5.067	2.889	2.937	20.740	
16	SCR =	0.126606	24.520854	4.9333701	10.4447	8.562519	48.588	
18								
19								
20		Y_{ij}	Y_{ij}	$Y_{ij} - \bar{Y}_i$	$Y_{ij} - \bar{Y}_j$		SCA =	37.46567
21		20	20.049	-0.198	0.002		SCV =	107.4453
22		18.8	20.049	-0.151	1.249		SCT =	176.7737
23		21.8	20.179	-0.021	-1.621			
24		20	20.236	0.036	0.236		SCGA =	11.12267
25		21.3	20.309	0.109	-0.991			
26		19.3	20.426	0.226	1.126			
27	$\Sigma(\text{Desv})^2$	6.62	0.126606	0.126606	6.493394			

Medias Ambientales ($\bar{Y}_{.j}$). Después de la última columna de datos del Cuadro 3 está otra columna, que son las medias ambientales y se calcula sumando los datos de todas las variedades en un ambiente y dividirlo entre el número de variedades (*ver Cuadro 1*), lo que se obtiene en la hoja de cálculo de EXCEL mediante la función “PROMEDIO”. Por ejemplo, para el caso del Cuadro 3, los datos de todas las variedades en el ambiente 1 están desde la celda B2 hasta la celda F2 (*ver Cuadro 4*), por lo que en la celda siguiente (G2), o en la que se prefiera, se escribe:

$$=PROMEDIO(B2:F2)$$

Siempre que se termine de insertar una función, se deberá aceptarla con “ENTER” o haciendo “clic” con el botón izq. del “ratón” en el icono “introducir”  (*ver Cuadro 4*). Para el ambiente 2, se escribe en la siguiente celda elegida se escribe:

$$=PROMEDIO(B3:F3)$$

Y así sucesivamente. Con un poco de experiencia, se coloca la flecha del “ratón” en la esquina inferior derecha de la celda que contenga la media 1 y se “arrastra” (haciendo un “clic” manteniendo oprimido el botón izquierdo) hacia abajo (o hacia la derecha, según sea el caso) tantas celdas como medias haya que calcular.

Medias Varietales (\bar{Y}_i). El mecanismo es igual, se obtiene sumando todos los rendimientos de cada variedad a través de ambientes y se divide entre el número de datos. En la hoja de cálculo EXCEL dado que los datos de la variedad 1 están de la celda B2 hasta B7 (*ver Cuadro 4*), se elige la celda inmediata inferior (B8 o donde se desee que aparezca el resultado) y se escribe:

$$=PROMEDIO(B2:B7)$$

De igual manera, se pone la flecha del “ratón” en la esquina inferior izquierda y se “arrastra” hacia la derecha tantas celdas como medias haya que calcular. Para la variedad 2, la función es igual, solo que el rango es (C2:C7) debido a que es donde se encuentran los datos para la ésta variedad. Así sucesivamente para todas las variedades.

Media General ($\bar{Y}_{..}$). Se obtiene calculando el promedio de todos los datos del experimento con procedimiento idéntico, o bien con la misma función “PROMEDIO” pero eligiendo todas las celdas de la manera siguiente:

$$=PROMEDIO(B2:F7)$$

Índice Ambiental (I_j). Se calcula simplemente extrayendo la media general de cada una de las medias ambientales:

$$I_j = \bar{Y}_{.j} - \bar{Y}_{..}$$

(En el Cuadro 3 están claramente señaladas las operaciones).

Para calcular el índice del ambiente 1, en la celda “H2” se escribe:

$$=G2-G8$$

Lo

anterior es dado que en G2 está la media del ambiente 1 y en G8 está la media general. En “H3” se escribe:

$$=G3-G8$$

Y así sucesivamente hasta calcular todos los índices ambientales. En éste caso *no se puede* hacer clic en la parte inferior derecha y *arrastrar* para calcular los siguientes índices por que G8 se mantiene constante, y si se hiciera, en H3, por ejemplo, aparecería: “=G3-G9” en lugar de “=G3-G8” y se produciría un error.

Desviaciones al cuadrado. La distancia que hay entre una observación y su media se le denomina *desviación* y normalmente se le designa con una letra minúscula. Si se obtienen las desviaciones de los datos de una variedad con su respectiva media, se elevan al cuadrado y se suman, se obtiene la “*Suma de las Desviaciones al Cuadrado*” y matemáticamente se expresan así:

$$\sum_j y_{ij}^2 = \sum_j (Y_{ij} - \bar{Y}_{i.})^2$$

Por ejemplo, para cada una de las variedades del Cuadro 3 se tienen las siguientes sumas de las desviaciones al cuadrado:

$$\sum_j y_{1j}^2 = \sum_j (Y_{1j} - \bar{Y}_{1.})^2 = (20-20.2)^2 + (18.8-20.2)^2 + (21.8-20.2)^2 + (20-20.2)^2 + (21.3-20.2)^2 + (19.3-20.2)^2 = 6.62$$

$$\sum_j y_{2j}^2 = \sum_j (Y_{2j} - \bar{Y}_{2.})^2 = (20-22.25)^2 + (19-22.25)^2 + (22-22.25)^2 + (24-22.25)^2 + (23.5-22.25)^2 + (25-22.25)^2 = 27.875$$

$$\sum_j y_{3j}^2 = \sum_j (Y_{3j} - \bar{Y}_{3.})^2 = 10.0$$

$$\sum_j y_{4j}^2 = \sum_j (Y_{4j} - \bar{Y}_{4.})^2 = 13.333$$

$$\sum_j y_{5j}^2 = \sum_j (Y_{5j} - \bar{Y}_{5.})^2 = 11.5$$

Los anteriores resultados se obtienen en el programa "EXCEL" con la función "DESVIA2", señalando las celdas con las que se va a trabajar de la misma manera que en los casos anteriores, por ejemplo, para la variedad 1 (que se encuentra entre el rango de celdas B2:B7), se escribe en la celda B11 en la que aparecerá el resultado:

$$=DESVIA2(B2:B7)$$

Para la variedad 2 se escribe en C11:

$$=DESVIA2(C2:C7)$$

O bien se "arrastra" la celda, como ya se explicó, para los siguientes cálculos.

La siguiente fila, $\sum_j Y_{ij} I_j$, cada dato de una variedad a través de ambientes se multiplica por su respectivo índice ambiental y se suman. Se calcula de la siguiente manera:

$$\sum_j Y_{1j} I_j = (20)(-1.52) + (18.8)(-1.16) + (21.8)(-0.16) + (20)(-0.28) + (21.3)(-0.84) + (19.3)(1.74) = 0.974$$

$$\sum_j Y_{2j} I_j = (20)(-1.52) + (19)(-1.16) + (22)(-0.16) + (24)(-0.28) + (23.5)(-0.84) + (25)(1.74) = 13.555$$

$$\sum_j Y_{3j} I_j = (22)(-1.52) + (24)(-1.16) + (25)(-0.16) + (23)(-0.28) + (24)(-0.84) + (26)(1.74) = 6.08$$

$$\sum_j Y_{4j} I_j = (23)(-1.52) + (24)(-1.16) + (23)(-0.16) + (25)(-0.28) + (26)(-0.84) + (27)(1.74) = 8.847$$

$$\sum_j Y_{5j} I_j = (24)(-1.52) + (25)(-1.16) + (24)(-0.16) + (26)(-0.28) + (26)(-0.84) + (28)(1.74) = 8.01$$

En EXCEL los resultados se obtienen mediante la función “SUMAPRODUCTO” y se seleccionan las celdas deseadas. La variedad 1 esta desde la celda B2 hasta la celda B7, y los índices ambientales están de la celda H2 hasta la celda H7. La función se escribe en la celda en que aparecerá el resultado y queda de la siguiente manera:

$$=SUMAPRODUCTO(B2:B7;H2:H7)$$

Para la variedad 2, en la celda elegida se escribe:

$$=SUMAPRODUCTO(C2:C7;H2:H7)$$

Y así sucesivamente. (En este cálculo no funciona el “arrastrar” la celda).

Coefficiente de Regresión. Hacia abajo, el siguiente renglón corresponde al coeficiente de regresión b_i que se calcula con la fórmula:

$$b_i = \frac{\sum_j Y_{ij} I_j}{\sum_j I_j^2}$$

$\sum_j Y_{ij} I_j$ ya se calculó previamente, y $\sum_j I_j^2$ se calcula sumando cada índice ambiental al cuadrado. (En los Cuadros 2 y 3 están calculados)

$$b_1 = 0.974 / 7.4934 = 0.130$$

$$b_4 = 8.847 / 7.4934 = 1.181$$

$$b_2 = 13.555 / 7.4934 = 1.809$$

$$b_5 = 8.01 / 7.4934 = 1.069$$

$$b_3 = 6.08 / 7.4934 = 0.811$$

En el programa EXCEL, el coeficiente de regresión, b , se obtiene con la función “ESTIMACION.LINEAL” señalando las celdas que contienen los datos. Como ejemplo, para la regresión de la variedad 1 (de B2 a B7) sobre los índices ambientales (de H2 a H7) se escribe en la celda en donde se desee que aparezca el resultado lo siguiente:

$$=ESTIMACION.LINEAL(B2:B7;H2:H7)$$

Para el cálculo de b para la variedad 2, se escribe en la celda elegida:

=ESTIMACION.LINEAL(C2:C7;H2:H7)

Coefficiente de Determinación. R^2 es el coeficiente de determinación, y se calcula, para el citado ejemplo, mediante la fórmula:

$$R_i^2 = \left[\sum_j Y_{ij} I_j \right]^2 / \left[\sum_j I_j^2 \sum_j y_{ij}^2 \right]$$

En la que ambos términos ya fueron calculados:

$$R_1^2 = (0.974)^2 / [(7.4931)(6.620)] = 0.019 \quad R_4^2 = (8.847)^2 / [(7.4931)(13.333)] = 0.783$$

$$R_2^2 = (13.555)^2 / [(7.4931)(27.857)] = 0.880 \quad R_5^2 = (8.010)^2 / [(7.4931)(11.500)] = 0.745$$

$$R_3^2 = (6.080)^2 / [(7.4931)(10.000)] = 0.493$$

El coeficiente de determinación, R^2 , se calcula en EXCEL mediante la función "COEFICIENTE.R2" seleccionando las celdas que contengan los datos. Por ejemplo para calcular el coeficiente de determinación, R^2 , para la regresión de la variedad 1 sobre los índices ambientales, se escribe en la celda en la que aparecerá el resultado lo siguiente:

=COEFICIENTE.R2(B2:B7;H2:H7)

Para la variedad 2 se escribe en la celda elegida:

=COEFICIENTE.R2(C2:C7;H2:H7)

Y así sucesivamente se calcula para todas las variedades.

Con los valores observados de cada variedad a través de ambientes, Y_{ij} (variable dependiente) y el índice ambiental de cada ambiente, I_j (variable independiente) se forman "pares" de datos son las coordenadas (x, y) de puntos y se localizan en un

I_j	Y_{ij}	
-1.52333	20	sistema de ejes coordinados para formar un "Diagrama de
-1.16333	18.8	<i>Dispersión</i> . Por ejemplo, para la variedad 1 se tienen los
-0.16333	21.8	pares de datos (coordenadas de los puntos a trazar) en la
0.27667	20	forma como aparece a la izquierda. Al colocar los puntos a
0.83667	21.3	los que corresponden las coordenadas (6 puntos en el
1.73667	19.3	

ejemplo), se forma el mencionado diagrama de dispersión, el cual se puede observar en la Figura 5-A, y en la misma figura, se puede observar una línea recta que se ajusta a los puntos de dispersión llamada línea de regresión, para la cual se hicieron ya los cálculos correspondientes y aparecen en el Cuadro 3. Se tiene el coeficiente de regresión, b , y la intersección de la línea con el eje “Y” es la media de cada variedad (debido a que la media de los índices ambientales es cero), por lo que la ecuación se presenta de la siguiente manera:

$$\hat{Y}_{ij} = b_i I_j + \bar{Y}_i \quad \text{donde:}$$

\hat{Y}_{ij} Es el valor esperado de Y_{ij} (los valores esperados se ubican sobre la línea de regresión)

\bar{Y}_i es la media de la variedad “ i ”

I_j es el índice ambiental del ambiente “ j ”

b_i es el coeficiente de regresión de la variedad “ i ”

Por ejemplo, para la variedad 1 la ecuación es: $\hat{Y}_{1j} = 0.13I_j + 20.2$

Y los valores esperados se pueden obtener en el programa EXCEL mediante la función “PRONÓSTICO”. Para calcular el valor estimado de “Y” para la variedad 1, en el ambiente 1 (el valor observado fue de 20), en la celda deseada se escribe:

$$=PRONOSTICO(H2;B2:B7;H2:H7)$$

En la celda H2 está el valor del índice del ambiente 1 para el cuál se está calculando el valor esperado. De H2 a H7 están los índices ambientales (variables independientes) y de B2 a B7 están los rendimientos observados de la variedad 1 (variable dependiente). De la misma variedad 1, en el ambiente 2, la fórmula insertada en la celda deseada será:

$$=PRONOSTICO(H3;B2:B7;H2:H7)$$

CUCRA



BIBLIOTECA CENTRAL

Se puede también obtener los diagramas de dispersión en el mismo programa EXCEL: en el menú “insertar” se escoge “gráfico” eligiendo enseguida “XY (DISPERSIÓN)” como gráfico deseado y se siguen las instrucciones que aparezcan. La figura 5 muestra los diagramas de dispersión, líneas de regresión, ecuación de regresión y coeficiente de determinación para cada una de las variedades.

5.1 *Análisis de Varianza*

El Análisis de varianza se hace desde dos puntos de vista: para el experimento en general y para la regresión de cada variedad sobre los índices ambientales

Análisis de varianza para el experimento en general:

5.1.1 *Suma de Cuadrados Total del experimento.*

Es la suma de las desviaciones al cuadrado de todas las observaciones con respecto a la media general.

$$SCT = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2$$

En la hoja de cálculo de EXCEL (*ver Cuadro 4*), los datos del experimento están en el rango de celdas B2:F7, por lo que en la celda en la que aparecerá el resultado se escribe:

$$=DESVIA2(B2:F7)$$

5.1.2 *Suma de Cuadrados de Variedades.*

Es la suma de las desviaciones al cuadrado de las medias varietales con respecto de la media general multiplicado por el número de ambientes:

$$SCV = a \sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 = 107.445333$$

Se puede obtener fácilmente en la hoja de cálculo de la siguiente manera. Se eligen las celdas en las que se encuentran las medias ambientales (B8:F8) con la función “DESVIA2” y se multiplica por el número de ambientes, es decir, en la celda en que se quiera el resultado se escribe:

$$=DESVIA2(B8:F8)*6$$

5.1.3 Suma de Cuadrados para Ambientes.

Es la suma de las desviaciones al cuadrado de las medias ambientales con respecto de la media general multiplicado por el número de variedades:

$$SCA = v \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 = 37.4656667$$

Se puede obtener fácilmente en la hoja de cálculo de la siguiente manera. Se eligen las celdas en las que se encuentran las medias ambientales (G2:G7), con la función “DESVIA2” y se multiplica por el número de ambientes, es decir, en la celda en que se quiera el resultado se escribe:

$$=DESVIA2(G2:G7)*5$$

5.1.4 Suma de Cuadrados para la interacción Variedad por Ambiente.

Se obtiene mediante la siguiente extracción:

$$SCVxA = SCT - SCV - SCA - SCDR$$

En el **Análisis de Varianza** de la regresión se lleva a efecto con cada una de las variedades contra los índices ambientales. El fundamento es el siguiente:

$$\sum_j (Y_{ij} - \bar{Y}_{i.})^2 = \sum_j (Y_{ij} - \bar{Y}_{i.})^2 + \sum_j (Y_{ij} - Y_{ij})^2$$

Suma de cuadrados totales.	Suma de cuadrados de la regresión.	Suma de cuadrados de las desviaciones de regresión.
----------------------------------	--	---

La anterior ecuación se entiende mejor como se muestra gráficamente en las figuras 2, 3 y 4. Por desarrollos matemáticos sencillos se obtienen fórmulas más fáciles de trabajar:

$$\sum_j (Y_{ij} - \bar{Y}_{i.})^2 = \sum_j y_{ij}^2$$

$$\sum_j (Y_{ij} - \bar{Y}_{i.})^2 = r_i^2 \sum_j y_{ij}^2 = b_i^2 \sum_j x_j^2 = \frac{(\sum_j xy)^2}{\sum_j x^2}$$

$$\sum_j (Y_{ij} - Y_{ij})^2 = (1 - r_i^2) \sum_j y_{ij}^2$$

Nótese que se está tratando con literales minúsculas, o sea desviaciones. (Ver Cuadro 3)

Para el caso de parámetros de estabilidad, $\sum_j x_j^2 = \sum_j I_j^2$

debido a que la media de los índices ambientales es igual a cero, por definición.

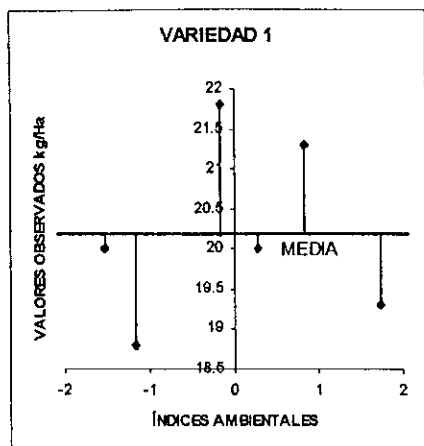


Figura 2. *Variación total*

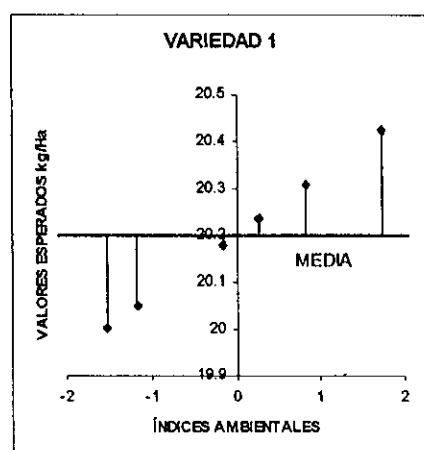


Figura 3. *Variación explicada*

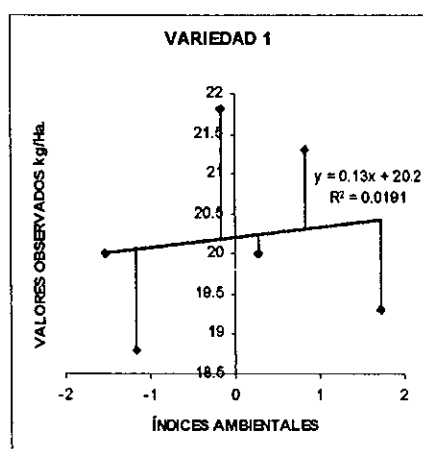


Figura 4. *Variación no explicada*

$\sum_j (Y_{ij} - \bar{Y}_{i.})^2$; Es la suma de los cuadrados de las distancias verticales que hay entre los valores observados (puntos del diagrama de dispersión) y la media de la misma variedad. En éste caso es *la variación total* de la variedad 1, es decir, la *Suma de Cuadrados Total* para la regresión de la variedad 1 sobre los índices ambientales. Su interpretación gráfica puede observarse en la figura 2.

$\sum_j (Y_{ij} - \bar{Y}_{ij})^2$; Es la suma de los cuadrados de las distancias verticales que hay entre los valores esperados para la variedad 1 (localizados sobre la línea de regresión) y la media de la misma variedad. Es *la variación explicada* por el modelo lineal utilizado y es la *Suma de Cuadrados para la Regresión*. Representa, pues, *la porción de la variación total explicada*. Su interpretación gráfica puede observarse en la figura 3.

$\sum_j (Y_{ij} - Y_{ij})^2$; Es la suma de los cuadrados de las distancias verticales que hay entre los valores observados (puntos del diagrama de dispersión) y los valores esperados para la variedad 1 (localizados sobre la línea de regresión). Es *la variación no explicada* por el modelo lineal y es la *Suma de Cuadrados de las Desviaciones de Regresión*. Representa la porción de la variación total no explicada. Su interpretación gráfica puede observarse en la figura 4.

En el Cuadro 3 aparecen los cálculos para las sumas de cuadrados de la regresión y de las desviaciones de regresión.

5.1.5 La suma de cuadrados de las desviaciones ponderadas SCDP.

Es la suma de las desviaciones de regresión de las variedades sobre los índices ambientales.

5.1.6 La suma de cuadrados del error ponderado.

Se obtiene de sumar las sumas de cuadrados del error experimental de los análisis de varianza efectuados para cada localidad.

Grados de libertad para el total: $(av-1) = (5)(6) - 1 = 29$

Grados de libertad para variedades: $(v-1) = (5-1) = 4$

Grados de libertad para VxA (lineal): $(v-1) = (5-1) = 4$

Grados de libertad para las desviaciones ponderadas: $v(a-2) = 5(4) = 20$

Grados de libertad para las desviaciones de regresión: $(a-2) = 4$

Grados de libertad para el error ponderado: $a(r-1)(v-1) = 72$

Los cálculos se disponen en el Cuadro 5 para el análisis de varianza



Cuadro 5. Análisis de varianza para el ejemplo.

FUENTE	G.L.	S.C.	C. MED.	Fc.	P(Fc)
TOTAL	29	176.77			
Variedades (V)	4	107.44	26.86	25.901	1.15E-07
Residual	25				
Amb. (lineal)	1	37.466			
VXA (lineal)	4	11.123	2.78075	2.681	0.06132
Desv Pond	20	20.740287	1.0370		
VAR 1	4	6.493394	1.6233	3.95537	0.0059
VAR 2	4	3.354146	0.8385	2.043135	0.0974
VAR 3	4	5.066630	1.2667	3.086272	0.0210
VAR 4	4	2.888636	0.7222	1.759575	0.1464
VAR 5	4	2.937481	0.7344	1.789328	0.1403
E. POND.	72	29.55	0.4104		

UNIVERSIDAD CENTRAL

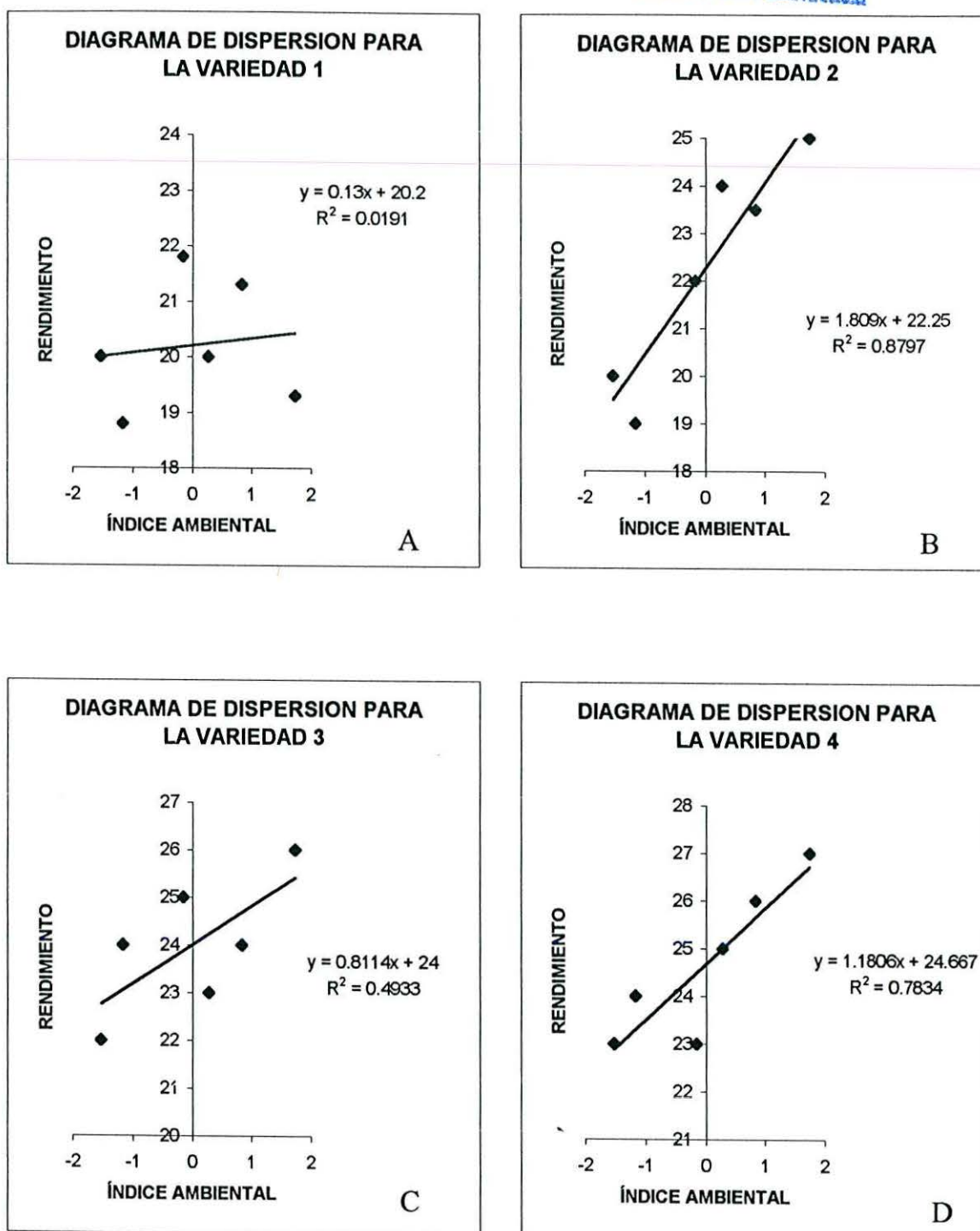


Figura 5. Diagramas de dispersión y líneas de regresión para el ejemplo.

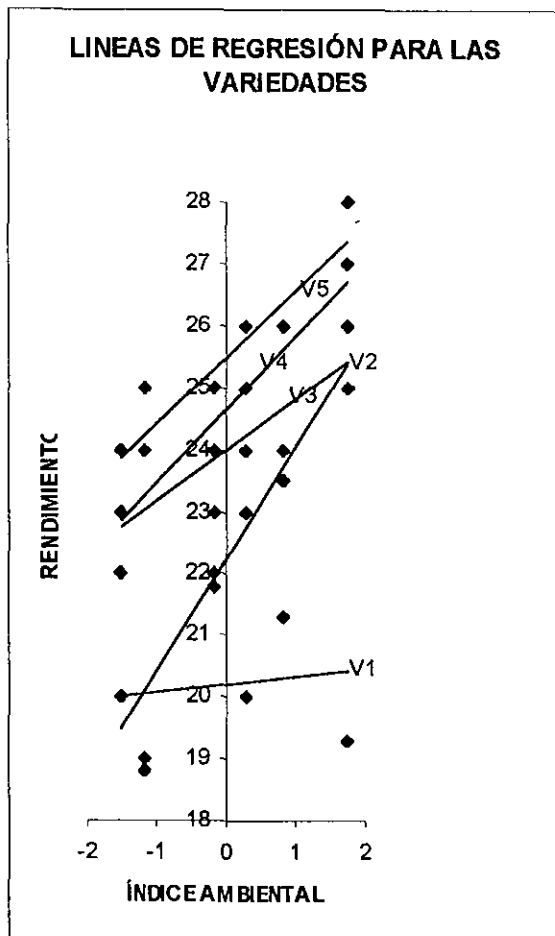
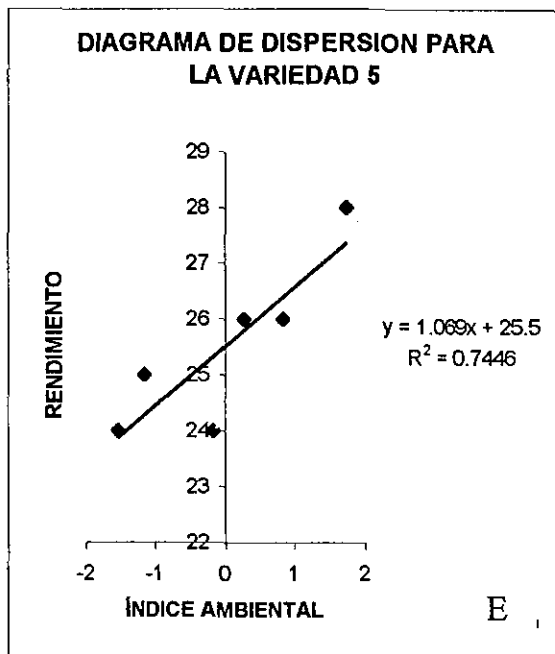


Figura 5. (Continuación.)

En el Cuadro 5, la diferencia entre variedades es altamente significativa, puesto que si la hipótesis nula fuese cierta, la probabilidad de que ocurra un valor de F mayor o igual que 25.9 es de 1.15×10^{-7} , por lo que se rechaza la hipótesis y se acepta que son diferentes las variedades. Si no hubiese interacción genotipo-ambiente, la probabilidad de que ocurra un valor de “F” mayor o igual que 2.68 es de .06, se acepta la hipótesis nula, no hay interacción. Para las variedades 1 y 3, las desviaciones de regresión son significativas.

Visualmente se puede hacer un primer análisis al percibir qué tan alejados están los puntos marcados en la dispersión de la línea de regresión y cómo se comporta la variedad a través de ambientes. El coeficiente de determinación, R^2 , indica la porción de la variación total que es explicada por el modelo. Para un mismo valor de “I_j” hay dos valores de “Y”: el observado y el esperado, por lo que la distancia es vertical (paralela al eje Y). Esta distancia, gráficamente hablando, es una línea vertical que une a los puntos de dispersión con la recta.

^

En el Cuadro 3 están calculadas los valores esperados (\hat{Y}_{ij}) para los rendimientos de la variedad 1 mediante su ecuación y la distancia entre ellos y la recta, es decir las desviaciones $d = (\hat{Y}_{ij} - Y_{ij})$. Las desviaciones se elevan al cuadrado y se suman, obteniendo la suma de los cuadrados de las desviaciones, que como se puede observar es igual a la *Suma de Cuadrados de las Desviaciones de Regresión* (SCDR), a veces llamado residual o error, y es la parte de la variación total **no explicada**.

$$SCDR_i = \sum_j (\hat{Y}_{ij} - Y_{ij})^2$$

El modelo explica y predice, mediante la ecuación de regresión, los valores esperados, pero la distancia entre los esperados y los observados quedan fuera de la explicación del modelo. R^2 establece el porcentaje de la variación total del experimento que es explicada por el modelo matemático utilizado (modelo lineal), de manera que $(1-R^2)$ sería la porción de la variación total no explicada. **No se**

olvide que solo se está considerando la variación independientemente para cada variedad. La variación total (de la variedad i) es $\sum_j y_{ij}^2 = \sum_j (Y_{ij} - \bar{Y}_{i.})^2$ por lo que la *Suma de Cuadrados de la Regresión* (porción de la variación total explicada) se puede calcular con la fórmula:

$$SCR = r_i^2 \sum_j y_{ij}^2$$

Y la *Suma de Cuadrados de las Desviaciones de Regresión* (porción de la variación total no explicada) mediante la fórmula:

$$SCDR = (1 - r_i^2) \sum_j y_{ij}^2$$

Con los valores esperados o bien con la ecuación obtenida, se traza una “línea de tendencia”. EXCEL proporciona también la línea de tendencia. Colocando la flecha del “ratón “ en un punto del diagrama de dispersión, se hace clic en el botón derecho y se elige “Agregar línea de tendencia”. Se pueden escoger varias opciones según las probables tendencias. En ése mismo cuadro de diálogo “Agregar línea de tendencia”, haciendo “clic” en la pestaña “opciones”, se puede agregar la ecuación de la línea de tendencia y el valor del coeficiente de determinación.

En el diagrama de dispersión para la variedad 1 (Figura5A) se establecen los puntos mencionados, la línea de tendencia, la ecuación de regresión y el coeficiente de determinación R^2 . Claramente se observa que los puntos de los valores observados están muy alejados de la línea de tendencia, excepto el punto correspondiente al ambiente 1, lo que genera unas desviaciones grandes y mucha variabilidad “no explicada” o no contemplada indica que el modelo lineal no está representando a la realidad. $R^2 = 0.019$ indica que sólo el 1.9 % de la variabilidad total de la variedad 1 es explicada por el modelo lineal, el resto de variabilidad queda sin explicación. La línea de tendencia con $b_1 = 0.13$ indica que la variedad rinde razonablemente igual en todos los ambientes, aunque ésta información ya no es confiable dadas las desviaciones tan grandes que presentó.

Para la variedad 2 (Figura 5B) se observa que los puntos están mas cerca de la línea de tendencia, el coeficiente de determinación $R^2 = 0.8797$ indica que casi el 88 % de la variación total de la variedad está explicada por el modelo matemático que se

utilizó, $\hat{Y}_i = \bar{Y}_i + b_1 I_j$. La pendiente $b_2 = 1.809$ sugiere que la variedad 2 rinde muy bien en ambientes favorables pero el rendimiento cae muy bruscamente en ambientes pobres, considerando a los ambientes ricos o pobres en relación directa a su producción (índice ambiental). La línea de tendencia forma un ángulo de unos 61° con respecto al eje de las abscisas. La mayoría de los investigadores prefieren un valor de R^2 mayor de 0.9 ($R^2 > 0.9$) y una pendiente de 1.

Redundando un poco, en el diagrama de dispersión para la variedad 3 (Figura 5C), se observan unas distancias de los puntos observados a la línea (desviaciones) considerables.

La línea forma un ángulo de unos 39° con respecto al eje de las abscisas, rinde más o menos bien en ambientes favorables y el rendimiento no se cae mucho en ambientes desfavorables. Presenta un coeficiente de determinación de 0.49 lo que sugiere que el 51 % de la variabilidad total de la variedad 3 del experimento es debida a las desviaciones también llamada variación “no explicada”.

La variación “explicada” debida a las diferencias entre los valores esperados y la media de la variedad, representa solo el 49 % de la variación total.

Para la variedad 4 (figura 4D) se tiene unos puntos en el diagrama de dispersión más o menos ajustados a la línea de tendencia, el 78 por ciento de la variabilidad total de la variedad está explicada por el modelo y la línea de tendencia forma un ángulo de unos 50° con el eje de las abscisas. Aparentemente puede haber una contradicción por que, como se puede observar en el Cuadro 3, no necesariamente la variedad con mayor coeficiente r^2 es el que presenta una suma de cuadrados para las desviaciones de regresión mas baja, esto es debido a que ello depende de la variación total, si esta

es pequeña o grande. Una porción pequeña de una variabilidad total grande puede ser mayor que una porción grande de una variabilidad total pequeña.

La variedad 5 (Figura 5E) presenta un 74 por ciento de la variabilidad total (de la misma variedad) explicada por el modelo, unas desviaciones de regresión no muy considerables y la línea de tendencia tiene unos 45° respecto a las abscisas.

El conjunto de líneas de regresión para las variedades se elabora sobreponiendo todas las líneas para cada una de las variedades. En la figura 5F se puede observar las características de las variedades, anteriormente establecidas, pero ya en conjunto.

	R^2	b_i	\bar{Y}_i	
V1	0.0191	0.1300	20.20	Los valores más bajos para el R^2 son para las variedades 1 y 3, que en el análisis de varianza fueron las variedades cuyas desviaciones de regresión fueron significativas. Las variedades 2 y 4 son las que presentan un valor de R^2 más alto (<i>ver figuras 5-B y 5-D</i>).
V2	0.8797	1.8090	22.25	
V3	0.4933	0.8114	24.00	
V4	0.7834	1.1806	24.67	
V5	0.7446	1.0690	25.50	

El rendimiento de la variedad 2 tiene un promedio bajo y desciende bastante en ambientes desfavorables, aunque rinde muy bien en ambientes buenos; aunque es la “consistente” no es “estable”. La variedad 5 es la de mayor media, tiene una desviación de regresión que no es significativa (consistente), y un coeficiente de regresión muy próximo a la unidad (estable).

Los cálculos se desarrollaron en una hoja de cálculo (EXCEL) y se explicó de manera sencilla e ilustrativa la manera de utilizarlo. Esto tiene mucha utilidad dado que, por lo general, las hojas de cálculo son de uso común y se encuentran en la mayoría de las computadoras comerciales. Las hojas de cálculo normalmente están integradas en paquetes computacionales que por lo general vienen integrados en los equipos de cómputo. De manera especial, en el presente escrito se trabajó sobre la versión **Microsoft Excel 97** aunque los cálculos se pueden realizar en versiones anteriores, solo de debe de tener cuidado de que el paquete esté instalado en forma completa y correcta.

6. APLICACIÓN DEL MÉTODO

Inicialmente se aplicarán los criterios de Eberhart y Russell señalados en su artículo original, detallando cada paso del cálculo y señalando la forma de realizarlo en una hoja de cálculo de fácil acceso como lo es EXCEL.

En el Cuadro 6 están las respectivas fórmulas para el análisis de varianza que se realizará y en el Cuadro 7, se presentan las medias de cuatro repeticiones por ambiente de cada uno de los genotipos y localidades que se proporcionan a continuación.

Datos de CIMMYT de las siguientes variedades de maíz:

- | | | |
|--------------------|---------------------|-----------------------|
| 1. Poza Rica 9022 | 7. Lodana 8929 | 13. Pop 22 X TSR 90 |
| 2. Across 8825 | 8. Poza Rica 8929 | 14. Poza Rica 90301 |
| 3. Baobab 8825 | 9. Across 8843 | 15. Suwan 8222 RE |
| 4. Poza Rica 9025 | 10. Mivuazi 8843 | 16. Farako Ba 8625 RE |
| 5. Guanacaste 8729 | 11. Guatacamas 9043 | |
| 6. Across 8929 | 12. Poza Rica 9043 | |

Los experimentos se realizaron el 1992 bajo el diseño de bloques al azar con 4 repeticiones y las localidades fueron las siguientes:

Turipana, Colombia	Querétaro, México
Antioquia, Colombia	Sinematiali, Costa de Marfil
Porto Viejo, Ecuador	Bako, Etiopia
Capitán Miranda, Paraguay	Potchefstrom, Rep. S. Africa
Maracay, Venezuela	Dholi, India
Central Farm, Belice	Rampur, Nepal
Alajuela, Costa Rica	Faisalabad, Pakistán
Guanacaste, Costa Rica	Song-Boy, Vietnam
Omonita, Honduras	Dan phuong, Vietnam
La Ceiba, Honduras	La Granja, Filipinas
Poza Rica, México	

Cuadro 6. Fórmulas para el Análisis de Varianza

FUENTE	G.L.	SUMA DE CUAD.	CUAD.
MED.			
TOTAL	$av-1$	$\sum_i \sum_j Y_{ij}^2 - C$	
Variedades (V)	$v-1$	$(1/a) \sum_i Y_i^2 - C$	
CM ₁			
VXA	$(v-1)(a-1)$		
Ambiente (lineal)	1	$1/v \left[(\sum_j Y_{.j} I_j)^2 / \sum_j I_j^2 \right]$	
VXA (lineal)	$(v-1)$	$\sum_i \left[(\sum_j Y_{ij} I_j)^2 / \sum_j I_j^2 \right] - \text{Amb}(\text{lin})$	
CM ₂			
Desv. pond.	$v(a-2)$	$\sum_i \sum_j \delta_{ij}^2$	C
M ₃			
Variedad 1	$a-2$	$\left[\sum_j Y_{1j}^2 - (Y_{1.})^2 / a \right] - \left[(\sum_j Y_{1j} I_j)^2 / \sum_j I_j^2 \right]$	
:	:	:	
:	:	:	
Variedad v	$a-2$	$\left[\sum_j Y_{vj}^2 - (Y_{v.})^2 / a \right] - \left[(\sum_j Y_{vj} I_j)^2 / \sum_j I_j^2 \right]$	
Error ponderado	$a(r-1)(v-1)$		

CUCBA



BIBLIOTECA CENTRAL

Cuadro 7. Datos de rendimiento de maíz en kg /Ha. (media de cuatro repeticiones) por localidad, y cálculos para su análisis.

	A	B	C	D	E	F	G
1							
2		POZA 1	ACROSS2	BAOBAB3	POZA_4	GUANAC_5	ACRS_6
3	DHOLI	3448	2233	1424	2504	2261	2782
4	POTCHEFS	3451	3508	3518	3741	3589	2828
5	TURIPANA	3257	3284	3162	3676	3468	3262
6	LA_CEIBA	3504	3240	3478	3853	4280	3097
7	FAISALAB	3493	4522	2987	4767	4088	4537
8	MARACAY	4110	4091	4251	4487	4353	4103
9	SONG_BOY	4764	3850	3978	4221	5502	3760
10	RAMPUR	4633	4672	3970	4777	4065	3713
11	ANTIOQIA	4948	4027	4329	4099	4730	4933
12	DAN_PONG	4613	4759	4191	4295	4323	4344
13	GUANACAS	5241	3790	4423	4616	4710	4890
14	CENTFARM	4621	4811	4668	4746	4763	4697
15	LA_GRANJ	4963	4428	4407	5128	5116	4593
16	BAKO	4961	4378	4892	5089	4979	5127
17	CAPMIRAN	4985	4997	4823	4813	5339	5226
18	SINEMATI	5405	4970	4890	5564	6147	5397
19	POZARICA	6029	5004	5659	5848	6110	5716
20	ALAJUELA	7702	5719	6076	5699	5519	6308
21	OMONITA	6669	6384	6786	6947	6906	6968
22	QUERETAR	8378	7868	7731	8399	8178	8036
23	PTOVIEJO	7641	6893	7673	8005	8259	8125
24	\bar{Y}_i	106816	97428	97316	105274	106685	102442
25	\bar{Y}_i	5086.47619	4639.42857	4634.09524	5013.04762	5080.2381	4878.19048
26	dat gral.	SCT=	730488463	SCV =	15071131	SCA =	648548600
27	$\sum_j Y_{ij} I_j =$	39817732	34825601	42448228	38545896	40078476	41884889
28	$(\sum_j Y_{ij} * I_j) / \sum_j I_j =$	39113843.75	29920902.3	44452540.3	36655043.9	39627790.6	43280492.1
29				$\sum (\bar{Y}_i - \bar{Y}_{..})^2 =$	717672.902		
30	REGRESIÓN						
31	$SCT_i = \sum_j y_{ij}^2 =$	43232969.24	33140125.1	45925529.8	39334043	42570833.8	46189181.2
32	$b_i =$	0.982322237	0.859164	1.04721781	0.95094544	0.98875492	1.03331997
33	$r^2 =$	0.904722582	0.90286026	0.96792656	0.93189108	0.93086715	0.93702661
34	$SCR_i =$	39113843.54	29920902.1	44452540.1	36655043.9	39627790.6	43280492.1
35	$SCDR_i =$	4119125.694	3219223.01	1472989.68	2678999.06	2943043.21	2908689.12
36							
37	SCT_i =	$\sum_j y_{ij}^2$	SCR_i =	$r^2 \sum_j y_{ij}^2$	SCDR_i =	$(1-r^2) \sum_j y_{ij}^2$	

Cuadro 6 (continuación)

	H	I	J	K	L	M	N
1							
2	LODANA_7	POZA_8	ACROS_9	MIVUAZ_10	CATACA_11	POZA_12	"POP__13
3	3483	2250	3053	1136	2523	2335	2389
4	2999	3071	3260	2752	3654	2695	3435
5	4073	3021	3996	3402	3323	4295	3739
6	3810	3592	4031	3385	4164	3207	3494
7	3500	3339	3315	3847	3413	3064	4113
8	4964	3807	3891	3788	4178	3924	4928
9	4687	4040	5684	4416	4436	4547	4191
10	3906	3290	4609	5149	4832	4179	5132
11	4473	4393	4858	4925	4771	5204	4036
12	4646	4799	4451	5242	4749	5085	4870
13	4777	4278	4829	4978	4968	5157	5174
14	4744	4802	4163	4710	4486	4668	4710
15	5065	5470	4311	4244	4737	5205	5034
16	4013	4310	4683	4954	5195	5396	4750
17	4846	4998	5407	5268	5445	5212	4719
18	5248	5413	6171	6470	6083	5816	5199
19	5960	5980	5634	6018	5854	6290	6040
20	6178	6879	6004	4738	6687	6695	6842
21	6674	6828	6906	6533	7049	7016	5650
22	7890	7421	7418	6956	7677	8283	8550
23	8508	7434	9018	8580	9452	8909	8144
24	104444	99415	105692	101491	107676	107182	105139
25	4973.52381	4734.04762	5032.95238	4832.90476	5127.42857	5103.90476	5006.61905
26	Y..=	1642856	Y²..=	8763154644		SCVxA(lin)=	3220356.28
27	38017518	41071283	39525438	41907479	44457667	47153044	39908553
28	35657014.4	41615392.4	38541696.6	43327190.4	48760797.2	54852564.1	39292478.2
29							
30							
31	39463747.2	45180105	43629757	50773643.8	51277133.1	57438631.8	43214731
32	0.93791011	1.01324793	0.97511119	1.03387728	1.09679163	1.16328786	0.98456284
33	0.90353848	0.92109995	0.88338096	0.85334018	0.95092674	0.95497686	0.90923806
34	35657014.4	41615392.4	38541696.6	43327190.4	48760797.2	54852564.1	39292478.2
35	3806732.87	3564712.54	5088060.36	7446453.39	2516335.99	2586067.74	3922252.78
36							
37							

Cuadro 6 (continuación)

	O	P	Q	R	S	T
1						
2	POZA_14	SUWAN_15	FARAKO16	\bar{Y}_j	I_j	Y_j
3	2376	1256	1013	2279.125	-2610.32738	36466
4	2873	2985	2989	3209.25	-1680.20238	51348
5	3258	3799	3513	3533	-1356.45238	56528
6	3051	3866	4102	3634.625	-1254.82738	58154
7	3738	3372	2709	3675.25	-1214.20238	58804
8	4544	5179	3060	4228.625	-660.82738	67658
9	3819	4108	3819	4363.875	-525.57738	69822
10	3108	5406	5273	4419.625	-469.82738	70714
11	4623	5127	3959	4589.6875	-299.76488	73435
12	4448	5076	3955	4615.375	-274.07738	73846
13	3639	4914	3585	4623.0625	-266.38988	73969
14	5105	5374	4416	4717.75	-171.70238	75484
15	4897	4067	3848	4719.5625	-169.88988	75513
16	4746	5299	5382	4884.625	-4.82738	78154
17	4229	5553	5029	5055.5625	166.11012	80889
18	5925	5555	4953	5575.375	685.92262	89206
19	5608	5759	5492	5812.5625	923.11012	93001
20	6704	6695	5452	6243.5625	1354.11012	99897
21	5894	7053	5689	6622	1732.54762	105952
22	8321	7974	6932	7875.75	2986.29762	126012
23	7235	7327	6801	8000.25	3110.79762	128004
24	98141	105744	91971			
25	4673.38095	5035.42857	4379.57143	4889.45238	$\sum_j Y_j \cdot I_j =$	648548602
26				$\sum_j I_j^2 =$	40534287.51	
27	40996811	41549501	36360485			
28	41464611.5	42590141.1	32616458.3	651768957		
29						
30				TOTAL		
31	46709547	48089105.1	39248249.1	715417332		
32	1.01141066	1.02504579	0.89703033			
33	0.8877117	0.88565052	0.83102964			
34	41464611.5	42590141	32616458.3	651768956		
35	5244935.5	5498964.09	6631790.86	63648375.9		
36						
37						

CUCBA



BIBLIOTECA CENTRAL

6.1 Suma de Cuadrados Total.

Es la variación total de un experimento e incluye a las demás variaciones (genotípica, ambiental, error, interacciones). Por definición, es la suma de cuadrados de las desviaciones de cada observación (Y_{ij}) con respecto de la media general ($\bar{Y}_{..}$):

$$SCT = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 \quad 1)$$

Para facilitar el cálculo se utiliza la fórmula que aparece en el Cuadro 6 y que se obtiene de la fórmula 1 (anterior) mediante simplificación algebraica:

$$SCT = \sum_i \sum_j Y_{ij}^2 - C \quad 1a)$$

donde C el factor de corrección que se calcula mediante la fórmula

$$C = \left(\sum_i \sum_j Y_{ij} \right)^2 / av = (Y_{..})^2 / av \quad 2)$$

En la cual a es el número de ambientes y v el número de variedades.

Para el ejemplo, del Cuadro 7, se obtiene la sumatoria general ($Y_{..}$), son 21 ambientes y 16 variedades, $n = 336$ por lo que el valor del factor de corrección es:

$$\sum_i \sum_j Y_{ij} = Y_{..} = 3448 + 3451 + 3257 + 3504 + \dots + 6932 + 6801 = 1\ 642\ 856$$

$$C = (1\ 642\ 856)^2 / 336 = 8\ 032\ 666\ 181$$

Del Cuadro 7 se obtiene que $\sum_i \sum_j Y_{ij}^2 = (3448)^2 + (2233)^2 + \dots + (6801)^2 = 8\ 763\ 154\ 644$

$$SCT = 8\ 763\ 154\ 644 - 8\ 032\ 666\ 181 = 730\ 488\ 463$$

Para el uso de la hoja de cálculo de EXCEL, se puede calcular la SCT mediante la función "DESVIA2" seleccionando todos los datos del experimento. Los datos del estudio están desde la celda B3 hasta la celda Q23 (de B3 a B23 están los datos de la variedad "Poza Rica 9022", de C3 a C23 está la variedad "Across 8825", etc.). por lo que en la celda en que aparecerá el resultado se escribe:

$$=DESVIA2(B3:Q23)$$

6.2 Suma de Cuadrados de Variedades

Es la porción de la variación total del experimento que se explica por el efecto de genotipos. Por definición, es la suma de los cuadrados de las desviaciones de la media de cada variedad (\bar{Y}_i) con respecto a la media general ($\bar{Y}..$) multiplicada por el número de ambientes:

$$SCV = a \sum_i (\bar{Y}_i - \bar{Y}..)^2 \quad 3)$$

$$SCV = 21(717672.9) = 15\ 071\ 130.9$$

Para facilitar el cálculo cuando se realiza de manera manual, se utiliza la fórmula presentada en el Cuadro 6. Se elevan al cuadrado las sumas de cada variedad presentadas en el Cuadro 7 (Y_i), se suman, se dividen entre el número de ambientes y se le resta el factor de corrección (C) calculado con la ecuación 2 mediante la fórmula:

$$SCV = \frac{1}{a} \sum_i Y_i^2 - C \quad 3a)$$

$$SCV = [(106816)^2 + (97428)^2 + (97316)^2 + \dots + (91971)^2] / 21 - 8\ 032\ 666\ 181$$

$$SCV = [169\ 002\ 483\ 500] / 21 - 8\ 032\ 666\ 181 = 15\ 071\ 129$$

Si se observa detenidamente la ecuación 3, se deduce que también se puede calcular obteniendo las desviaciones de las medias varietales elevadas al cuadrado y multiplicadas por el número de ambientes.

En la hoja de cálculo de EXCEL con la función “DESVIA2” se puede realizar lo anteriormente expuesto, se seleccionan las celdas donde se encuentran las medias varietales y se multiplica por el número de ambientes, se escribe en la celda:

$$=DESVIA2(B25:Q25)*21$$

6.3 Suma de Cuadrados del Residual.

Esta suma de cuadrados del residual es igual a la suma de cuadrados total menos la suma de cuadrados de variedades, es decir que incluye la variación debida al ambiente y a la interacción genético-ambiental:

$$SC_{resid} = SCT - SCV = 730\ 488\ 463 - 15\ 071\ 131 = 715\ 417\ 332$$

6.4 Suma de cuadrados de ambientes (lineal)

Es la suma de cuadrados de la regresión de las sumas de ambientes ($Y_{.j}$) como variable dependiente sobre los índices ambientales I_j como variable independiente, dividida entre el número de variedades probadas en cada ambiente. En el Cuadro 6 se presenta la fórmula para su cálculo:

$$SCA(\text{lin}) = 1/v \left[(\sum_j Y_{.j} I_j)^2 / \sum_j I_j^2 \right] \quad 4)$$

Los índices ambientales se obtienen de restarle a cada una de las medias ambientales ($\bar{Y}_{.j}$) la media general ($\bar{Y}_{..}$). Si se considera que son los mismos genotipos en todos los ambientes, la diferencia entre las medias ambientales se deberá básicamente al efecto ambiental, por lo que restado de la manera anteriormente explicada se obtendría un índice ambiental, un parámetro del cómo está influyendo el ambiente sobre el rendimiento de una variedad.

Esto tiene la gran desventaja de que, como ya se discutió anteriormente, existe auto-regresión dado que los índices ambientales están dependiendo directamente de las mismas medias ambientales de rendimiento y matemáticamente impone ciertos criterios de precaución en su manejo.

$$I_j = (\bar{Y}_{.j} - \bar{Y}_{..}) \quad 5)$$

$$I_1 = 2279.125 - 4889.45 = -2610.32$$

$$I_2 = 3209.25 - 4889.45 = -1680.20$$



$$I_3 = 3533 - 4889.45 = -1356.45$$

: : : : :

$$I_{21} = 8000.25 - 4889.45 = 3110.80$$

$$\sum_j I_j^2 = (-2610.33)^2 + (-1680.20)^2 + (-1356.45)^2 + \dots + (3110.8)^2 = 40\,534\,287.5$$

En el Cuadro 7 se presenta el índice mencionado para cada ambiente en la columna "S", de la celda 3 a la celda 23. La suma de los cuadrados de los índices ambientales ($\sum_j I_j^2$), se obtiene insertando en la celda en que aparecerá el resultado la función:

$$=SUMA.CUADRADOS(S3:S23)$$

$\sum_j Y_{.j} I_j$ Es la suma de los productos del índice del ambiente "j" por la suma de todas las variedades en el ambiente "j". Las sumas de ambientes están en la columna "T".

$$\sum_j Y_{.j} I_j = (36466)(-2610.33) + (51348)(-1680.20) + \dots + (128004)(3110.8) = 648548$$

602

Para obtener lo anterior en una hoja de cálculo se escribe en la celda en la que debe aparecer el resultado lo siguiente:

$$=SUMAPRODUCTO(S3:S23;T3:T23)$$

De la celda S3 a S23 están los índices ambientales, y de T3 a T23 están las sumas de ambientes.

$$SCA(\text{lin}) = 1/16 [(648\,548\,602)^2 / (40\,534\,288)] = 648\,548\,603$$

Aquí cabe hacer la aclaración que si se utiliza la fórmula clásica para la suma de cuadrados de ambientes, que por definición es la suma de los cuadrados de las desviaciones de la media de cada localidad ($\bar{Y}_{.j}$) con respecto a la media general ($\bar{Y}_{..}$) multiplicada por el número de variedades se obtiene el mismo resultado:

$$SCA = v \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 \quad 6)$$

$$SCA = 16(40\ 534\ 288) = 648\ 548\ 600$$

mas comúnmente calculada con la fórmula

$$SCA = (1/v) \sum_j Y_{.j}^2 - C \quad 6a)$$

$$SCA = [(36466)^2 + (51348)^2 + (56528)^2 + \dots + (128004)^2] / 16 - 8\ 032\ 666\ 181$$

$$SCA = (8\ 681\ 214\ 781 - 8\ 032\ 666\ 181) = 648\ 548\ 600$$

Pero hay que tomar en cuenta que para el primero (ecuación 4) existe solo un grado de libertad, mientras que para el segundo (ecuación 6) habría 20 grados de libertad.

En sí, esta prueba es solo para demostrar que la suma de cuadrados para el ambiente lineal solo tiene un grado de libertad. Si se observa la formula 4, $\sum_j I_j^2$ son,

propiamente dicho, las desviaciones al cuadrado de los ambientes por lo que si se multiplica por el número de variedades se obtiene la suma de cuadrados de ambientes, lo que está expresado en la ecuación 4 la cual empieza precisamente con el recíproco de v . Por otro lado, la expresión $\sum_j Y_{.j} I_j$ También resulta la suma de cuadrados de ambientes (SCA) dado que:

$$\sum_j Y_{.j} I_j = \sum_j Y_{.j} (\bar{Y}_{.j} - \bar{Y}_{..}) = \sum_j Y_{.j} \left(\frac{Y_{.j}}{v} - \frac{Y_{..}}{av} \right) = (1/v) \sum_j Y_{.j}^2 - C$$

Y la última parte (encerrada en el rectángulo) corresponde a la ecuación 6a.

Para calcular la SCA en EXCEL, se obtiene mediante la función "DESVIA2" señalando las celdas en las que se encuentran las medias ambientales y se multiplica por el número de variedades. En la celda en la que debe aparecer el resultado se escribe:

$$=DESVIA2(R3:R23)*16$$

6.5 Suma de Cuadrados de Variedad por Ambiente (lineal)

$$SCV_{xA} \text{ (lineal)} = \sum_i \left[\frac{(\sum_j Y_{ij} I_j)^2}{\sum_j I_j^2} \right] - SCA(\text{lin}) \quad 7)$$

De donde $\left[\frac{(\sum_j Y_{ij} I_j)^2}{\sum_j I_j^2} \right]$ es la suma de cuadrados de la regresión de la variedad "i" sobre I_j . En el Cuadro 7, la fila $\sum_j Y_{ij} I_j$ está formada de igual manera que para el ejemplo sencillo con el que se inició. Cada una de las observaciones (Y_{ij}) de cada variedad se multiplicó por su respectivo índice ambiental y se sumaron todos los productos. Obsérvese a continuación:

$$\sum_j Y_{1j} I_j = Y_{11} I_1 + Y_{12} I_2 + Y_{13} I_3 + \dots + Y_{1,21} I_{21}$$

$$\sum_j Y_{1j} I_j = (3448)(-2610.33) + (3451)(-1680.2) + \dots + (7641)(3110.8) = 39\,817\,732$$

$$\sum_j Y_{2j} I_j = (2233)(-2610.33) + (3508)(-1680.2) + \dots + (6893)(3109.9) = 34\,825\,600$$

$$\sum_j Y_{3j} I_j = 42\,448\,228$$

$$\sum_j Y_{10j} I_j = 41\,907\,479.1$$

$$\sum_j Y_{4j} I_j = 38\,545\,896$$

$$\sum_j Y_{11j} I_j = 44\,457\,667.3$$

$$\sum_j Y_{5j} I_j = 40\,078\,476.3$$

$$\sum_j Y_{12j} I_j = 47\,153\,044.6$$

$$\sum_j Y_{6j} I_j = 41\,884\,888.9$$

$$\sum_j Y_{13j} I_j = 39\,908\,553.2$$

$$\sum_j Y_{7j} I_j = 38\,017\,518.1$$

$$\sum_j Y_{14j} I_j = 40\,996\,810.7$$

$$\sum_j Y_{8j} I_j = 41\,071\,283$$

$$\sum_j Y_{15j} I_j = 41\,549\,501$$

$$\sum_j Y_{9j} I_j = 39\,525\,437.6$$

$$\sum_j Y_{16j} I_j = 36\,360\,485.5$$

Este cálculo se puede efectuar mediante la función "SUMAPRODUCTO". Por ejemplo, para obtener $\sum_j Y_{1j} I_j$, en la celda en la que debe de aparecer el resultado se escribe:

$$=SUMAPRODUCTO(B3:B23;S3:S23)$$

Para el caso $\sum_j Y_{2j} I_j$ se escribe:

$$=SUMAPRODUCTO(C3:C23;S3:S23)$$

Y así sucesivamente.

En el siguiente renglón, del mismo Cuadro 7, se encuentra $\left[\frac{(\sum_j Y_{ij} I_j)^2}{\sum_j I_j^2} \right]$, en el que cada una de las sumas anteriores es elevada al cuadrado y dividida entre $\sum_j I_j^2$, como se indica enseguida:

$$\left(\frac{\sum_j Y_{1j} I_j}{\sum_j I_j} \right)^2 = \frac{(39817730)^2}{40534287.5} = 39113843.8 \quad \left(\frac{\sum_j Y_{9j} I_j}{\sum_j I_j} \right)^2 = 38541696.8$$

$$\left(\frac{\sum_j Y_{2j} I_j}{\sum_j I_j} \right)^2 = 29920902.3 \quad \left(\frac{\sum_j Y_{10j} I_j}{\sum_j I_j} \right)^2 = 43327190.4$$

$$\left(\frac{\sum_j Y_{3j} I_j}{\sum_j I_j} \right)^2 = 44452540.3 \quad \left(\frac{\sum_j Y_{11j} I_j}{\sum_j I_j} \right)^2 = 48760797.2$$

$$\left(\frac{\sum_j Y_{4j} I_j}{\sum_j I_j} \right)^2 = 36655043.9 \quad \left(\frac{\sum_j Y_{12j} I_j}{\sum_j I_j} \right)^2 = 54852564.1$$

$$\left(\frac{\sum_j Y_{5j} I_j}{\sum_j I_j} \right)^2 = 39627790.6 \quad \left(\frac{\sum_j Y_{13j} I_j}{\sum_j I_j} \right)^2 = 39292478.2$$

$$\left(\frac{\sum_j Y_{6j} I_j}{\sum_j I_j} \right)^2 = 43280492.1 \quad \left(\frac{\sum_j Y_{14j} I_j}{\sum_j I_j} \right)^2 = 41464611.5$$

$$\left(\frac{\sum_j Y_{7j} I_j}{\sum_j I_j} \right)^2 = 35657014.4 \quad \left(\frac{\sum_j Y_{15j} I_j}{\sum_j I_j} \right)^2 = 42590141.3$$

$$\left(\frac{\sum_j Y_{8j} I_j}{\sum_j I_j} \right)^2 = 41615392.4 \quad \left(\frac{\sum_j Y_{16j} I_j}{\sum_j I_j} \right)^2 = 32616458.3$$

Todos estos resultados se suman para obtener finalmente:

$$\sum_i \left[\frac{(\sum_j Y_{ij} I_j)^2}{\sum_j I_j^2} \right] = 651\,768\,960$$

$$SCV_{xA}(\text{lin}) = \sum_i \left[\frac{(\sum_j Y_{ij} I_j)^2}{\sum_j I_j^2} \right] - SCA(\text{lin}) = 651768960 - 648548603 = 3220356.3$$

CUCBA



BIBLIOTECA CENTRAL

6.6 Suma de Cuadrados de las Desviaciones Ponderadas

La SCDP es la suma de las desviaciones de regresión de cada variedad sobre los índices ambientales.

$$SCDP = \sum_i \sum_j \delta_{ij}^2 = SC_{resid} - Amb(lin) - SC_{VxA}(lin) \quad 8)$$

$$SCDP = 715417332 - 648548603 - 3220356.3 = 63648372.7$$

La SCDP se descompone en las sumas de cuadrados de desviaciones de regresión para cada una de las variedades.

La suma de cuadrados de las desviaciones de regresión para la i -ésima variedad es igual a la variación total para la variedad " i " ($\sum_j y_{ij}^2 = SCT_i$), menos la suma de cuadrados de la regresión de la misma variedad sobre el índice ambiental (SCR_i) (ver Cuadro 7.) Lo anterior está explicado detenidamente en las páginas 30 y 31 del presente trabajo.

$$SCDR_i = \left[\sum_j Y_{ij}^2 - (Y_{i.})^2 / a \right] - \left[(\sum_j Y_{ij} I_j)^2 / \sum_j I_j^2 \right] \quad 8a)$$

$$SCDR_i = SCT_i - SCR_i = \sum_j d_{ij}^2 = \sum_j (Y_{ij} - \hat{Y}_{ij})^2$$

Y_{ij} es el valor observado e \hat{Y}_{ij} es el valor esperado por la ecuación de regresión.

La suma de cuadrados total para la i -ésima variedad, se puede calcular mediante la fórmula definición:

$$SCT_i = \sum_j y_{ij}^2 = \sum_j (Y_{ij} - \bar{Y}_{i.})^2 = \sum_j Y_{ij}^2 - Y_{i.}^2 / a$$

Siguiendo la fórmula anterior, se tiene para la primera variedad:

$$SCT_1 = \sum_j Y_{1j}^2 - Y_{1.}^2 / a = (3448^2 + 3451^2 + 3257^2 + \dots + 7641^2) - (3448 + 3451 + 3257 + \dots + 7641)^2 / 21 = 43\,232\,969$$

El cálculo se realiza fácilmente escribiendo en la celda deseada:

$$=DESVIA2(B3:B23)$$

Ya que se están calculando las desviaciones al cuadrado de los datos que se encuentran en el rango de celdas (B3:B23). Para las siguientes variedades, se puede "arrastrar" el puntero del ratón, como se explico en la página 22, tantas celdas como datos haya que calcular

En el Cuadro 7, en el renglón "SCT_i", están los resultados del anterior cálculo y para las siguientes variedades; y en el renglón "SCDR_i" están las sumas de cuadrados de las desviaciones de regresión para cada variedad ($\sum_j d_{ij}^2$) que se calculan mediante la fórmula 8a, teniendo en cuenta que ya se calcularon las sumas de cuadrados de la regresión, $\left(\sum_j Y_{ij} I_j\right)^2 / \sum_j I_j^2$, para cada variedad en el punto 6.5:

$$\sum_j d_{1j}^2 = 43\,232\,969 - 39\,113\,823.39 = 4\,119\,125.7 \quad \sum_j d_{9j}^2 = 5\,088\,060.4$$

$$\sum_j d_{2j}^2 = 33\,140\,125 - 29\,920\,887 = 3\,219\,223 \quad \sum_j d_{10j}^2 = 7\,446\,453.4$$

$$\sum_j d_{3j}^2 = 1\,472\,989.7 \quad \sum_j d_{11j}^2 = 2\,516\,336$$

$$\sum_j d_{4j}^2 = 2\,678\,999 \quad \sum_j d_{12j}^2 = 2\,586\,067.7$$

$$\sum_j d_{5j}^2 = 2\,943\,043.2 \quad \sum_j d_{13j}^2 = 3\,922\,252.8$$

$$\sum_j d_{6j}^2 = 2\,908\,689 \quad \sum_j d_{14j}^2 = 5\,244\,935.5$$

$$\sum_j d_{7j}^2 = 3\,806\,732.9 \quad \sum_j d_{15j}^2 = 5\,498\,964$$

$$\sum_j d_{8j}^2 = 3\,564\,712.5 \quad \sum_j d_{16j}^2 = 6\,631\,790.9$$

La suma general es **63 648 372.6** que es, como se mencionó con anterioridad, la suma de cuadrados de las desviaciones ponderadas.

6.7 Cuadrado Medio del Error Ponderado (CMEP).

Se obtiene de sumar las sumas de cuadrados del error experimental de los análisis de varianza efectuados para cada localidad y dividirla entre la suma de los grados de libertad de cada uno de los experimentos. Para éste caso, se presentan los coeficientes de variación para cada localidad. Sabiendo que:

$$C.V. = (See/\bar{Y})(100) \quad 9)$$

$$\text{se despeja } See = (\bar{Y})CV/100$$

See se eleva al cuadrado para obtener el cuadrado medio del error y se multiplica por los grados de libertad para obtener la suma de cuadrados del error ponderado. Como se mencionó con anterioridad, en 21 ambientes se probaron 16 variedades con 4 repeticiones por ambiente. Para cada localidad, por lo tanto, hay $(16 \times 4) - 1 = 63$ grados de libertad total, 15 para variedades y 3 para repeticiones, por lo tanto hay $63 - 15 - 3 = 45$ grados de libertad para el error en cada experimento, o bien $(16-1)(4-1) = 45$. Por ejemplo, para Dholi, India se tiene una media de 2279.125 y un C.V.= 24.4: $2279.125 \times 24.4/100 = 556.1065$ y $(556.1065)^2 \times 45 = 13\,916\,449.77$

556.1065 que sería la desviación estándar del error, se eleva al cuadrado para obtener el cuadrado medio del error y se multiplica por 45 para finalmente tener la suma de cuadrados del error para la localidad mencionada.

Posteriormente se suman y se dividen entre la suma de los grados de libertad, obteniendo así el Cuadrado Medio para el Error Ponderado (*ver Cuadro 8*).

$498465119.8 / 945 = 527\,476.3172$, que obviamente se podía calcular simplemente dividiendo la suma de S_{ee}^2 entre el número de variedades, 21, obteniendo el mismo resultado:

$11077002,66 / 21 = 527476.3172$ por lo que es pertinente señalar que no importa el número de repeticiones que halla sido en los experimentos por localidad, el resultado es el mismo para el Cuadrado Medio del Error Ponderado.

Cuadro 8. Cálculo del cuadrado medio para el error ponderado (CMEP).

LOCALIDAD	MEDIA	C.V.	S_{ee}	S_{ee}^2	SC _{ee}	GL.
DHOLI	2279.1	24.4	556.1	309254.4	13916449.8	45
POTCHEFSTROM	3209.3	16.7	535.9	287236.8	12925654.9	45
TURIPANA	3533.0	19.7	696.0	484417.4	21798782.6	45
LA_CEIBA	3634.6	17	617.9	381783.4	17180253.8	45
FAISALABAD	3675.3	21.5	790.2	624382.5	28097210.6	45
MARACAY	4228.6	24.3	1,027.6	1055871.1	47514198.4	45
SONG-BOY	4363.9	17	741.9	550354.4	24765948.2	45
RAMPUR	4419.6	22.3	985.6	971360.8	43711235.6	45
ANTIOQUIA	4589.7	18	826.1	682513.5	30713107.3	45
DAN-PHUONG	4615.4	10.4	480.0	230399.0	10367956.8	45
GUANACASTE	4623.1	15.6	721.2	520126.2	23405678.8	45
CENTRAL FARM	4717.8	12.2	575.6	331275.6	14907404	45
LA GRANJA	4719.6	18.6	877.8	770600.7	34677029.3	45
BAKO	4884.6	19.4	947.6	897978.5	40409030.4	45
CAP. MIRANDA	5055.6	8.5	429.7	184661.7	8309776.3	45
SINEMATIALI	5575.4	8.9	496.2	246222.8	11080023.8	45
POZA RICA	5812.6	7.4	430.1	185011.3	8325510.08	45
ALAJUELA	6243.6	12	749.2	561341.8	25260383.1	45
OMONITA	6622.0	8.9	589.4	347342.9	15630428.4	45
QUERETARO	7875.8	7.4	582.8	339662.3	15284801.3	45
PORTO VIEJO	8000.3	13.2	1,056.0	1115205.7	50184256.4	45
				11077002.7	498465120	945
				CMEP = 527476.317		

CUCBA



BIBLIOTECA CENTRAL

6.8 Cuadrados Medios y valores de "F".

Aceptando que las poblaciones con las que se trabaja se distribuyen normalmente, se procede a lo siguiente: los cuadrados medios o varianzas para cada fuente de variación, se obtienen de dividir las sumas de cuadrados entre sus respectivos grados de libertad. El cuadrado medio para variedades, se obtiene de dividir la suma de cuadrados de variedades entre los grados de libertad de variedades:

$$CM_1 = \frac{15071131}{15} = 1004742$$

De igual manera para el cuadrado medio de la interacción VXA (CM_2) y el cuadrado medio para las desviaciones ponderadas (CM_3).

Los valores de "Fc", para variedades se obtiene de dividir el CM_1 entre el CM_3 , y la "Fc" para la interacción se obtiene dividiendo CM_2 entre CM_3

La probabilidad de que un valor aleatorio de "F" ocurra, aceptando la hipótesis nula, $P(Fc)$, se obtiene en el programa EXCEL mediante la función "DISTR.F". El cuadro de análisis de varianza (Cuadro 9) se colocó en una hoja de cálculo y la función se inserta de la siguiente manera: en la celda en la que aparecerá el dato buscado se escribe:

$$=DISTR.F(E3;B3;B7)$$

En la celda E3 está el valor de Fc para el cual se busca la probabilidad; en la celda B3 están los grados de libertad del numerador, en éste caso los grados de libertad para las variedades (15); y en la celda B7 están los grados de libertad para el denominador, o sea los grados de libertad para las desviaciones ponderadas (304). También se puede escribir en la celda correspondiente insertando el valor numérico en lugar de las celdas, por ejemplo:

$$=DISTR.F(4.8;15;305)$$

El valor que arroja es $2.3 \text{ E-}8 = 2.3 \times 10^{-8}$, que es la probabilidad de que un valor aleatorio para "F" sea igual o mayor que 4.8 siendo cierta la hipótesis nula. Por ello, se infiere que las variedades no provienen de la misma población y que la diferencia es "altamente significativa".

Cuadro 9. Análisis de varianza del experimento.

FUENTE	G.L.	S.C.	C.M.	Fc	P(Fc)
TOTAL	335	730488463			
Variedades (V)	15	15071131	1004742.1	4.80	2.3E-08 **
Residual		715417332			
Ambiente (lineal)	1	648548600			
VXA (lineal)	15	3220356	214690.4	1.03	0.43 N.S
Desv. pond.	304	63648376	209369.7		
VAR 1	19	4119126	216796.1	0.41	0.98815 N.S.
VAR 2	19	3219223	169432.8	0.32	0.99758 N.S.
VAR 3	19	1472990	77525.8	0.15	0.99999 N.S.
VAR 4	19	2678999	141000.0	0.27	0.99933 N.S.
VAR 5	19	2943043	154897.0	0.29	0.99870 N.S.
VAR 6	19	2908689	153088.9	0.29	0.99880 N.S.
VAR 7	19	3806733	200354.4	0.38	0.99273 N.S.
VAR 8	19	3564713	187616.4	0.36	0.99523 N.S.
VAR 9	19	5088060	267792.7	0.51	0.96033 N.S.
VAR 10	19	7446453	391918.6	0.74	0.77557 N.S.
VAR 11	19	2516336	132438.7	0.25	0.99958 N.S.
VAR 12	19	2586068	136108.8	0.26	0.99948 N.S.
VAR 13	19	3922253	206434.4	0.39	0.99123 N.S.
VAR 14	19	5244936	276049.2	0.52	0.95343 N.S.
VAR 15	19	5498964	289419.2	0.55	0.94064 N.S.
VAR 16	19	6631791	349041.6	0.66	0.85832 N.S.
E. POND.	945	498465120	527476.3		

CUCBA



BIBLIOTECA CENTRAL

6.9 Cálculo de los Coeficientes de Regresión. (b_i)

$$b_i = \frac{\sum_j Y_{ij} I_j}{\sum_j I_j^2}$$

$\sum_j I_j^2$ se calculó en el punto 6.4 y $\sum_j Y_{ij} I_j$ en el punto 6.5:

$$b_1 = \frac{\sum_j Y_{1j} I_j}{\sum_j I_j^2} = 39817732/40534287.51 = 0.9823$$

$$b_2 = \frac{\sum_j Y_{2j} I_j}{\sum_j I_j^2} = 34825801/40534287.51 = 0.8592$$

$$b_3 = 42448228/40534287.51 = 1.0472$$

$$b_{10} = 41907479/40534287.51 = 1.0339$$

$$b_4 = 38545896/40534287.51 = 0.9509$$

$$b_{11} = 44457667/40534287.51 = 1.0968$$

$$b_5 = 40078476/40534287.51 = 0.9888$$

$$b_{12} = 47153044/40534287.51 = 1.1633$$

$$b_6 = 41884889/40534287.51 = 1.0333$$

$$b_{13} = 39908553/40534287.51 = 0.9846$$

$$b_7 = 38017518/40534287.51 = 0.9379$$

$$b_{14} = 40996811/40534287.51 = 1.0114$$

$$b_8 = 41071283/40534287.51 = 1.0132$$

$$b_{15} = 41549501/40534287.51 = 1.0250$$

$$b_9 = 39525438/40534287.51 = 0.9751$$

$$b_{16} = 36360485/40534287.51 = 0.8970$$

En la hoja de Cálculo "EXCEL", el cálculo se realiza mediante la función "ESTIMACIÓN LINEAL", escribiendo para la variedad 1 en la celda deseada:

$$=ESTIMACION.LINEAL(B3:B23;S3:S23)$$

Lo anterior dado que en el rango de celdas (B3:B23) están los rendimientos de la variedad 1 y en (S3:S23) están los índices ambientales. Para la variedad 2 se escribe:

$$=ESTIMACION.LINEAL(C3:C23;S3:S23)$$

6.10 Cálculo de las Desviaciones de Regresión. ($S^2_{d_i}$)

Las desviaciones de regresión se calculan mediante la fórmula:

$$S^2_{d_i} = \sum_j \frac{\hat{d}_{ij}^2}{(a-2)} - S^2_{e/r} \quad (15)$$

Donde $\sum_j \frac{\hat{d}_{ij}^2}{(a-2)}$ es el cuadrado medio de la desviación de regresión de cada variedad (desviaciones ponderadas, obtenidas en el punto 6.6) y $S^2_{e/r}$ es el cuadrado medio del error ponderado previamente calculado en el punto 6.7.

$$S^2_{d_1} = 216\,796.3 - 527\,476.32 = 0$$

$$S^2_{d_9} = 267\,792.9 - 527\,476.32 = 0$$

$$S^2_{d_2} = 169\,433.0 - 527\,476.32 = 0$$

$$S^2_{d_{10}} = 391\,918.8 - 527\,476.32 = 0$$

$$S^2_{d_3} = 77\,526.0 - 527\,476.32 = 0$$

$$S^2_{d_{11}} = 132\,439.0 - 527\,476.32 = 0$$

$$S^2_{d_4} = 141\,000.2 - 527\,476.32 = 0$$

$$S^2_{d_{12}} = 136\,109.1 - 527\,476.32 = 0$$

$$S^2_{d_5} = 154\,897.2 - 527\,476.32 = 0$$

$$S^2_{d_{13}} = 206\,434.6 - 527\,476.32 = 0$$

$$S^2_{d_6} = 153\,089.1 - 527\,476.32 = 0$$

$$S^2_{d_{14}} = 276\,049.4 - 527\,476.32 = 0$$

$$S^2_{d_7} = 200\,354.6 - 527\,476.32 = 0$$

$$S^2_{d_{15}} = 289\,419.4 - 527\,476.32 = 0$$

$$S^2_{d_8} = 187\,616.7 - 527\,476.32 = 0$$

$$S^2_{d_{16}} = 349\,041.8 - 527\,476.32 = 0$$

Los residuales no son significativos. Se concluye que el modelo utilizado sí representa a la realidad observada en los experimentos, es decir, los datos analizados se ajustan al modelo lineal. La diferencia entre las medias de variedades es altamente significativa y la interacción no es significativa, por lo que se puede decir que el modelo puede servir para realizar las inferencias debidas.

Los coeficientes de regresión son casi iguales a 1 (*ver Cuadro 10*), pero no existen pruebas estadísticas para probar si son iguales a 1 o no. Con base en las pruebas anteriores, se puede decir que todas las variedades son consistentes dado que las desviaciones fueron iguales a cero. Se puede también decir que las variedades rinden de acuerdo al ambiente. Observando el Cuadro 10 se tienen valores para los coeficientes de regresión, b , muy cercanos a la unidad, aunque las variedades 11 y 12 presentan valores de b poco mayores de la unidad, por lo que se infiere que rinden mejor en ambientes favorables y su rendimiento baja más que el promedio en ambientes menos favorables, aunque las medias de éstas dos variedades son las más altas, en especial la variedad 11 cuya media es la mayor y $b = 1.097$, las hace recomendables para ambientes potencialmente buenos (*ver figuras 6-11 Y 6-12*). Las variedades 2 y 16, por presentar los valores más bajos de b , presentan menos disminución de su producción en ambientes desfavorables aunque sus medias son de las menores, pero no responde muy bien a buenos ambientes. La elección de la variedad adecuada depende de los propósitos, de los ambientes, de los parámetros de estabilidad y de las medias de las variedades.

Cuadro 10. Medias y coeficientes de regresión para las variedades

Variedades	Media	Bi
V1	5086.5	0.98
V2	4639.4	0.86
V3	4634.1	1.05
V4	5013.0	0.95
V5	5080.2	0.99
V6	4878.2	1.03
V7	4973.5	0.94
V8	4734.0	1.01
V9	5033.0	0.98
V10	4832.9	1.03
V11	5127.4	1.10
V12	5103.9	1.16
V13	5006.6	0.98
V14	4673.4	1.01
V15	5035.4	1.03
V16	4379.6	0.90

CUCBA



BIBLIOTECA CENTRAL

7. UN MÉTODO SENCILLO.

Con base en los análisis de varianza de las regresiones de cada variedad sobre los índices ambientales, se obtienen las sumas de cuadrados del análisis de varianza del Cuadro 9. Los cálculos están hechos en el Cuadro 7 y se ordenan en el Cuadro 11. Antes de proceder con los cálculos, es conveniente establecer lo siguiente:

SCT_i	Es la variación total de la i -ésima variedad. (Suma de Cuadrados Total para la i -ésima variedad).
SCR_i	La Suma de Cuadrados de la Regresión para la i -ésima variedad sobre los índices ambientales.
$SCDR_i$	Es la Suma de Cuadrados de las Desviaciones de Regresión para la i -ésima variedad sobre los índices ambientales.
$CMDR_i$	Es el Cuadrado Medio de las Desviaciones de Regresión que se obtiene de dividir la $SCDR$ entre los grados de libertad ($a-2$)
r_i^2	Es el Coeficiente de Determinación, que especifica la porción de la variación total que es explicada por el modelo utilizado.
b_i	Es el Coeficiente de Regresión del rendimiento sobre el índice ambiental, la relación de cambio del rendimiento al variar el ambiente.
F_c (Reg)	Es la F calculada para la regresión, se obtiene de dividir el cuadrado medio de la regresión, que es el mismo valor de la SCR dado que hay un solo grado de libertad para la regresión, entre $CMDR$
P (F)	Probabilidad de un valor aleatorio de F igual o mayor que el observado
F_c (DReg)	Valor calculado de "F" para las desviaciones de regresión. Se obtiene de dividir $CMDR$ entre el cuadrado medio del error ponderado (CMEP)
CMEP	Cuadrado medio del Error Ponderado. (Punto 6.7).

Cuadro 11. Análisis de regresión de cada variedad sobre los índices ambientales.

Variedades	SCT _i	³ SCR _i	SCDR _i	CMDR _i	Fc (Reg)	P(F) (Reg)	Fc (DReg)	P(F) (DReg)
V1	43232969	39113844	4119126	216796	180.4	3.774E-11	0.4110	0.98815
V2	33140125	29920902	3219223	169433	176.6	4.541E-11	0.3212	0.99758
V3	45925530	44452540	1472990	77526	573.4	1.179E-15	0.1470	0.99999
V4	39334043	36655044	2678999	141000	260.0	1.535E-12	0.2673	0.99933
V5	42570834	39627791	2943043	154897	255.8	1.77E-12	0.2937	0.99870
V6	46189181	43280492	2908689	153089	282.7	7.271E-13	0.2902	0.99880
V7	39463747	35657014	3806733	200354	178.0	4.247E-11	0.3798	0.99273
V8	45180105	41615392	3564713	187616	221.8	6.24E-12	0.3557	0.99523
V9	43629757	38541697	5088060	267793	143.9	2.603E-10	0.5077	0.96033
V10	50773644	43327190	7446453	391919	110.6	2.332E-09	0.7430	0.77557
V11	51277133	48760797	2516336	132439	368.2	6.757E-14	0.2511	0.99958
V12	57438632	54852564	2586068	136109	403.0	2.975E-14	0.2580	0.99948
V13	43214731	39292478	3922253	206434	190.3	2.374E-11	0.3914	0.99123
V14	46709547	41464611	5244936	276049	150.2	1.813E-10	0.5233	0.95343
V15	48089105	42590141	5498964	289419	147.2	2.157E-10	0.5487	0.94064
V16	39248249	32616458	6631791	349042	93.4	9.06E-09	0.6617	0.85832
Total	715417332	651768956	63648376					

³ El cuadrado medio para la regresión (CMR_i) es igual que la suma de cuadrados de la regresión (SCR_i) dado que hay un solo grado de libertad para la regresión.

Hablando en términos de desviaciones, es muy conveniente tener presente que las letras cursivas minúsculas simbolizan desviaciones:

$$SCT_i = \sum_j y_{ij}^2 = \sum_j (Y_{ij} - \bar{Y}_{i.})^2 = \sum_j Y_{ij}^2 - Y_{i.}^2 / a \quad 10)$$

Dado que $\sum_j I_j = 0$, si se habla en términos de desviaciones, $\sum_j i_j^2 = \sum_j I_j^2$

$$r_i^2 = (\sum_j Y_{ij} I_j)^2 / \sum_j y_{ij}^2 \sum_j I_j^2 \quad 11)$$

$$b_i = \sum_j Y_{ij} I_j / \sum_j I_j^2 \quad 12)$$

$$SCR_i = r_i^2 \sum_j y_{ij}^2 \quad 13)$$

$$SCDR_i = (1 - r_i^2) \sum_j y_{ij}^2 \quad 14)$$

El cuadrado medio de la regresión (CMR_i) es igual que la suma de cuadrados de la regresión (SCR_i) dado que solo hay un grado de libertad

$CMDR_i$ Se obtiene al dividir $SCDR_i$ entre los grados de libertad, que para todos los casos son 19, ($a-2$).

7.1 Cálculos para la regresión de variedades sobre índices ambientales.

7.1.1 Suma de Cuadrados Total para las regresiones (SCT_i).

Para la Suma de Cuadrados Total para la Variedad 1, se sustituyendo la "i" por "1" en la ecuación 10):

$$SCT_1 = \sum_j y_{1j}^2 = \sum_j Y_{1j}^2 - (Y_{1.}^2 / a) = 586\,550\,010 - (106816)^2 / 16 = 43\,232\,969.24$$

El cálculo anterior es muy sencillo, como ya se ha explicado, en EXCEL mediante la función "DESVIA2". Basándose en el Cuadro 7, en la celda que se quiere que aparezca el resultado se escribe:

$$=DESVIA2(B3:B23)$$

Dado que los datos de la variedad uno están en las celdas de B3 a B23.

Los cálculos para las siguientes variedades se puede realizar de igual forma o bien, se coloca la punta del cursor del “ratón” en la esquina inferior derecha de la celda en la que aparece el primer resultado y, manteniendo presionado el botón izquierdo del “ratón”, se “arrastra” hacia la derecha tantas celdas como variedades haya. Los resultados se disponen en el Cuadro 11.

7.1.2 Coeficiente de Determinación (r^2).

Para el cálculo de éste coeficiente, se tiene la ecuación 11:

$$r_1^2 = \left(\sum_j Y_{1j} I_j \right)^2 / \sum_j y_{ij}^2 \sum_j I_j^2 = (39817732.1)^2 / (43232969.24)(40534287.5) = 0.90472259$$

El coeficiente de determinación (r^2) se obtiene en EXCEL escribiendo en la celda en la que se desea que aparezca el resultado lo siguiente:

$$=COEFICIENTE.R2(B3:B23;S3:S23)$$

Dado que los datos de la variedad 1 están en las celdas de B3 a B23 y los índices ambientales están desde S3 hasta S23 (ambos con referencia al Cuadro 7).

Para obtener el coeficiente de determinación r^2 de la variedad 2 se escribe en la celda correspondiente:

$$=COEFICIENTE.R2(C3:C23;S3:S23)$$

Dado que la variedad 2 está en las celdas de C3 hasta C23 y así sucesivamente se calculan para todas las variedades. En éste caso, el “arrastrar” con el ratón la celda *no* funciona.

7.1.3 Coeficiente de Regresión (b_i).

Para el cálculo de éste parámetro de estabilidad para la variedad 1, se tiene fórmula 12:

$$b_1 = \sum_j Y_{1j} I_j / \sum_j I_j^2 = 39\,817\,732.1 / 40\,534\,287.5 = 0.98232224$$

Para obtenerlo en EXCEL, en la celda en que se desea que aparezca el coeficiente de regresión (b_i), se escribe lo siguiente.

$$=ESTIMACION.LINEAL(B3:B23;S3:S23)$$

Dado que, como se acaba de indicar, B3:B23 contiene los datos de la variedad 1 y S3:S23 contiene los índices ambientales. Para obtener el coeficiente de regresión (b) de la variedad 2 se escribe:

$$=COEFICIENTE.R2(C3:C23;S3:S23)$$

Y sucesivamente se obtienen todos los coeficientes para las variedades.

7.1.4 *Suma de Cuadrados de la Regresión (SCR_i).*

Para la variedad 1 se calcula con la fórmula 13:

$$SCR_1 = r_1^2 \sum_j y_{1j}^2 = 0.90472259 (43\ 232\ 969.24) = 39\ 113\ 843.8$$

Ambos resultados ya se tienen por lo que se pide al programa que los multiplique escribiendo en la celda en que aparecerá el resultado lo siguiente:

$$=B31*B33$$

En la celda B31 está la SCT para la variedad 1 y en la celda B33 está el coeficiente de determinación para la misma variedad 1.

Para la variedad 2 se escribe en la celda deseada:

$$=C31*C33$$

Para éste cálculo sí funciona el “arrastrar” la celda como ya se explicó. Los resultados se disponen también en el Cuadro 11.

7.1.5 *Suma de Cuadrados de las Desviaciones de Regresión (SCDR_i).*

Se calcula con la ecuación 14:

$$SCDR_1 = (1 - r_1^2) \sum_j y_{1j}^2 = (1 - 0.90472259) (43\ 232\ 969.24) = 4\ 119\ 125.49$$

Este último también se puede calcular con la diferencia de la [SCT₁– SCR₁].

Para resolver el cálculo en EXCEL, se escribe en la celda deseada:

$$= B31 - B34$$

La suma de éstas desviaciones ($SCDR_i$) es la Suma de Cuadrados de las Desviaciones Ponderadas. Ya concentrados los resultados anteriores en el Cuadro 11, se procede al siguiente paso.

7.2 Análisis de varianza partiendo de las regresiones.

En el Cuadro 11, el total de la columna de SCT_i (715417332) es la suma de cuadrados del residual del Cuadro 9, lo que viene siendo la variación total del experimento menos la suma de cuadrados de las variedades ($SCT - SCV$ calculados en los puntos 6.1 y 6.2). Esto es explicable por que la Suma de Cuadrados Total para cada variedad (SCT_i) del Cuadro 11 se calcularon para cada regresión, en otras palabras, es la variación *dentro* de cada una de las variedades y al sumarse se anula la variación *entre* variedades. Por lo que, la diferencia entre la suma de cuadrados total del experimento y el total de SCT_i es precisamente la variación atribuible a las variedades. De éste hecho se puede calcular la Suma de Cuadrados de Variedades (SCV_i).

7.2.1 Suma de Cuadrados Total

La suma de cuadrados total se calcula igual que en el punto 6.1. Por definición:

$$SCT = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = 730\ 488\ 463$$

Para su cálculo, en la celda deseada se escribe:

$$=DESVIA2(B3:Q23)$$

Lo anterior debido a que todos los datos están desde la celda B3 hasta la celda Q23, y de ellos se necesita la suma de sus desviaciones al cuadrado.

7.2.2 Suma de Cuadrados de Variedades.

La suma de cuadrados total calculada para una variedad, (SCT_i), es la variación *dentro* de genotipos y excluye la variación *entre* genotipos. Por lo que si se resta de la Suma de Cuadrados Total del experimento, calculada en el punto anterior, el total de SCT_i , ordenados en el Cuadro 11, se obtiene la Suma de Cuadrados de Variedades:
 $SCV = SCT - \text{total } (SCT_i) = 730\ 488\ 463 - 715\ 417\ 332 = 15\ 071\ 130$

También se puede calcular como en el punto 6.2.

7.2.3 Suma de Cuadrados de Ambientes (lineal).

$$SCA = v \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 = 16(40\ 534\ 288) = 648\ 548\ 600$$

Debe recordarse que el valor de la SCA y SCA(lin) son iguales, la única diferencia es que para el primero hay 20 grados de libertad ($a - 1$), 4 mientras que para la segunda hay uno solo grado de libertad. Para calcular la SCA en EXCEL, se obtiene mediante la función "DESVIA2":

$$=DESVIA2(R3:R23)*16$$

7.2.4 Suma de Cuadrados de Variedad por Ambiente (lineal).

$$SCVxA \text{ (lineal)} = \sum_i \left[\frac{(\sum_j Y_{ij} I_j)^2}{\sum_j I_j^2} \right] - SCA(\text{lin}) \quad 7)$$

De donde $\sum_i \left[\frac{(\sum_j Y_{ij} I_j)^2}{\sum_j I_j^2} \right]$ es el total de las SCR_i . Del Cuadro 11 se obtiene el total de las SCR_i y del punto anterior la $SCA(\text{lin})$:

$$SCVxA \text{ (lineal)} = 651\ 768\ 956 - 648\ 548\ 600 = 3\ 220\ 356$$

También se puede calcular con la siguiente fórmula:

$$SCVxA(\text{lin}) = \text{Total SCT}_i - SCA - \text{Total SCDR}_i$$

$$SCVxA = 715\ 417\ 332 - 648\ 548\ 603 - 63\ 648\ 372 = 3\ 220\ 357$$

7.2.5 Suma de Cuadrados de las Desviaciones Ponderadas. (SCDP)

La SCDP es el total de $SCDR_i$ del Cuadro 11 (63 648 376), que a su vez se descompone en las desviaciones de regresión de las variedades, calculadas ya en el punto 7.1.5. Lo siguiente se extrae, también, directamente del Cuadro 11.

$\sum_j d_{1j}^2 = 4\ 119\ 125.7$	$\sum_j d_{7j}^2 = 3\ 806\ 732.9$	$\sum_j d_{13j}^2 = 3\ 922\ 252.8$
$\sum_j d_{2j}^2 = 3\ 219\ 223$	$\sum_j d_{8j}^2 = 3\ 564\ 712.5$	$\sum_j d_{14j}^2 = 5\ 244\ 935.5$
$\sum_j d_{3j}^2 = 1\ 472\ 989.7$	$\sum_j d_{9j}^2 = 5\ 088\ 060.4$	$\sum_j d_{15j}^2 = 5\ 498\ 964$
$\sum_j d_{4j}^2 = 2\ 678\ 999$	$\sum_j d_{10j}^2 = 7\ 446\ 453.4$	$\sum_j d_{16j}^2 = 6\ 631\ 790.9$
$\sum_j d_{5j}^2 = 2\ 943\ 043.2$	$\sum_j d_{11j}^2 = 2\ 516\ 336$	
$\sum_j d_{6j}^2 = 2\ 908\ 689$	$\sum_j d_{12j}^2 = 2\ 586\ 067.7$	

CUCRA



BOLETA CENTRAL

7.2.6 Cuadrados medios y valores observados de "F".

Con las Sumas de Cuadrados así obtenidas, se aplica el análisis de varianza presentado en el Cuadro 6. En la inteligencia de que son los mismos valores, el cuadro de análisis de Varianza queda igual que el Cuadro 9.

En el Cuadro 11 se realiza, también, los valores de "F" para la regresión y para las desviaciones de regresión. Los cuadrados medios se obtienen dividiendo la Suma de Cuadrados entre los grados de libertad. Los grados de libertad para la regresión es el número de variables menos uno, para éste caso (2-1). Los grados de libertad para las desviaciones de regresión son iguales al número de pares menos dos, para éste caso (21-2). El cuadrado medio para la regresión es igual que la suma de cuadrados de la regresión, dado que existe solo un grado de libertad. El cuadrado medio para las desviaciones de regresión se obtiene dividiendo cada una de las sumas de cuadrados entre 19, cálculos ya realizados en el Cuadro 11.

Los valores de "F", para la regresión, se calcula dividiendo el cuadrado medio de la regresión entre el cuadrado medio de las desviaciones de regresión. Por ejemplo:

Varietad 1: $F = (39\ 113\ 844)/(216\ 796) = 180.4$

Varietad 2: $F = (29\ 920\ 902)/(169\ 433) = 176.6$

Varietad 3: $F = (44\ 452\ 540)/(77\ 526) = 573.4$

Varietad 4: $F = (36\ 655\ 044)/(141\ 000) = 260.0$

CUCBA



BIBLIOTECA CENTRAL

Los siguientes resultados se presentan en el Cuadro 11.

Los valores para "F" calculados para la regresión, si se acepta la hipótesis nula, tienen una probabilidad mucho menor de 0.01 para ocurrir (*ver Cuadro 11*). Por lo que se infiere que el modelo de regresión lineal sí está explicando lo observado.

Si se acepta la hipótesis nula, es decir, que no existe relación alguna entre los rendimientos y los índices ambientales, y que la aparente relación es debida al azar, ¿Cuál es la probabilidad de que un valor de "F" aleatorio fuese igual o mayor que el

observado?. La respuesta viene dada para cada variedad en el Cuadro 11 en la columna “P(F) (Reg)”, el cual para la variedad 1(F=180.4) es de 3.774×10^{-11} , probabilidad extremadamente baja, por lo que se rechaza la hipótesis nula y se infiere que existe una relación real entre los valores de rendimiento y los índices ambientales. La hipótesis nula se rechaza en las demás variedades también, el modelo lineal es el adecuado para explicar el fenómeno. Los valores de “F” observados para las desviaciones de regresión, se calculan dividiendo el cuadrado medio de las desviaciones de regresión entre el Cuadrado Medio del Error Ponderado calculado en el punto 6.7. Por ejemplo: Variedad 1:

$$F = (216\ 796) / (527\ 476) = 0.411$$

Variedad 2: $F = (169\ 433) / (527\ 476) = 0.411$

Variedad 3: $F = (77\ 526) / (527\ 476) = 0.411$

Variedad 4: $F = (141\ 000) / (527\ 476) = 0.411$

Los siguientes cálculos están en el Cuadro 11. Si se considera cierta la hipótesis nula, es decir que ambos cuadrados medios son estimaciones de la misma varianza poblacional, la probabilidad de que un valor aleatorio de “F” sea igual o mayor que el observado viene dado en la columna P(F)(DReg). Por ejemplo: para la variedad 1 el valor observado de F es de 0.41 y la probabilidad de que ocurra éste valor o uno mayor es de 0.988. Por lo anterior se acepta la hipótesis nula y se infiere que las desviaciones de regresión no son significativas, prueba redundante dado que ya se estableció que la relación entre las variables no es aleatoria.

Con los valores calculados se puede realizar un análisis de varianza con los mismos valores que el Cuadro 9, llegando por consiguiente a las mismas conclusiones. Para éste caso, resultó igual el cálculo utilizando el Cuadrado Medio del Error Ponderado que únicamente el análisis de varianza de la regresión.

7.2.7 Desviaciones de Regresión ($S^2_{d_i}$).

El parámetro de estabilidad $S^2_{d_i}$ se calcula igual que en el punto 6.9 con los datos se obtienen del Cuadro 11 de la siguiente manera:

$$S^2_{d_i} = \text{CM de las desviaciones de regresión} - \text{CM del error ponderado.}$$

Los cuadrados medios de las desviaciones de regresión para cada variedad están el Cuadro 11, y los cálculos están explicados en el punto 7.2.5. El cuadrado medio del Error Ponderado está calculado en el punto 6.7.

Ejemplificando para las primeras 4 variedades:

$$S^2_{d_1} = 216\,796.3 - 527\,476.32 = 0$$

$$S^2_{d_3} = 77\,526.0 - 527\,476.32 = 0$$

$$S^2_{d_2} = 169\,433.0 - 527\,476.32 = 0$$

$$S^2_{d_4} = 141\,000.2 - 527\,476.32 = 0$$

El parámetro de estabilidad $S^2_{d_i}$ fue estadísticamente igual a cero en todos los casos, lo que indica que las variedades son “consistentes”.

En las siguientes páginas se presentan las gráficas de dispersión para cada variedad que incluye la línea de tendencia, la ecuación de la línea y el coeficiente de determinación R^2 en las que se puede observar que los puntos están muy cerca de las líneas de regresión y las pendientes tienen valores muy cercanos a la unidad.

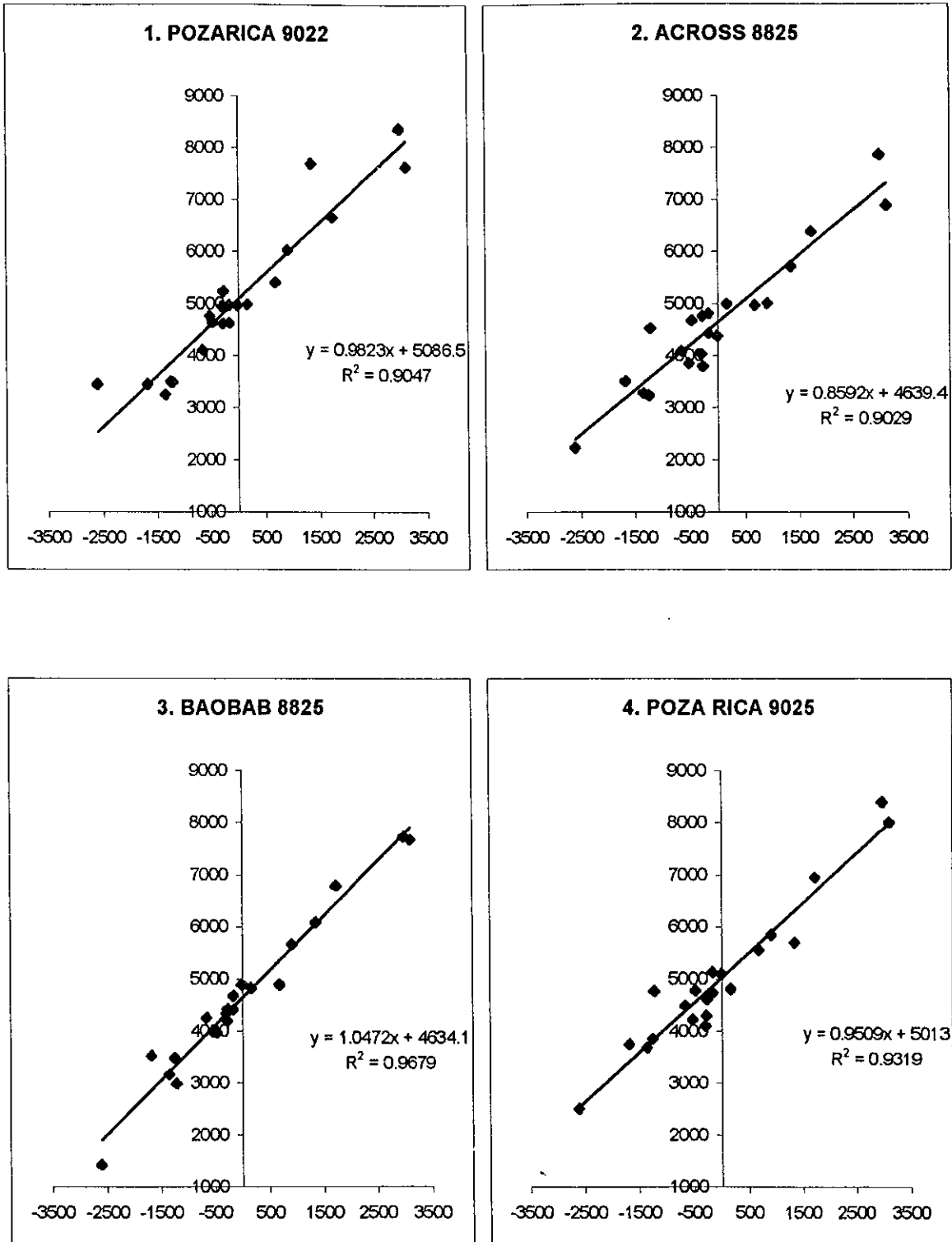


Figura 6. Diagramas de dispersión y líneas de regresión para el experimento.

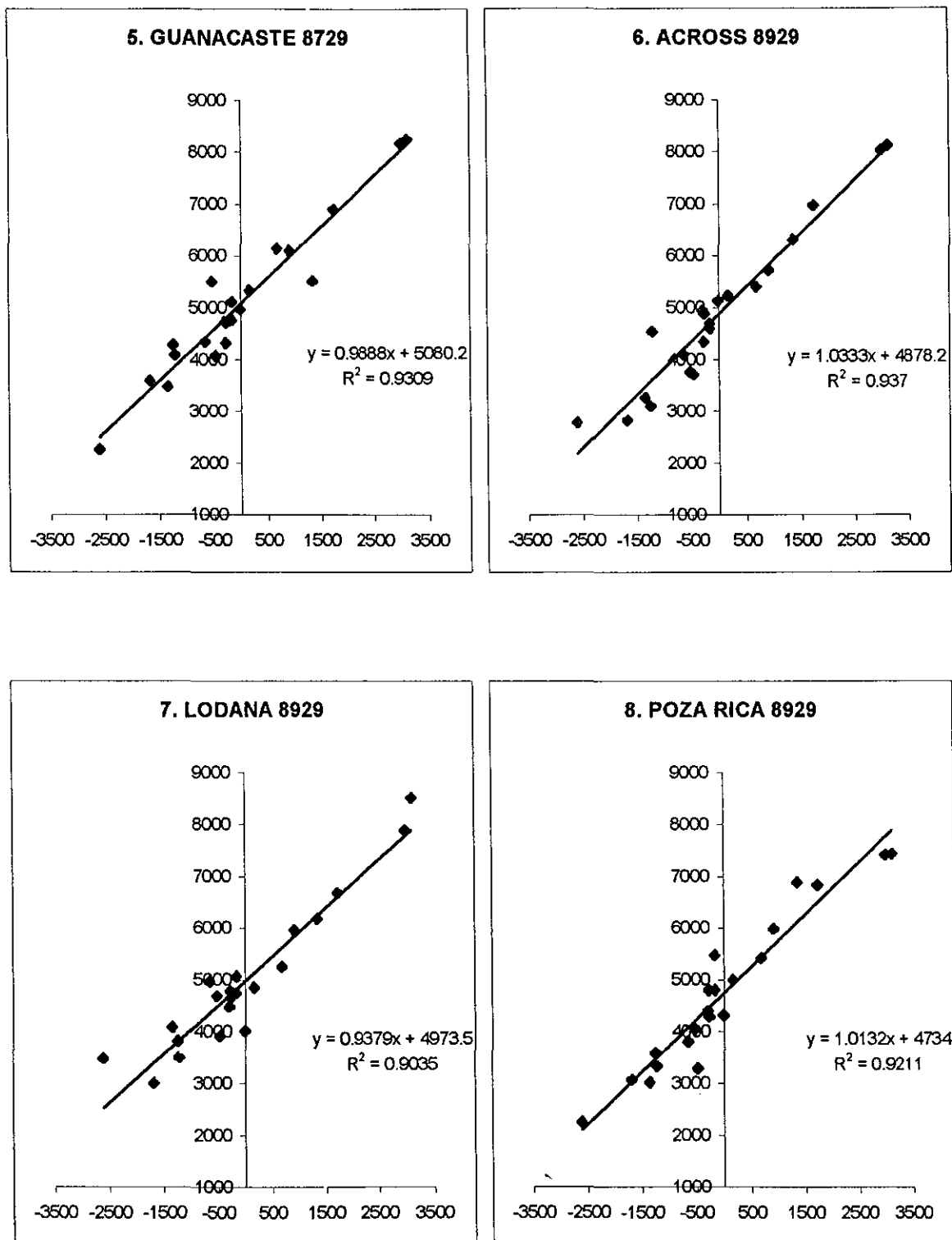


Figura 6. (Continuación)

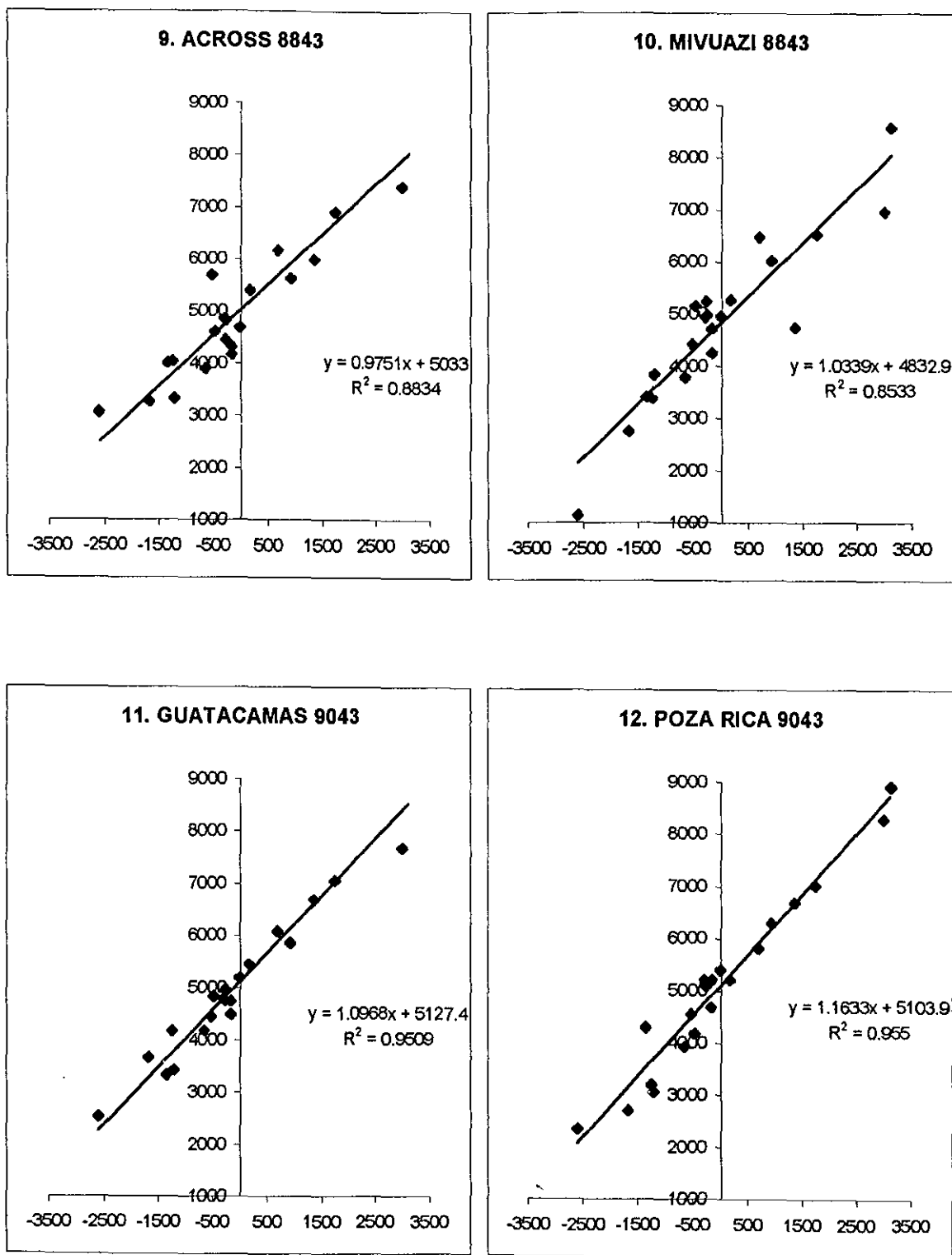


Figura 6. (Continuación)

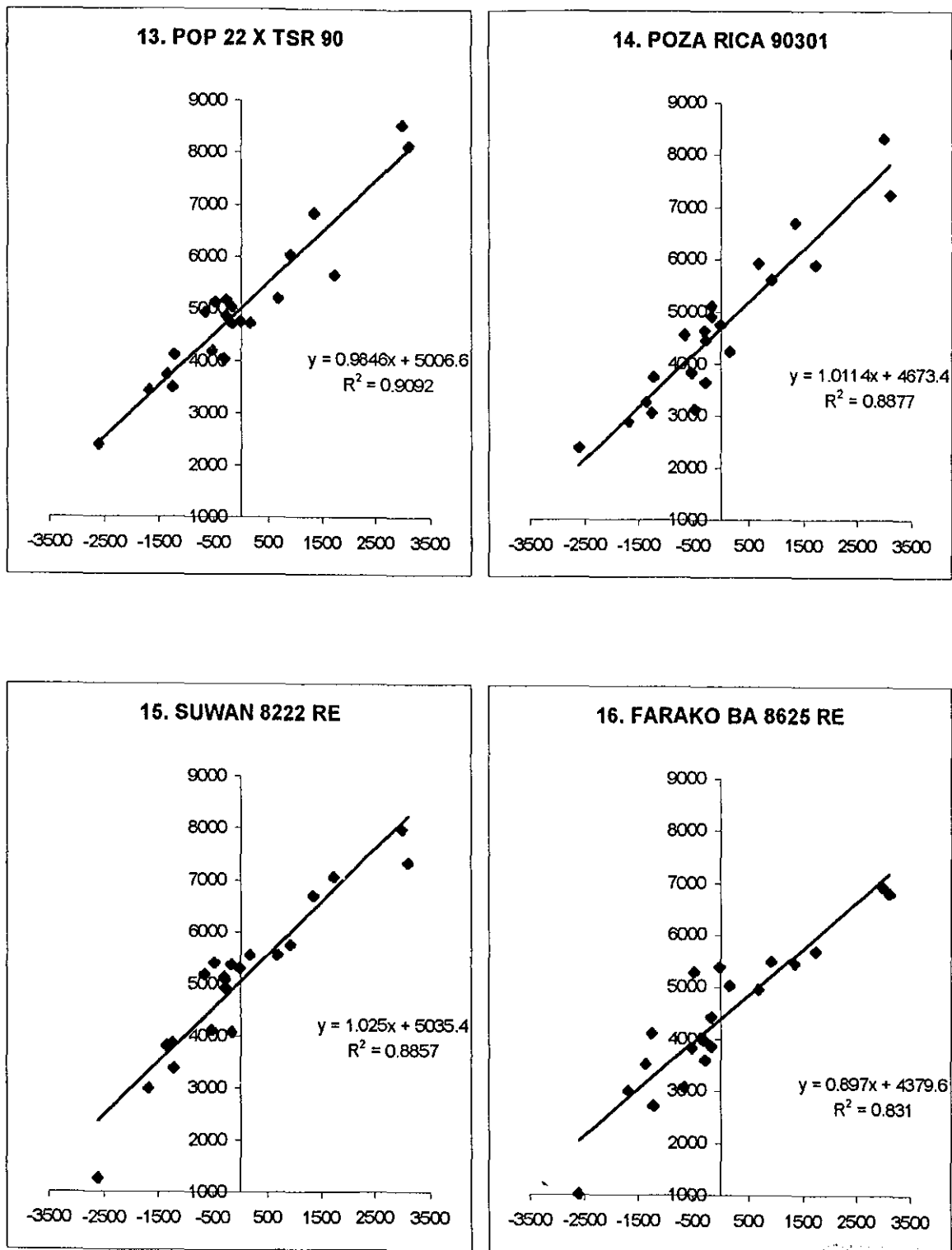


Figura 6. (Continuación)

8. RESULTADOS Y DISCUSIÓN DEL ESQUEMA PRESENTADO.

En el Cuadro 9 se observa que la interacción genotipo–ambiente no es significativa, por lo que el análisis de parámetros de estabilidad tiene poco caso realizarlo dado que todas las variedades presentarán coeficientes de regresión muy cercanos a la unidad, por lo que la elección de variedades se basará principalmente en las medias de variedades. Sin embargo, dado que el principal objetivo de la tesis es el ejemplificar dicho análisis, se procedió con el desarrollo del análisis de parámetros de estabilidad de Eberhart y Russell.

Si se efectúa una regresión de una serie de valores sobre sí mismos (que sean los mismos valores en el eje de las abscisas y de las ordenadas) se obtendría una serie de puntos perfectamente alineados sobre una recta con pendiente de igual a la unidad (formando un ángulo de 45 con respecto al eje de las abscisas), y unas desviaciones de regresión iguales a cero. Los parámetros de estabilidad para los datos serían: $b = 1$ y $S_d^2 = 0$. Derivado de lo anterior, cuando la regresión de una variedad sobre sus índices ambientales (o sobre la media ambiental) arroja los mismos resultados, $b = 1$ y $S_d^2 = 0$, se considera que no hay nada “extraño” en los resultados y que la relación entre las variables y los ambientes es directamente proporcional y lineal, o sea que la variación de los rendimientos de una variedad a través de ambientes, está bien explicada por el cambio de ambientes, es decir, la variedad está “accionando” con el ambiente pero no “interaccionando”. La línea que representa al *promedio* de todas las líneas de regresión de cada variedad del experimento, es la trazada con las medias ambientales en las ordenadas (eje Y) y los índices ambientales en las abscisas (eje X). Esta línea va a presentar también una pendiente $b = 1$ dado que prácticamente se está presentando el caso de auto-regresión, o sea la regresión es sobre ella misma, teniendo en cuenta que los índices ambientales se calculan con base en las estas mismas medias. Por lo anterior, el promedio de todos los coeficientes de regresión de las variedades sobre los índices ambientales, b_i , siempre es la unidad.

La media ambiental ($\bar{Y}_{.j}$) puede ser una buena calificación cuantitativa de un ambiente pero también puede también no serlo. Si el cultivo con el que se está realizando el experimento es el maíz, por ejemplo, la media de todas las variedades en un ambiente será una calificación del ambiente desde el punto de vista exclusivo del maíz. El hecho de que se reste la media general ($\bar{Y}_{..}$) de la media ambiental ($\bar{Y}_{.j}$), obedece a que el resultado así obtenido es el “efecto neto” que ejerce un ambiente “sobre” la media general. También puede obedecer a que los índices ambientales, así obtenidos son, en otras palabras, las desviaciones de las medias ambientales con respecto de la media general: $\bar{y}_{.j} = I_j = (\bar{Y}_{.j} - \bar{Y}_{..})$ y los cálculos, son más rápidos y sencillos. Si se hiciera una prueba similar a la anterior pero tomando como variable dependiente a las ganancias netas, en dinero, habría algunos casos paradójicos. Por ejemplo, si se probaran variedades criollas o sintéticas (que en experimentos conjuntos con muchas variedades presentaran coeficientes de regresión b cercanos a cero) en varios ambientes, es probable que los “mejores” ambientes, desde éste punto de vista, no sean los que produzcan más sino que los más baratos en propiedad, en renta o en tecnología sean los más rentables.

La línea que correspondería a las medias ambientales graficadas contra los índices ambientales, es una línea de 45° con respecto al eje de las abscisas con $b = 1$. Contra ella se comparan y trazan las demás líneas. Las que tengan valores para b mayores que la unidad es que forman un ángulo con las abscisas mayor de 45° y las variedades que presenten valores menores que la unidad para b , formarán ángulos menores de 45° (el coeficiente de regresión es precisamente la tangente del ángulo que forma la línea de regresión con el eje de las abscisas). Si el coeficiente de regresión, b_i , tiene un valor cercano a cero, quiere decir que la línea de regresión es paralela al eje de las abscisas.

Si de un experimento con varios genotipos se extraen dos variedades, V1 y V2, con valores de $b_1 = 1$ y $b_2 \approx 0$, y se hace el cálculo de los parámetros de estabilidad para

las dos variedades, las medias ambientales de los ambientes superiores bajarán con respecto del análisis inicial y las medias de ambientes más pobres se mantendrán. Lo anterior ocasiona que la línea de referencia, la que se traza con las medias ambientales y que siempre va a tener $b = 1$, aumente ambos valores de manera que $b_1 > 1$ y $b_2 > 0$, de hecho cuando solo se prueban 2 variedades la relación entre ellas es $b_1 = 2 - b_2$ (la suma de los valores de los coeficientes de regresión de las variedades b_i , es igual al número de genotipos probados, por que el valor promedio de todos los coeficientes b_i es, como ya se mencionó, la unidad).

Por éste motivo, cuando una variedad, que se probó con otras variedades, exhibe un coeficiente de regresión $b \approx 0$, de hecho, rindió un poco menos en ambientes buenos que en los adversos debido a la “inercia” de la mencionada “línea de equilibrio”.

Se puede inferir que el error que pueda traer el problema de la “auto-regresión” se disuelve al aumentar el número de variedades. La calificación cuantitativa de un ambiente, mediante el índice ambiental utilizado, aumenta en precisión al aumentar el número de variedades probadas.

De acuerdo a lo anterior, una variedad cuya línea de regresión sea paralela a la línea “promedio” o de equilibrio (cuya $b = 1$) y que las desviaciones de regresión no sean significativas, se considerará estable.

Observando el Cuadro 10, se tienen valores para los coeficientes de regresión (b), muy cercanos a la unidad. Las variedades 11 y 12 presentan valores de b poco mayores de la unidad por que rindieron mejor en ambientes favorables y su rendimiento bajó más que el promedio en ambientes menos favorables. Las medias de éstas dos variedades son las más altas, en especial la variedad 11, que las hace recomendables para ambientes potencialmente buenos (*ver figuras 6-11 Y 6-12*). Las variedades 2 y 16, por presentar los valores más bajos para b , presentan menos disminución de producción en ambientes desfavorables, pero no responden muy bien a buenos ambientes, además sus medias son de las menores.

La elección de la variedad adecuada depende de los propósitos, de los ambientes, de los parámetros de estabilidad y de las medias de las variedades.

Considerando que en el experimento analizado las desviaciones de regresión no fueron significativas, se puede decir que todas las variedades probadas son consistentes. Todas las variedades responden al ambiente pero no interactúan.

En el Capítulo 6 se explicó cada uno de los cálculos tal como lo propusieron los autores, y, en el Capítulo 7 se presenta la alternativa de obtener idénticos resultados partiendo de los análisis de varianza de las regresiones. Ésta manera ofrece mejor comprensión, igual precisión y menor probabilidad de un error.

CUCBA



BIBLIOTECA CENTRAL

9. CONCLUSIONES.

Los parámetros de estabilidad de Eberhart y Russell tienen una gran aplicabilidad en la experimentación agrícola, tanto para evaluación y mejoramiento de plantas como para determinar localidades idóneas para un programa de mejoramiento genético de cultivos. Básicamente, los parámetros de estabilidad de Eberhart y Russell auxilian en la descripción del comportamiento de un genotipo en diversos ambientes, y que tanto el modelo utilizado sirvió para explicar lo observado.

Debido a que el método realiza la regresión los rendimientos sobre un índice calculado a partir de las medias ambientales de los rendimientos, se cae en una contradicción estadística llamada “auto-regresión”. Los errores que la mencionada auto-regresión pueda traer consigo, se diluyen conforme se aumenta el número de ambientes y de variedades que se incluyan en el experimento.

El trabajo se desarrolló en forma progresiva, introduciendo al lector, de manera lógica, al análisis de parámetros de estabilidad de Eberhart y Russell tal como ellos lo propusieron originalmente. Para cada estadístico, se estableció una definición para su comprensión y una explicación detallada para su cálculo. Para cada cálculo se estableció el procedimiento respectivo en la hoja de cálculo de EXCEL.

El cálculo para el análisis de parámetros de estabilidad es más sencillo, rápido y con menos probabilidades de error, si se efectúa únicamente con los análisis de regresión de las variedades sobre su índice ambiental, pudiendo llegar a idénticas conclusiones que con el método descrito por Eberhart y Russell en su artículo original. Los cálculos requeridos son fácilmente realizables por programas de cómputo que son de muy fácil acceso, pero por no ser programas muy especializados, se requiere de los fundamentos del análisis, los cuales también se presentaron. Para una mejor interpretación del análisis de los parámetros de estabilidad, la comprensión matemática del procedimiento, es de gran ayuda en la toma de decisiones.

BIBLIOGRAFÍA

- Cochran W. G. y G.M. Cox, 1983, *Diseños experimentales*. Octava edición. Trillas, Mex.
- Cruz M., R. y A. Hernández J. 1994. *Análisis computacional de la interacción genotipo-ambiente con el modelo AMMI*. Rev. Fitotec. Méx. Vol. 17:103-115.
- Cruz M.,R. 1989. *Un ejemplo de la prueba exacta de los parámetros de estabilidad de Eberhart y Russell*. Rev Fitotec. Mex. 12:147-155.
- Eberhart, S.A. y W.A. Russell. 1966. *Stability parameters for comparing varieties*. Crop Sci. 6: 36-40.
- Finaly, K.W. y G.N. Wilkinson. 1963. *The analysis of adaptation in a plant breeding programme*. Aust. J. Agric. Res. 14:742-754
- Jalaluddin Md. And S.A. Harrison. 1993. *Repeatability of Stability Estimators for Grain Yield in Wheat*. Crop Sci. 33:720-725.
- Little T. y F. Jackson Hills, 1978, *Métodos estadísticos para la investigación en la agricultura*. Editorial Trillas México.
- Marquez S.F. 1974. *El problema de la interacción genético-ambiental en genotécnia vegetal*. PATENA A.C. Fitotécnia ENA. Chapingo México.
- Molina G., J. D. 1992. Respuesta lineal y no lineal de variedades a los efectos del ambiente y su importancia en la genotécnia del maíz. Memorias Simposio Interacción Genotipo ambiente U. de G., INIFAP. SOMEFI.
- Palomo. G., A. y R. Prado M. (1975) *Estimación de los parámetros de estabilidad y su aplicación en investigación agrícola con algodónero*. CIANE, INIA, SAG.
- Palomo. G.,A. 1974. *Interacción genotipo-ambiente y parámetros de estabilidad en variedades de algodónero (Gossypium hirsutum L.) para la Comarca Lagunera*. Tesis M.C. Chapingo, Méx. Colegio de Postgraduados.

- Pham H.N. and M.S. Kang. 1988. *Interrelationships among and Repeatability of Several Stability Statistics Estimates from International Maize Trials*. 1988. *Crop Sci.* 28:925-928.
- Piepho H.P. 1994. *Partitioning Genotype-Environmental Interaction in Regional Yield Trials Via a Generalized stability Variance*. *Crop Sci.* 34: 1682-1685.
- Shukla, G.K., 1972 *Some statistical aspects of partitioning genotype-environmental components of variability*. *Heredity* 29: 237-245
- Spiegel M.R., 1979, *Estadística*. serie de Schaum McGraw-Hill de México.
- Steel R.G.D. Y J.H. Torrie, 1960, *Principles and Procedures of Statistics With Special References to the Biological Sciences*. McGraw-Hill Book Co., Inc., New York.
- Cruz M.,R. 1989. *Análisis Estadístico de la Interacción genotipo-ambiente*. ITA No. 21 DGETAM SEP COSNET.
- Ayres, F. Jr. 1962. *Matrices*. Serie de Schaum. McGraw-Hill de México

CUCBA



BIBLIOTECA CENTRAL