IICA

INSTITUTO INTERAMERICANO DE COOPERACION PARA LA AGRICULTURA
OEA

METODOS ESTADISTICOS APLICADOS A LA INVESTIGACION AGRICOLA

El Instituto Interamericano de Cooperación para la Agricultura (IICA) es el organismo especializado en agricultura del Sistema Interamericano. Sus orígenes se remontan al 7 de octubre de 1942 cuando el Consejo Directivo de la Unión Panamericana aprobó la creación del Instituto Interamericano de Ciencias Agrícolas.

Fundado como una institución de investigación agronómica y de enseñanza de posgrado para los trópicos, el IICA, respondiendo a los cambios y las nuevas necesidades del Hemisferio, se convirtió progresivamente en un organismo de cooperación técnica y fortalecimiento institucional en el campo agropecuario. Estas transformaciones fueron reconocidas formalmente con la ratificación, el 8 de diciembre de 1980, de una nueva convención, la cual estableció como los fines del IICA los de estimular, promover y apoyar los lazos de cooperación entre sus 29 Estados Miembros para lograr el desarrollo agrícola y bienestar rural.

Con un mandato amplio y flexible y con una estructura que permite la participación directa de los Estados Miembros en la Junta Interamericana de Agricultura y en su Comité Ejecutivo, el IICA cuenta con una extendida presencia geográfica en todos los países miembros para responder a sus necesidades de cooperación técnica.

Los aportes de los Estados Miembros y las relaciones que el IICA mantiene con 12 Países Observadores, y con numerosos organismos internacionales, le permiten canalizar importantes recursos humanos y financieros en favor del desarrollo agrícola del Hemisferio.

El Plan de Mediano Plazo 1987-1991, documento normativo que señala las prioridades del Instituto, enfatiza acciones dirigidas a la reactivación del sector agropecuario como elemento central del crecimiento económico. En función de esto, el Instituto concede especial importancia al apoyo y promoción de acciones tendientes a la modernización tecnológica del agro y al fortalecimiento de los procesos de integración regional y subregional.

Para lograr esos objetivos el IICA concentra sus actividades en cinco áreas fundamentales que son: Análisis y Planificación de la Política Agraria; Generación y Transferencia de Tecnología; Organización y Administración para el Desarrollo Rural; Comercialización y Agroindustria; y Sanidad Vegetal y Salud Animal.

Estas áreas de acción expresan, de manera simultánea, las necesidades y prioridades fijadas por los mismos países miembros y los ámbitos de trabajo en los que el IICA concentra sus esfuerzos y su capacidad técnica, tanto desde el punto de vista de sus recursos humanos y financieros como de su relación con otros organismos internacionales.

Instituto Interamericano de Cooperación para la
Agricultura IICA

PROMECAFE

METODOS ESTADISTICOS APLICADOS A LA INVESTIGACION AGRICOLA

F. Omar Osorio
Ing. Agrónomo MG. SC.

Edgar Lionel Ibarra
Ing. Agrónomo M. S.
Especialista en Investigación
IICA-Honduras

Tegucigalpa, D. C. HONDURAS Septiembre, 1987

Segunda Edición auspiciada por: AID-ROCAP-COFEE PEST CONTROL PROJECT 590-0090

PUBL. MISC. 004 AI/HN 87 ISSN 0253-4746

PREFACIO

Los artículos sobre metodología estadística presentados en esta publicación, constituyen material de referencia de un Curso-Taller realizado por la Secretaría de Recursos Naturales de Honduras y el IICA a finales de 1982, dictados por los Ingenieros F. Omar Osorio y Edgar Lionel Ibarra; dirigido a personal profesional del Programa Nacional de Investigación Agrícola de la mencionada Secretaría de Estado.

Posteriormente, en abril de 1984, se realizó también en Honduras el Seminario sobre Métodos Estadísticos Aplicados a la Investigación Cafetera, auspiciado por el Instituto Hondureño del Café y el PROMECAFE. Gran parte del contenido de esta publicación fue utilizada en dicho seminario, el cual tuvo asistencia de técnicos de Investigación del IHCAFE y del Progrma de Roya de Guatemala; y fue dirigido por el Dr. Julio Henao (CATIE) y el Ingeniero Edgar Lionel Ibarra del IICA-Honduras.

Por su utilidad, como material de consulta para investigadores agrícolas, PROMECAFE realiza la publicación de esta 2da. edición revisada por los Doctores Gilberto Vegarano (PROMECAFE) y Pedro Oñoro (IICA), quienes han contado para ello con el permiso de los autores.

Tegucigalpa, D.C. Septiembre 1987

C O N T E N I D O

C O N T E N I D O Páginas

C O N T E N I D O

Páginas

1. CONCEPTOS FUNDAMENTALES SOBRE DISEÑO DE EXPERIMENTOS

F. Omar Osorio

1.1 Introducción al Diseño de Experimentos

Muchas definiciones se han dado al término experimento. Para nuestro propósito entenderemos por tal un ensayo físico planeado con el propósito de confirmar o rechazar los resultados de experiencias previas de tal forma que haga posible la toma de decisiones.

Cada experimento se planea con el propósito de proveer respuestas a una o más preguntas con este criterio, el investigador decide qué comparaciones de tratamiento produce la información más relevante al objeto de estudio. Se conduce entonces, un experimento, para probar hipótesis sobre diferencias entre tratamientos bajo condiciones comparables.

Para el estadístico el experimento es un conjunto de reglas usadas para inferir hacia la población bajo estudio. En este contexto entenderemos por diseño experimental la forma o procedimiento por el cual se asignan los tratamientos a las unidades experimentales.

1.2 Objetivos de un Experimento

Para diseñar un experimento se deben definir claramente los objetivos; que pueden tomar la forma de preguntas que deben de responderse o de efectos que deben estimarse.

Es deseable clasificar los objetivos en forma jerarquizada, especialmente cuando se acude al uso de diseños esperimentales que dan mayor precisión a algunas comparaciones que a otras.

El experimento debe suministrar datos que permitan dilucidar un problema o decidan un curso de acción; consecuentemente un buen diseño experimental solo resultará de una correcta apreciación del problema a ser resuelto; ningún refinamiento o técnica estadística puede sustituir este aspecto, a pesar de que lo que pueda decirse a favor de un correcto análisis de datos.

1.3 La Replicación y su Función

Cuando un tratamiento aparece más de una vez en un experimento, se dice que el tratamiento se ha replicado.

La replicación de los tratamientos cumple las siguientes funciones :

1) Proveer de un estimador del error experimental:
2) Mejorar la precisión de un experimento al reducir la desviación estándar del promedio del tratamiento;*
3) Incrementar el alcance de la inferencia de un experimento al seleccionar y usar el número apropiado de unidades experimentales.

La estimación del error experimental, es necesaria para efectuar las pruebas de hipótesis y para establecer los intervalos de confianza para los promedios.

Un experimento en el que los tratamientos aparecen una sola vez nos permite la estimación del error experimental y por lo

---

* La precisión, sensibilidad o cantidad de información se mide como el valor recíproco de la varianza de la media. Sea I=información; entonces $I=1/\sigma^2_{\bar{x}}$

tanto hace imposible observar la diferencia debida al tratamiento y la debida a la naturaleza de la unidad o parcela experimental. En otras palabras : si no se posee un estimador del error experimental, no hay forma de determinar si las diferencias observadas son debidas a los tratamientos o son una consecuencia de la variación inherente a la unidad experimental.

Cuando el número de repeticiones se incrementa, los estimadores de los promedios poblacionales se hacen más precisos.

En cierto tipo de experimentos, la replicación indica un alcance mayor de la inferencia estadística, la muestra poblacional es menos restringida por definición y la inferencia se verá favorecida. Supóngase que se desea determinar si existe diferencia real en el comportamiento de dos cultivares en una área donde hay principalmente dos tipos de suelo. Si el objetivo del experimento es realizar inferencias para ambos tipos de suelos, es obvio que ambos deben estar en el experimento. Es también importante que el área incluida dentro de cada repetición, o sea, la pareja de parcelas en que se planten los dos cultivares, sea de un solo tipo de suelo y lo más uniforme posible.

En algunos experimentos de campo el experimento se repite por varios años. La razón para ello es que las condiciones varían de un año a otro, lo cual hace importante conocer el efecto de los años sobre las diferencias entre tratamientos. Asimismo, se usan diferentes localidades para evaluar los tratamientos bajo diferentes condiciones ambientales en las cuales se utilizarán las recomendaciones producidas por la experimentación.

Las repeticiones en tiempo (años) y en el espacio (localidades) tienen por propósito incrementar el alcance de la inferencia.

En el caso de los cultivos perennes la repetición en el tiempo es una característica intrínsica de la experimentación en tanto que la repetición en el espacio es frecuente, el primer caso hace necesario una serie de consideraciones relativas al análisis de tales experimentos.

1.4 Control de Error

Por error experimental se entiende la no obtención de resultados idénticos en dos unidades experimentales tratadas idénticamente.

El control del error puede realizarse mediante :

1) El diseño experimental,
2) El uso de información concomitante,
3) Selección del tamaño y forma de la unidad experimental.

El uso del diseño experimental adecuado ha sido ampliamente investigado, aquí solo se señalan los aspectos más relevantes. El control del error experimental por medio del diseño consiste en diseñar el experimento de tal forma que la variación natural entre el conjunto de unidades experimentales sea manejada de tal manera que no contribuya a la diferencia entre medias de tratamientos.

Cuando se utiliza el Diseño de Bloques Completos al azar, el error experimental se basa en la variación entre unidades dentro de la repetición o bloque, puesto que la variación entre bloques puede ser estimada por un procedimiento aritmético. En este sentido puede avanzarse ampliado las características de cada diseño experimental.

En algunos experimentos la precisión puede incrementarse mediante el uso de información accesoria y de la técnicadenominada "Análisis de Covarianza". Este análisis se usa cuando la variación entre unidades experimentales es debida, en parte, a la variación en otros caracteres medibles que no son directamente controlados por el diseño experimental.

En la investigación agrícola son numerosos los casos en los que el análisis de covarianza y el uso de información accesoria puede ayudar a mejorar el control del error experimental. Una situación se da en los ensayos que se evalúa la efectividad de diversos fungicidas sobre una determinada enfermedad; en tales casos, generalmente, los resultados se han visto influenciados por el nivel de infección presente al momento de iniciar la aplicación de los tratamientos y los conteos de índices de infección deberían corregirse por la covariable "índice inicial de infección".

1.5 Tamaño y forma de las parcelas

Como regla general, puede decirse que parcelas o unidades experimentales grandes conducen a menor variación que las unidades pequeñas, usualmente el incremento del tamaño de la parcela resulta en un menor número de repeticiones. Generalmente es más fácil realizar una buena repetición de parcelas pequeñas que la correspondiente repetición para parcelas grandes.

La mayoría de los investigadores agrícolas recomiendan que las parcelas sean rectangulares; alargadas y estrechas en dirección a un gradiente observable en el terreno. Si no hay gradiente en ninguna dirección, no importa la forma de la parcela. Cochran, citados por Federer [2] presenta un ejemplo hipotético para mos-

trar esto, el caso incluye nueve tratamientos a comparar; se desea seleccionar una forma de parcela con la menor varianza del error experimental, en promedio, cuando la dirección del gradiente de fertilidad es desconocido.

A Plan (a) B

-3G	0	3 G
-3G	0	3 G
-3G	0	3 G

C D

A Plan (b) B

-4G -3G -2G -1G 0 1G 2G 2G 4G D

Suponiendo que en la realidad el gradiente de fertilidad es paralelo a AB, las parcelas en el Plan (b) están perpendiculares al gradiente. La suma de cuadrados entre las nueve parcelas sería $8\sigma^2 + 60\ G^2$. Si el gradiente de fertilidad estuviera paralelo a A C todas las parcelas estarían igualmente afectadas. En este caso la suma de cuadrados entre las nueve parcelas sería $8\sigma^2$.

Si σ^2 es la varianza aleatoria dentro de bloques, independientes de la forma de parcela, los cuadrados medios para los dos arreglos serían :

Gradiente de Fertilidad	G.L.	Cuadrados Medios Esperados Plan (a)	Plan (b)
Paralelo a AC	8	$\sigma^2 + \frac{54}{8} G^2$	σ^2
Paralelo a AB	8	$\sigma^2 + \frac{54}{8} G^2$	$\sigma^2 + \frac{60}{8} G^2$
Promedio	8	$\sigma^2 + \frac{54}{8} G^2$	$\sigma^2 + \frac{30}{8} G^2$

Si el valor de G es pequeño, la selección de la forma de la parcela es cuestión de preferencia. Cuando G es grande se debe seleccionar parcelas largas y estrechas. Ciertamente que el cultivo y sus labores culturales tienen mucho que ver en la selección de la forma de la parcela.

En relación a la forma del bloque y especialmente en ensayos con cultivos perennes es conveniente tener bloques compactos en vez de alargados.

En investigaciones realizadas en el país sobre la estimación de tamaño de parcela, solo se logró encontrar dos referencias, sobre arroz y maní o cacahuate, trabajos realizados por: J Posner y por M.A. Escoto y F.O. Osorio respectivamente (XXV Reunión Anual PCCMCA, Tegucigalpa, D.C., Honduras), es evidente que más investigación es necesaria sobre este aspecto.

En general, los tamaños de parcela que se usan en el país tienen diferentes orígenes: la adopción del tamaño de parcela que se usa en otros países, el criterio del investigador frente a diversos factores (costo, cultivo, naturaleza de los tratamientos), y los productos de investigaciones nacionales (muy poco).

Al no disponer de la investigación necesaria, los criterios más importantes a considerar en la selección del tamaño de parcela son:

a) Tipo de Cultivo. Cultivos anuales requieren tamaños de parcela menores que los cultivos perennes (frutales);
b) Tipo de investigación. La investigación de sistemas de riego y sistemas de cultivo requieren áreas mayores que los ensayos varietales.

c) Aspectos de tipo práctico. Estudios de pastoreo requieren parcelas muy grandes, generalmente definidas por el tamaño del potrero;

d) Costo. Al aumentar el tamaño de la parcela aumenta el costo del experimento. El investigador debe transigir entre información y costo, seleccionando el menor tamaño de parcela que le reporte el máximo de información.

1.6 Selección de Variables

Conviene definir detenidamente las variables que se evaluarán, si existen claros objetivos en un experimento, no se debe tener dificultad en definir claramente las variables de interés.

Un experimento es muy costoso para no obtener toda la información necesaria y posible del mismo. Algunos investigadores solamente toman en cuenta la variable rendimiento, argullendo que el fin último de cualquier investigación es incrementar rendimientos; aunque esta afirmación no siempre es válida debe tenerse en cuenta que existe una serie de variables que pueden ayudar a mejorar la interpretación de los resultados de un experimento; altura de plantas, vigor, fenología, índice de infección foliar de una enfermedad, índice de infección de nemátodos, etc.

Con alguna frecuencia, la selección inadecuada de los tratamientos imposibilita el correcto análisis estadístico de los resultados. Una situación de este tipo, se presenta por ejemplo cuando al evaluar el efecto de nitrógeno y fósforo sobre la producción de un cultivo, el experimentador selecciona los mismos niveles de ambos factores, ocasionando una perfecta correlación entre las dos variables independientes (en este caso nitrógeno y fósforo).

En tal situación el uso del análisis de regresión por el método de mínimos cuadrados es impracticable debiendo hacerse uso de algunas modificaciones que están fuera del alcance de este curso.

Asimismo, con frecuencia el espacio de exploración solo evalúa una región limitada (muy alto o muy bajo) del espacio de respuesta, concluyéndose que no hay respuesta del cultivo a la aplicación del factor considerado. Usualmente el espacio de exploración se ha restringido mucho; esto es particularmente inconveniente cuando se trata de ensayos exploratorios en los cuales, los limites deben ser razonablemente amplios.

1.7 Selección de Tratamientos

En cierto tipo de experimentos, los tratamientos tienen un efecto sustancial sobre la precisión, esto es especialmente cierto en el caso de los experimentos factoriales.
En algunos experimentos, las dosis de un factor son muy importantes, cuando se está midiendo el efecto de cierto nutriente sobre la respuesta de un cultivo, deberán incluirse diversos niveles de uso del nutriente a fin de determinar si la respuesta es de naturaleza lineal o curvilínea; en este caso el número de niveles y su espaciamiento es particularmente importante si se desea conocer las respuestas.

1.8 Técnica Experimental de Campo

Es importante usar una técnica cuidadosa en la conducción de un experimento. El investigador es responsable de que cada paso de la experimentación se ejecute con el mayor cuidado posible para asegurarse buenos resultados.

En general la variación resultante de la no adecuada atención del experimento, no es una variable aleatoria y no está sujeta a las

a las leyes de probabilidad sobre las que se basa la inferencia estadística. Esta variación puede denominarse deficiencia técnica en contraste con las variaciones aleatorias previamente mencionadas.

Asi, además de ser cuidadoso en la selección del diseño experimental, el investigador debe serlo también en la aplicación de los tratamientos y en las prácticas culturales a realizar en el experimento.

Una de las deficiencias técnicas más frecuentes se da en la falta de uniformidad entre las plantas que integran el experimento. Con frecuencia en un solo ensayo se encuentra una diversidad de tallos verticales por planta que genera modificaciones en las respuestas individuales de las plantas a los tratamientos, con lo cual se está introduciendo una fuente de variación no planeada en el experimento.

1.9 Aleatorización

El objeto de la aleatorización es asegurar la obtención de estimadores insesgados o válidos del error experimental, de las medias de tratamiento y de sus diferencias. La aleatorización es una de las características de los diseños experimentales modernos, introducida por R.A. Fisher [1]. La aleatorización generalmente involucra el uso de procedimientos tales como el lanzamiento de una moneda o el uso de una tabla de números aleatorios.

Para evitar el sesgo en la comparación entre medias de tratamiento, es necesario asegurarse que un tratamiento en particular no será consistentemente favorecido en las varias repeticiones por alguna fuente de variación conocida o no. En otras palabras, cada tratamiento debe tener la misma oportunidad de ser asignado a una parcela, sea ésta favorable o no, la aleatorización provee esta oportunidad. Cochran y Cox [1] establecen que "la aleatorización

es algo análogo a un seguro, en el sentido de que es una precaución contra disturbios que pueden incurrir o no y que pueden o no ser serios si ocurren".

La aleatorización tiende a eliminar la correlación entre errores que se presentan en ensayos sistemáticos, otrogando validez a las pruebas de hipótesis.

1.10 Submuestreo en parcelas Experimentales

Puede haber situaciones en las que en lugar de cosechar toda la parcela, sea más aconsejable cosechar solo una muestra de plantas. En estos casos lo importante no es la reducción en trabajos y costo, sino que existe la posibilidad de obtener la información valiosa al analizar los datos obtenidos en plantas individuales.

Ejemplo :

Con la finalidad de evaluar el contenido de cafeína en cuatro variedades de café, se realiza un ensayo utilizando un diseño en bloques completos al azar, con cuatro repeticiones. Cada parcela se muestrea tomando información de veinte plantas por parcela, el cuadro del análisis de varianza para el Modelo 1 (efectos fijos) sería el siguiente :

F. de V.	G.L.	C.M. Esperados
Bloques	3	$\sigma^2+20\sigma_e^2+\frac{80}{3}\sum\tau^2$
Variedades	3	$\sigma^2+20\sigma_e^2+\frac{80}{3}\sum\tau^2$
Error	9	$\sigma^2+20\sigma_e^2$
Error de Muestreo	304	σ^2
TOTAL CORREGIDO	319	

La prueba de hipótesis para variedades se hace utilizando el CM del error experimental, no el de el error de muestreo. El cálculo de los estimadores de σ^2 y σe^2 (s^2 y se^2 respectivamente) da información muy útil sobre la variación, de parcela a parcela y de la variación de planta a planta dentro de una parcela. Se espera que las muestras provean estimaciones insesgadas de dichos errores.

1.11 El Concepto de Significancia

El uso del diseño experimental adecuado permite reducir el error experimental, con ello se reduce la probabilidad de cometer error al hacer comparaciones entre tratamientos.

Supóngase que como resultado de la prueba de F en el análisis de varianza de un experimento, se encuentra que existe diferencia estadísticamente significativa entre tratamientos al nivel de $\alpha = 0.01$, esto significa que el investigador debe seleccionar entre dos alternativas.

a) Aceptar que la diferencia efectivamente se debe a los tratamientos (su probabilidad de cometer error es sólo de 1%);
b) Aceptar que la diferencia no se debe a los tratamientos, sino es producto de una coincidencia o error. La probabilidad de ocurrencia de esta coincidencia es solo del 1%.

Con lo anterior se quiere establecer que aunque el nivel de significancia se determina en forma objetiva (mediante un procedimiento aritmético), su interpretación tiene ciertos elementos de subjetividad.

En el caso expuesto, el investigador puede aceptar la segunda alternativa y considerar que aunque la probabilidad de ocurrencia de una coincidencia del tipo de la obtenida, es solo de 1%, en su caso lo que ocurrió fue precisamente esa coincidencia, por lo tanto no acepta la diferencia como producto de la naturaleza de los

tratamientos, obviamente para tomar esta decisión el investigador recurre al conocimiento que sobre la naturaleza del experimento (y obviamente de sus tratamientos) posee.

No es correcto lo que se observa con frecuencia, que aunque se acepte que efectivamente no existe diferencia estadísticamente significativa como producto del análisis de varianza, algunos investigadores, proceden a efectuar un análisis económico en el cual utilizan las diferencias entre tratamientos que previamente se ha aceptado no son significativas. Obviamente en tales casos corresponde, casi siempre, seleccionar el tratamiento que involucre el menor costo de producción.

Sin embargo la probabilidad de error nunca desaparece, recuérdese en principio que el diseño de experimentos se basa en el uso de modelos probablilísticos. El análisis de varianza permite someter a prueba hipótesis del tipo :

Ho: ti = t2..... tn = 0: Hipótesis nula

Hi: ti ≠ t2..... tn ≠ 0: Hipótesis alternativa

El procedimiento de fijar un nivel de significancia (α) a la decisión final es enteramente objetiva y, si decidimos rechazar Ho cuando ésta es verdadera y por lo tanto debe aceptarse, decimos que cometemos un error tipo I, por otro lado cuando aceptamos Ho y ésta es falsa estamos cometiendo error tipo II.

a= Probabilidad (error I)

B= Probabilidad (error II)

A la cantidad a se le denomina nivel de la prueba, en tanto

que a 1-B se le denomina potencia de la prueba.

Lo ideal sería tener el nivel pequeño y la potencia de la prueba grande, sin embargo, generalmente para aumentar la potencia se requiere aumentar el nivel de la prueba. El procedimiento clásico consiste en especificar el tamaño aceptable y buscar la mejor (más potente) prueba del tamaño seleccionado.

R E F E R E N C I A

1.- COCHRAN, W.G. and G.M. COX. Diseños experimentales. Editorial Trillas, México, 1965.

2.- FEDERER, W.T. Experimental design. Macmillan Co. N.Y. 1955.

3.- OSTLE, B. Estadística Aplicada. México, Limusa-Wiley, 1970.

4.- STEEL, R.G.D. and J.H. TORRIE. Principles and procedures of statistics, a biometrical approach. 2nd. edition, N.Y. Mac - Graw-Hillo 1980.

2. ANALISIS DE SERIES O GRUPOS DE EXPERIMENTOS.

Edgar Lionel Ibarra.

2.1. INTRODUCCION.

En los programas de investigación agrícola a nivel de una región del país, es común encontrar series o grupos de experimentos similares, repetidos tanto en espacio como a través del tiempo. Ello es así por diversas razones, dentro de las que sobresalen el interés de los investigadores de incrementar la confiabilidad de los resultados experimentales y el de realizar inferencias aplicables a toda una región o un periodo futuro de mediano plazo.

Con esa intención es bastante justificable la realización de series de experimentos similares, ya que en cuanto a confiabilidad, un análisis combinado provee una mayor precisión en las estimaciones y pruebas de hipótesis y se pueden aislar algunos efectos de medio ambiente (de localización y de tiempo) que interacionan con los efectos de los tratamientos bajo estudio. Por otra parte, las inferenicas sobre resultados de experimentos individuales tienen en la mayoría de casos, una aplicacion restringida a las particulares condiciones de lugar y tiempo en que estos fueron ejecutados.

Lo anterior tiene consecuencias importantes, particularmente en experimentos agrícolas, en los cuales los efectos ambientales tienen mayor influencia, en comparación con investigaciones de laboratorio, en las cuales el investigador puede ejercer un mayor control de factores exógenos.

La metodología de análisis e interpretación de grupos de experimentos similares, se basa en una extensión del procedimiento para casos Individuales. Por ello principiaremos con una revisión del modelo lineal para un experimento simple.

2.2 EL MODELO LINEAL DE UN EXPERIMENTO SIMPLE.

Para exponer este tema tomeremos el caso bastante común de un experimento bajo el diseño de bloques al azar. En el mismo, una observación experimental Y_{ij} de cualquier parcela, se asume que tiene la siguiente forma lineal.

$$Y_{ij} = \gamma + \tau_i + \beta_j + \varepsilon_{ij}$$

donde γ = media verdadera de la población

τ_i = efecto del tratamiento i aplicado en la parecela observada

β_j = Efecto del bloque j particular donde esté situada la parcela observada; y

ε_{ij} = Variación no explicada o error experimental en la parcela.

Además de lo anterior, constituyen supuestos básicos para el análisis da varianza, las siguientes condiciones:

a) Los efectos de tratamiento τ_i y de bloques β_j pueden ser fijos o aleatorios, dependiendo de consideraciones de diseño y objetivos del experimento. Si por ejemplo, los tratamientos son escogidos con un interés particular en determinados factores y nó consituyen una muestra aleatoria de todos los tratamientos posibles, entonces son fijos (caso contrario son aleatorios); y este es un caso muy generalizado en investigación agrícola. Lo mismo se puede decir de bloques.

En todas las observaciones experimentales :

$$\Sigma\tau_i = 0 \,, \qquad \Sigma\beta_j = 0$$

b) Se asume también que los errores ε_{ij} representan una variable aleatoria independiente, normalmente distribuida con media cero y varianza σ_e^2.

Para una apropiada interpretación de los resultados experimentales, debe establecerse claramente el modelo de análisis especificando si los efectos son fijos, aleatorios o si ocurren ambos tipos (modelo mixto). En el siguiente cuadro mostramos la estructura del análisis de varianza, indicando para cada componente el modelo de estimación o valor esperado del cuadrado medio; lo cual es importante para decidir sobre la relación de F de varianza adecuada para las pruebas de significancia.

Cuadro 1. Estructura del ANVA y valores esperados de sus componentes de varianza, Diseño en Bloques Completos al Azar.

FUENTE	G.L.	MODELO FIJO (τ_i, β_j)	MODELO ALEATORIO (τ_i, β_j)
Bloques	r-1	$\sigma_e^2 + \frac{t}{r-1}\Sigma\beta_i^2$	$\sigma_e^2 + t\sigma_b^2$
Tratamientos	t-1	$\sigma_e^2 + \frac{r}{t-1}\Sigma\tau_i^2$	$\sigma_e^2 + r\sigma_t^2$
Error	(r-1) (t-1)	σ_e^2	σ_e^2

FUENTE	G.L.	MODELO MIXTO
Bloques	r-1	$\sigma_e^2 + \frac{t}{r-1}\Sigma\beta_i^2$
Tratamientos	t-1	$\sigma_e^2 + r\,\sigma_t^2$
Error	(r-1)(t-1)	σ_e^2

n caso de efectos fijos, la prueba de hipótesis nula sobre tratamientos e plantea en términos de igualdad de efectos,
n tanto que si los efectos son aleatorios, la hipótesis nula se plantea

en términos de la varianza de efectos de tratamientos

$$H_0 : \sigma_t^2 = 0 \ , \ H_1 : \sigma_t^2 > 0$$

Examinando el Cuadro 1, se ve que la prueba de significancia adecuada para tratamiento, es la relación $F = \frac{\text{CM Tratamiento}}{\text{CM Error}}$ para cualquiera de los modelos presentados.

2.3 CASO DE EXPERIMENTOS REPETIDOS EN ESPACIO

Al tratar este caso (que también podría ser el de un solo experimento repetido en años), haremos algunas consideraciones de aplicación general en el análisis de series de experimentos.

2.3.1 Condiciones sobre uniformidad de los experimentos.

Al hablar en términos de un análisis combinado implicamos que ello trata del análisis de un grupo de experimentos similares en cuanto a su diseño y manejo, siendo las únicas diferencias el sitio distinto donde cada uno está localizado y el sorteo de tratamientos en cada lugar.

En cuanto a diseño se entenderá que hay igualdad en los tratamientos, número de bloques (o de repeticiones en cada lugar); tamaño y población de cada parcela experimental, misma forma de parcelas y manejo general de los experimentos. Sobre esto último es importante señalar que se han tomado similares medidas para reducir al máximo el error experimental y errores accidentales en cada experimento y que las variables de respuesta son medidas con procedimientos uniformes que garanticen la no inclusión de sesgos de estimación.

Las anteriores consideraciones, son altamente deseables para proceder a un análisis combinado de grupo de experimentos. Aunque pueden haber algunas diferenicas permisibles (y que ocurren en la práctica) tal es el caso de diferencias en el número de repeticiones, pues hay un procedimiento analítico para obviar esta dificultad, aunque ello puede incidir en diferencias en precisión y posiblemente en heterogenidad de variancias del error, que efecta la validez del análisis combinados como veremos más adelante.

Para cada localidad o sitio es necesario contar por lo menos con dos repeticiones o bloques; aunque pueden existir ensayos extensivos (tipo demostrativo) en fincas que utilizan un solo bloque, pero el análisis combinado de estos es muy impreciso.

También existe la posibilidad de ajustar por diferencias de población de parcelas entre lugares mediante un análisis de covarianza. Ello sin embargo complica demasiado e innecesariamente el análisis combinado, de tal manera que la mejor recomendación que se puede dar es que los experimentos repetidos en espacio, sean similares en todos sus especificaciones de diseño; incluyendo en ello un sorteo diferente de los tratamientos de cada localidad.

2.3.2 Consideraciones sobre los objetivos de los experimentos e inferencias sobre los resultados.

Si los objetivos de los experimentos dentro de una región dada consideran la obtención de recomendaciones agrónomicas aplicables a los sitios especificos de cada experimento, no tiene caso la realización de un análisis combinado. Sin embargo los investigadores no se limitan a este escaso aprovechamiento de la información experimental y más bien desean aumentar la confiabilidad para desarrollar recomendaciones aplicables a toda la región; a determinar la consistencia de ciertos efectos (tratamientos) en las variadas condiciones ambientales

dentro de la región y determinar la interacción de dichas condiciones con los tratamientos bajo experimentación.

Estas consideraciones repercuten en el modelo de análisis y la formulación de inferencias sobre los resultados, pudiéndose presentar entonces las siguientes variantes, según la estructura del análisis combinado.

Cuadro 2. Estructura de ANVA combinado, varias localidades, valores esperados de los componentes de varianza, diseño en Bloques completos al azar.

FUENTE	G.L.	MODELO FIJO	MODELO MIXTO (Trat.fijo.Loc.aleatorio)
Localidades o experimentos (l)	L-1	----------	$\sigma_e^2 + (r/(t-1))\sigma_{tL}^2 + rt\sigma_l^2$
Repeticiones en localidades (R)	L(r-1)	----------	----------
Tratamientos (T)	t-1	$\sigma_e^2 + rL/t-1 \sum \tau_i^2$	$\sigma_e^2 + t\,\sigma_{tL}^2 + (rL/(t-1))\sum \tau_i^2$
LxT	(L-1)(t-1)	$\sigma_e^2 + (r/(t-1)(L-1))\sum \lambda^2$	$\sigma_e^2 + t\sigma_{tL}^2$
Error Combinado	L(r-1)(t-1)	σ_e^2	σ_e^2

La situación del modelo fijo, aunque muy común en la realidad puesto que se supone que las localidades y los tratamientos no fueron determinados al azar como una muestra aleatoria de una población finita de lugares y tratamientos posibles; supone la dificultad teórica de que las inferencias que se hagan han de suscribirse a los lugares específicos y no a la generalidad de la región bajo estudio (tal cmo se supone en el caso de realizarse un análisis individual de los experimentos). Por lo anterior

y aunque forzando un poquito la situación, se considera que un modelo mixto en que se supone los efectos de tratamientos fijos (que es lo usual) y los de localidades al azar, es un modelo apropiado para aprovechar mejor la información experimental; es decir derivar inferencias aplicables a toda la región bajo estudio.

2.3.3 Análisis Combinado

Para ilustrar la metodología de un análisis basado en la estructura indicada en el Cuadro 2, vamos a utilizar un ejemplo de un ensayo hipotético de tres variedades de arroz, repetido en 3 localidades del Departamento de Olancho. El Cuadro 3 muestra los resultados de rendimiento observados en cada parcela experimental. Diseño Bloques al azar con tres repeticiones.

Cuadro 3. Resultados de rendimiento (TM/Ha) de tres variedades de arroz en las localidades que se indican. 1981

Localidades y repetición	VARIEDAD 4440	CICA-9	Blue Bonnet-50	Totales
Concepción				
I	4.40	3.83	3.60	11.83
11	5.60	5.00	4.90	15.50
111	4.10	4.50	4.10	12.70
Sub-total	14.10	13.33	12.60	40.03 ($\Sigma X_{..1}$)
La Bomba				
1	4.61	4.20	3.59	12.40
11	5.20	5.25	4.50	14.95
111	5.30	4.75	5.60	15.65
Sub-total	15.11	14.20	13.69	43.00 ($\Sigma X_{..2}$)
Las Delicias 1	9.10	7.41	4.00	20.51
11	6.99	7.01	4.92	18.92
111	7.15	6.18	3.72	17.05
Sub-Total	23.24	20.60	12.64	56.48 ($\Sigma X_{..3}$)
TOTAL ($\Sigma X_{i..}$)	52.45	48.13	38.93	139.51

	ΣX^2_{ijk}	CF	SCTOT.	Reps en L.
Concepción	181.2289	178.0445	3.1844	2.4518
La Bomba	208.6452	205.4444	3.2008	1.9506
Las Delicias	379.0780	354.4434	24.6346	1.9996
	768.9521			178.4196
				205.7894
				374.7424
				758.9514

Cuadro 4. Análisis de varianza individuales

		Concepción		La Bomba		Las Delicias	
FUENTE	G.L.	SC	CM	SC	CM	SC	CM
Bloques	2	2.4518		1.9506		1.9996	
variedades	2	0.3751	0.1876	0.3450	0.1725	20.2990	10.1495*
Error	4	0.3575	0.0894	0.9052	0.2263	2.3360	0.5840
Total	8	3.1844		3.2008		24.6346	

Cuadro 5. Análisis Combinado

FUENTE	G.L.	SC	CM	"F"
Localidades (L)	2	17.0791	8.5396	
Repetición en localidades	6	6.4020	1.0670	
Variedades (v)	2	10.5961	5.2980	$CM_v/CM_{vL} = 2.03$
V x L	4	10.4229	2.6057	$CM_{VL}/Cme = 8.69$ *
Error Combinado	12	3.5987	0.2999	
TOTAL	26	48.0988	------	

ANALISIS COMBINADO

Sumas de cuadrados (componentes "cruzados").

TOTALES $SC_{Tot} = \Sigma X^2_{ijk} - FC_{G.}$

$$SC_{Tot} = 768.9521 - \frac{(139.51)^2}{27}$$

$$= 768.9521 - 720.8533 = 48.0988$$

Variedades

$$SC_v = \frac{1}{rL} \Sigma X^2_{i..} - FC_G$$

$$SC_v = \frac{1}{9} (54.25^2 + 48.13^2 + 38.93^2) - 720.8533$$

$$= 731.4494 - 7208533 = 10.5961$$

Localidades

$$SC_L = \frac{1}{rv} \Sigma X^2_{..k} - FC_G$$

$$= \frac{1}{9} (40,03^2 + 43.00^2 + 56.48^2) - 720.8533$$

$$= 737.9324-720.8533 = 17.0791$$

Variedades x Localidades

$$SC_{VL} = \frac{1}{r} \Sigma X^2_{i.k} - FC_G - SC_v - SC_L$$

$$= \frac{1}{3} (14.10^2 + \ldots + 12.64^2) - 720.8533-10.5961-17.0791$$

$$= 758.9514-720.8533-10.5961-17.0791$$

$$= 10.4229$$

Componentes "anidadas"

SC bloques en localidades

$SC_{(r)}$ = SCR en los análisis individuales
= 2.4518 + 19506 + 19996 = 6.4020

SC error combinado

$SC_{(e)}$ = SC_e en los análisis individuales
= 0.3575 + 0.9052 + 3.3360 = 3.5987

Pruebas de significancia e Interpretación

Según los valores esperados de componentes de varianza, en el modelo mixto asumido, presentado en el Cuadro 5, se puede decir que la prueba de significación para la interacción VXL se prueba contra el error combinado. En este caso.

$$"F" = \frac{\sigma_e^2 + T\sigma^2 tL}{\sigma_e^2} = \frac{CM\ VXL}{CME} = \frac{2.6057}{0.2999}$$

"F" = 8.69

La cual es significativa aún al nivel P= 0.25 sugerido por Bancroft (1), para una prueba preliminar.

En cuanto a rendimiento, las tres variedades no responden igual en las localidades; ó por lo menos hay una de ellas que no responde igual en las tres localidades.

Se demuestra así una interacción del ambiente (localidades), manifestada en una respuesta diferencial de las variedades según la localidad. En efecto, ello se demuestra también en los análisis individuales en los que aparece una respuesta significativa de variedades en una sola localidad. En las Delicias hay diferencias significativa en tanto que no la

hay en las otras dos localidades, entre el rendimiento varietal.

La prueba de significancia para el efecto de variedades, se hizo según el Cuadro 5, calculando una "F" así :

$$F = \frac{\sigma_e^2 + t\sigma^2 tL + rL/t-1 \Sigma \tau_i^2}{\sigma_e^2 + t\sigma^2 tL} = \frac{5.2980}{2.6057} = 2.03$$

No significativa.

Ello indica, según el modelo, que cualquiera que sea la localidad dentro de la región bajo estudio, que no hay diferencia en rendimiento entre las tres variedades estudiadas.

Hay que considerar desde luego que en este ejemplo, para la facilidad de cómputo se utilizaron pocas repeticiones y localidades, lo cual incide en una baja precisión. Sin embargo si ese fuese el caso real, el resultado indica que para la región se puede inferir igualdad de respuesta para las variedades.

Pero hay que considerar también si se cumple otra condición del análisis combinado, la cual supone homogenidad de varianzas del error dentro de la región. Dichas varianzas estimadas a través de los errores experimentales individuales exhiben aparentes diferencias como se puede apreciar en el Cuadro 4.

Una prueba de homogenidad es la desarrollada por Bartlett (2) definida mediante la distribución de χ^2 por :

$$\chi^2 = 1 \{f_t \log_e s_p^2 - \Sigma f_i \log_e S_i^2\} \text{ para}$$

K-1 grados de libertad.

donde:

K = Número de varianza del error comparadas

C = Factor de correción

$$C = 1 + \frac{1}{3(K-1)} \left\{ \frac{1}{F_i} - \frac{1}{F_t} \right\}$$

F_i = Grados de libertad asociadas a cada varianza del error.

F_t = Total de grados de liberta del conjunto de varianza del error.

S_i^2 = Varianzas del error (cuadrados medios) individuales.

S_p^2 = Error Combinado

Así tendríamos

Localidad	Grados de Libertad (F_i)	Cuadrado (S_i^2) medio error.	$\text{Log}_e S_i$
Concepcion	4	.0894	-2.4146
La bomba	4	.2263	-1.4859
Las Delicias	4	.5840	-0.55379
TOTAL	12 (F_t)		-4.4384

$$\chi^2 = \frac{1}{C} \left\{ 12 \ (\log_e .2999) - (-17.4384) \right\}$$

$$= \frac{1}{C} (-14.4517 + 17.4384) = \frac{1}{C} (2.9867)$$

$$C = 1 + \frac{1}{3(3-1)} \left(3 \frac{(1)}{4} - \frac{1}{12} \right)$$

$$= 1 + \frac{1}{6} \left(\frac{3}{4} - \frac{1}{12} \right) = 1 + \frac{1}{6} \frac{2}{3} = 1.11$$

$$\chi^2 = 2.9867/1.11 = 3.42$$

Este valor de X^2 para 2 grados libertad indica que la probabilidad de un valor de X^2 mayor que 3.42 está entre 0.20 y 0.30.

$X^2 = 3.42$ no es significativa; los datos entonces indican que se acepta la hipótesis de que las varianzas observadas (del error) provienen de una misma población, es decir que hay evidencia de homogenidad de varianzas del error y por lo tanto se cumple una de las condiciones del análisis combinado.

Otras condiciones de validez son las que también se refieren a los análisis individuales de varianza, tales como :

a) Independencia de varianzas del error y medias de tratamientos.
b) Aditividad de los efectos en el modelo lineal.

Estas condiciones, en caso de no presentarse, pueden ser satisfechas mediante trasformación de los datos experimentales.

2.4 CASO DE EXPERIMENTOS REPETIDOS EN ESPACIO Y TIEMPO

Este caso es muy frecuente y para su análisis son aplicables las condiciones de uniformidad de diseño y homogenidad de varianza ya antes presentadas. Ocurre a menudo en la investigación con plantas anuales en que experimentos similares se ubican en varias localidades o estaciones de una región y a su vez se repiten igualmente durante un período de varios años. También ocurre el caso en experimentos con plantas perennes ubicadas en varias localidades y las respuestas se registran anualmente. Estas modalidades imprimen variantes en el análisis estadístico; sin embargo el procedimiento de cálculo es solo una extensión del método empleado en el caso de experimentos repetidos en espacio.

Los efectos del medio ambiente en este caso ya no constituyen los efectos de sitios diferentes, sino también los de las condiciones ambientales (particularmente climáticas) de los ciclos anuales que duran los expe

rimentos y sus respectivas interacciones con las respuestas de los tratamientos.

2.4.1 Experimentos con Cultivos Anuales.

Lo más frecuente es encontrar dos modalidades :

a) Los experimentos se sitúan en varias localidades y allí mismo, se continúan en años sucesivos (aunque cambiando el sorteo de tratamiento en cada uno, que es lo apropiado); y

b) Los experimentos se sitúan en las localidades pero en los años sucesivos estos se repiten en diferentes localidades dentro de la región.

La diferencia de análisis en ambos casos, será que en el primero de ellos el efecto de localidades será identificable (sera un efecto "cruzado") a lo largo de todos los años; y el mismo se podrá medir a través de los grandes totales (contrastes que se comparan) de cada localidad sobre todo el período de años que dure la experimentación.

En la segunda modalidad, este efecto no se puede determinar a lo largo de todo el período porque las localidades son distintas cada año. En cambio si se podrá medir para cada año en particular (es un efecto "anidado") y en el análisis global se sumarán los efectos parciales por año. Es decir, la fuente de variación a incluirse en el análisis de varianza será: Localidades en años.

Con estas consideraciones preliminares, entonces podríamos ilustrar el análisis de experimentos repetidos en espacio y tiempo con la primera modalidad.

Para ese caso especial y por cierto muy frecuente, la estructura del análisis de varianza tendrá la estructura que se presenta en el Cuadro 6, donde también se exhiben los valores esperados de los cuadra-

dos medios, para dos modelos lineales (fijos y mixto) de los efectos de tratamientos, repeticiones, años y localidades.

Cuadro 6. Estructura del análisis de varianza de series de experimentos repetidos en espacio y tiempo.

FUENTE DE VARIANZA	G.L.	VALORES ESPERADOS DE LOS CUADRADOS MEDIOS	
		MODELO FIJO	MODELO MIXTO ($_i$Fijo)
Años (a)	a-1		
Localidades (L)	L-1		
LxA	(a-1)(L-1)		
Repeticione en localidades y años	aL(r-1)		
Tratamientos (T)	t-1	σ_e^2 + rLa(T)*	$\sigma_e^2+\sigma_{tLa}^2+ra\sigma_{tL}^2+rL\sigma_{ta}^2+La(T)$
TxA	(t-1)(a-1)	σ_e^2 + rL(TA)	$\sigma_e^2+\sigma_{tLa}^2+rL\sigma_{ta}^2$
TxL	(t-1)(L-1)	σ_e^2 + ra (TL)	$\sigma_e^2+ r\sigma_{tL}^2$
TxAXL	(t-1)(a-1)(L-1)	σ_e^2 + (TLA)	$\sigma_e^2+ r\sigma_{tLa}^2$
Error Combinado	aL(r-1)(t-1)	σ_e^2	σ_e^2

*Los simbolos de los efectos fijos aquí presentados se han simplificado, ejemplo (T) represente $(L/t-1)\Sigma\tau^2$

Los modelos para cuadrados medios esperados que se indican en el cuadro 6 tienen importantes implicaciones en la interpretación y las inferencias derivadas del análisis estadístico. En el caso del modelo fijo, donde se asume que las repeticiones, tratamientos, localidades y años tienen efectos constantes y que la inclusión de estos en los experimentos fue porque es lo que tiene especificado y no interesa investigar otras alternativas, las inferencias se limitan a los años, localidades y tratamientos particulares que

que se investigan. Es decir que los resultados no se pueden extender a toda la región ni a un período (futuro o pasado) a través del tiempo.

Aunque la interpretación de estos modelos no es siempre clara, el concepto del modelo fijo es que él efecto es único (de una localidad, tratamiento o año particular). Aunque es desconocido, es estimable y dicha estimación es aplicable entonces a las particulares condiciones de la observación experimental, las cuales constituyen el reducido universo del investigador.

Por otra parte, el modelo aleatorio interpreta que los efectos de localidades , años,tratamientos y repeticiones, son variables y que el experimento lo que provee es una estimación de la varianza de dichos efectos, considerado éste experimento como una muestra aleatoria de universos más amplios de espacio (localidades), tiempo (años) y tratamientos posibles, Por lo tanto las inferencias pueden extenderse a dichos universos, en el sentido más interesante desde el punto de vista agrónomico.

Lo que sucede en la realidad de la experimentación agrícola es que generalmente los tratamientos constituyen una variable con efectos fijos, particularmente en los trabajos con dosificaciones de productos químicos y fertilizantes, en que estos son escogidos por un interés particular y con antecedentes de experimentación previa y por lo tanto están muy lejos de considerarase una muestra de una población finita de tratamientos.
Sin embargo, es muy razonable considerar que a través del tiempo, los ciclos anuales tienen un efecto aleatorio, Un período de tres años por ejemplo, puede considerarse una muestra de un período de tiempo mayor, a pesar de ciertas fluctuaciones climáticas que ocurren por ciclos. Así también, él efecto de localidades puede considerarse aleatorio, cuando se intalan varios experimentos dentro de un área o región geografica, aunque la

escogencia de las mismas no ocurra al azar debido a razones de facilidad y otras conveniencias para el investigador.

Por lo tanto, un modelo mixto, con efectos fijos para tratamientos y aleatorios para tiempo y espacio, como el que se presenta en el cuadro 6, es el que resulta recomendable a la gran mayoría de situaciones del Programa Nacional de Investigación Agrícola.

2.4.2 Pruebas de Significancia e Interpretaciones

Las pruebas de significancia para : años, localidades y la interacción LXA, pueden efectuarse calculando una relación de varianza "F" contra repeticiones en localidades y años, en el modelo fijo; y así también aunque en forma aproximada en el modelo mixto. En ese sentido el esquema de este análisis tiene cierta semejanza con el diseño de "parcelas divididas" en el cual el error "b" para pruebas de tratamientos principales juega el mismo papel de "repeticiones en localidades y años" de éste análisis global.

Los tratamientos (T) y las interacciones TXA, TXL y TXAXL se prueban calculando "F" contre el error combinado, cuando se trata del modelo fijo, como podra deducirse al observar los cuadrados medios esperados en el cuadro 6.

En el modelo mixto, que es el más aplicable en los trabajos actuales de investigación del PNIA, la situación es diferente. En este caso la interacción TXAXL se prueba calculando "F" contra el error combinado; y a su vez, las interacciones TXA y TXL se prueban contra la interacción TXAXL :

$$"F"_{TxA} = \frac{\sigma_e^2 + r\sigma^2_{tLa} + rL\sigma^2_{ta}}{\sigma_e^2 + r\sigma^2_{tLa}}$$

$$"F"\ TxAXL = \frac{\sigma_e^2 + r\sigma^2_{tLa} + ra\sigma^2_{tL}}{\sigma_e^2 + r\sigma^2_{tLa}}$$

Por otra parte, de la observación del modelo mixto (cuadro 6), se deduce que no hay divisor apropiado para la prueba de "F" de tratamientos, cuyo cuadrado medio esperado es:

$$\sigma_e^2 + r\sigma_{tLa}^2 + ra\sigma_{tL}^2 + rL\sigma_{ta}^2 + rLa\Sigma\tau_i^2$$

donde el término que interesa aislar es $rLa\Sigma\tau_i^2$ y para lo cual se necesita un divisor :

$\sigma_e^2 + r\sigma_{tLa}^2 + ra\sigma_{tL}^2 + rL\sigma_{ta}^2$, que no lo hay según el cuadro 6.

Solo en el evento de que las interacciones TxL y TxA no hubiesen sido significativas, puede probarse contra la interacción TxAxL en una forma aproximada.

Ante este probelma, Cochran y Cox (2) desarrollaron una prueba para tratamientos también aproximada, donde se calcula "F" así :

$$\text{"F"} = \frac{CM_T + CM_{TxAxL}}{CM_{TxL} + CM_{TxA}}$$

, donde las iniciales CM indican los respectivos cuadros medios

Esta "F" calculada, se comapra, con los valores de "F" en las tablas al nivel de significancia deseado, con los grados de libertad dados por las siguientes expresiones desarrolladas por Satterthwaite (3) :

$$F_1' = \frac{(CM_T + CM_{TxAxL})^2}{\dfrac{(CM_T)^2}{f_t} + \dfrac{(CM_{TxAxL})^2}{f_{tLa}}}$$

$$F_2' = \frac{(CM_{TxL} + CM_{TxA})^2}{\dfrac{(CM_{TxL})^2}{f_{tL}} + \dfrac{(CM_{TxA})^2}{f_{ta}}}$$

donde f_t, f_{tLa}, f_{ta} son los correspondientes grados de libertad

de los cuadrados medios utilizados en estas dos expresiones, según el Cuadro 6.

Interpretación de las interacciones :

A continuación se presenta una serie de interpretaciones al alcanzarse niveles de significancia estadística en las siguientes comparaciones :

1) Localidades x Años :

Las diferencias entre tratamientos no están involucradas en esta Interacción, por lo tanto se interpreta que cualquiera que sean los tratamientos, las respuestas globales (por ejemplo el rendimiento promedio de un cultivo anual) fueron mejores en algunas localidades en ciertos años, en comparación con la respuesta de las mismas localidades en otros años

2) Tratamientos por Años :

Algunos tratamientos, en toda la región cualquiera que sea la localidad, respondieron mejor en ciertos años que en otros.
Es decir que hubo cambio de posición o "ranking" de los tratamientos al compararse la respuesta de estos en varios años. Esto dificultará hacer recomendaciones de ciertos tratamientos para largos períodos de años; excepto en el caso de que el CM de tratamiento haya sido significativamente superior al CM de TXA, con lo cual si se podrán recomendar algunos de los mejores tratamientos.

3) Tratamientos x Localidades

Durante todo el período de años, algunos tratamientos tuvieron mejor respuesta que otros, pero en ciertas localidades la posición o "ranking" de los tratamientos varió con respecto a otras localidades.

Cuando el CM de tratamientos excede significativamente el CM TXL, ello evidencia que la respuesta general de los tratamientos fue suficientemente consistente para demostrar que algunos de estos tratamientos fueron superiores en todas las localidades, cualquiera que haya sido el año de prueba; y lo cual es importante para producir una recomendación generalizada.

Por el contrario, si CM_{τ} /TXL no es significativa, se sugiere que las recomendaciones de tratamientos se hagan para cada localidad en particular.

4) Tratamientos x Localidades x Años:

Es un tanto más complicada de interpretar. Indica una respuesta diferencial de los tratamientos, tanto entre localidades como en diferentes años. Es decir el "ranking varió" según la localidad y el año.

Esta interacción implica que a menos que CM de tratamientos haya excedido significativamente al CM_{TXA} o a CM_{TXL}; no se podrán generalizar recomendaciones, tanto a toda la región como a cualquier período de tiempo.

5) Tratamientos.

No por ser menos importante se dejó de último la consideración de la significancia del efecto de tratamiento, sino por el hecho de que la misma debe hacerse conjuntamente con las interpretaciones de las interacciones antes presentadas, solo en el caso de significancia de tratamientos y no así de TXA, TXL y TXLXA, diremos finalmente que se interprete que los tratamientos tuvieron una respuesta significativamente cualquiera que haya sido la localidad y el año de prueba; y por lo tanto las recomendaciones sobre los mejores tratamientos podrán extenderse a toda la región y a cualquier período de tiempo.

2.4.3 Experimentos con Cultivos Perennes

2.4.3.1 Repeticione en tiempo.

En el caso más simple, si se tratara de un solo experimento que tiene permanencia a lo largo del tiempo por tratarse de un cultivo perenne, como en los experimentos con cafetos por ejemplo, hay tres alternativas de análisis :

a) Sumar las observaciones anuales de una variable y esta suma, o cualquier otra forma lineal, constituye la nueva variable a ser analizada como si se tratará de un experimento simple. Esta alternativa sin embargo no permite deslindar o conocer el efecto ambiental de los diferentes años o períodos.

b) Las observaciones anuales de una misma variable son consdieradas como variables independientes. El conjunto se puede someter a un análisis multivariado para probar hipotesos sobre efectos simultáneos de esas variables.

En este caso, si por ejemplo se midió rendimiento por parcela durante cuatro años, las variables x_1, x_2, x_3, x_4 que son los datos de rendimiento de cada año, constituyen un conjunto de variables independientes sobre las que se formulan hipótesis de efectos simultáneos y se utilizan criterios de prueba estadística como el de Wilks (5).

c) Realizar un análisis de varianza combinado, de los datos individuales por año, conforme el esquema de análisis del diseño en Bloques Divididos (Split-block).

Conforme a la última alternativa, los componentes del análisis de varianza se indican en el Cuadro. 7. Dicho análisis corresponde a un modelo lineal para una observación experimental cualquiera :

$X_{ijk} = \mu + A_i + \beta_j + A\beta_{ij} + \gamma_k + a\gamma_{i.k} + B\gamma_{.jk} +$ ijk, donde u es la media; y A_i, β_j, γ_k son los efectos del tramiento i, el bloque j y el año k, que corresponden con la observación. También se consideran las interacciones de dos factores; y e_{ijk} es el error experimental.

Cuadro 7. ANVA para el Diseño de Bloques Divididos

Fuente de Variación	Grado de Libertad	Valor esperado del cuadrado Medio*
Total	rta-1	----
Repeticiones (R)	r-1	$\sigma e^2 + a\sigma_{tr}^2 + t\sigma_{ta}^2 + at\sigma_r^2$
Tratamientos (T)	t-1	$\sigma e^2 + a\sigma_{rt}^2 + r\sigma_{ta}^2 + \frac{ra\Sigma t^2}{t-1}$
Tx R : Error a	(r-1)(t-1)	$\sigma e^2 + a\sigma^2 rt$
Años (A)	a-1	$\sigma e^2 + t\sigma_{ra}^2 + rt\sigma_a^2$
Ax R : Error b	(a-1)(r-1)	$\sigma e^2 + t\sigma_{ra}^2$
A x T	(a-1)(t-1)	$\sigma e^2 + r\sigma_{ta}^2$
AxTxR = Error c	(a-1)(t-1)(r-1)	σe^2

A la parcela principal se adjudica el factor tratamientos (t), en consideración de que éstos, ocupan fisicamente la parcela y permanecen durante todos los años del experimento; y luego los datos de cada año representarían las subparcelas, con la limitación de que los Años (A) no tienen realmente un sorteo sino que ocurren en una secuencia fija a lo largo (o en faja) de las parcelas principales. Esto último impone restricciones a las inferencias sobre el efecto de años; sin embargo esta alternativa de análisis es recomendable porque permite separar y conocer el efecto ambiental de años y la interacción años por tratamientos, com -

*Los sufijos: e.r.t.a corresponden con error experimental, repeticiones, tratamientos y años respectivamente.

ponentes que no contribuyen al error experimental. Al resultar no significativa la interacción AxT, ello indica cierta estabilidad o consistencia del efecto de tratamientos a lo largo del tiempo.

Cuadro 8. Resultados de un ensayo de rendimiento de variedades de café en los años que se indican. Peso de café una en kg. por parcela (X_{ijk}). Campamento, Olancho.

Año/Repetición		V_1	V_2	V_3	TOTAL
1983	I	4.40	3.83	3.60	11.83
	II	5.60	5.00	4.90	15.50
	III	4.10	4.50	4.10	12.70
Sub-total		14.10	13.33	12.60	40.03
1984	I	4.61	4.20	3.59	12.40
	II	5.20	5.25	4.50	14.95
	III	5.30	4.75	5.60	15.65
Sub-Total		15.11	14.20	13.69	43.00
1985	I	9.10	7.41	4.00	20.51
	II	6.99	7.01	4.92	18.92
	III	7.15	6.18	3.72	17.05
Sub-total		23.24	20.60	12.64	56.48
TOTAL		52.45	48.13	38.93	139.51

Considerando que el período ó años de ejecución del experimento es una muestra aleatoria de las condiciones ambientales en un lapso más prolongado de tiempo, se estima como aleatorio el efecto de años (τ_k), también el de repeticiones o bloques (B_j). Por otra parte, los efectos de tratamientos son generalmente fijos (Ai), debido a que estos no constituyen muestras aleatorias de una población de tratamientos. Los valores esperados de los componentes de varianza en este tipo de análisis son los que se indican en el cuadro 7, con lo cual se pueden identificar los divisores apropiados para las pruebas "F" de significancia estadística.

Presentamos a continuación un ejemplo numérico de este tipo de análisis. Se trata de un ensayo de rendimiento de tres variedades de café V_1, V_2 y V_3, que se condujo en una localidad bajo el diseño de Bloques al azar con tres repeticiones, durante cinco años; aunque para abreviar el ejemplo se consideran los datos de tres años, que se presentan en el cuadro 8.

El cálculo de los componentes de varianza indicados en el Cuadro 7, principia con el correspondiente a la variación total, cuya suma de cuadrados (SC total) se obtiene de las observaciones individuales del Cuadro 8, así :

SC total $= \sum X^2_{ijk} - FC$, $FC = \frac{(X....)^2}{rta}$, donde : X... es el gran total,

$i = 1,2,\ldots, r$ número de repeticiones.

$j = 1,2,\ldots, t$ número de tratamientos principales (variedades).

$k = 1,2,\ldots, a$ número de subtratamientos (años).

$$SC\ total = 4.40^2 + \ldots + 3.72^2 - \frac{(139.51)^2}{3(3)(3)}$$

$$= 768.9521 - 720.8533 = 48.0988$$

Prosigue el cálculo de la variación entre parcelas principales, para lo cual se agrupan los datos en un cuadro de tratamiento x repeticiones (Cuadro 9).

Cuadro 9. Totales por parcela (X_{ij}) en los tratamientos y repeticiones que se indican.

Tratamiento	Rep. I	Rep. II	Rep. III	Total ($X_{i..}$)
V_1	18.11	17.79	16.55	52.45
V_2	15.44	17.26	15.43	48.13
V_3	11.19	14.32	13.42	38.92
Total ($X_{.j.}$)	44.74	49.37	45.40	139.51 ($X_{...}$)

En el cuadro 9, también se muestran los totales marginales para tratamientos ($X_{i..}$) y repeticiones ($X_{.j.}$), que sirvan de contraste para calcular las respectivas sumas de cuadrados (tratamientos y repeticiones). Entonces :

$$\text{SC parcelas Principales} = \frac{\sum X_{ij.}^2}{a} - FC$$

$$= \frac{18.11^2 + \ldots + 13.42^2}{3} = 720.8533 = 13.5199$$

Se puede advertir que esta SC cuyo valor es 13,5199, se reparte entre los componentes: repeticiones, tratamientos y error a , indicados en el Cuadro 7. Además, como se conoce la SC de variación total que es 48.0988, por diferencia

con respecto a 13.5199 se puede calcular la variación restante que corresponde a : subtratamientos (años), error b, interacción TxA y error C; es igual a :

SS subparcelas = 48.0988 - 13.5199 = 34.5789

$$\text{SC Tratamientos} = \frac{\sum X_{i}^{2}\ldots - FC}{ra}$$

$$SC\ T = \frac{52.45^2 + 48.13^2 + 38.93^2}{(3)\ (3)} - 720.8533 = 10.5961$$

$$\text{SC repeticiones} = \frac{\sum X.j^2, - FC}{ta}$$

$$= \frac{44.74^2 + 49.37^2 + 45.40^2}{(3)\ (3)} - 720.8533 = 1.3939$$

Un resumen parcial de la variación entre parcelas, se muestra así :

Fuente	G.L.	SC.
Repeticiones	r-1=2	1.3939
Tratamientos	t-1=2	10.5961
Error a	(r-1)(t-1)=4	1.5298
TOTAL	rt-1=9	13.5199

La suma de cuadrados por error a fue obtenida por diferencia, así

$$SC_{e\ (a)} = 13.5199-10.5961 - 1.3939 = 1.5298$$

Un procedimiento similar se utiliza para obtener las sumas de cuadrados del resto de componentes, conociendo que se tiene que hacer un reparto de la variación entre subparcelas, en este caso el valor 34.5789.

Las sumas de cuadrados para subtratamientos (años) y error b se obtienen del Cuadro 10 de Repeticiones por subtratamientos (años) :

Cuadro 10. Totales en las combinaciones ($X._{jk}$) de repeticiones y años que se indican.

Año	Rep. I	Rep. II	Rep. III	TOTALES (X..k)
1983	11.83	15.50	12.70	40.03
1984	12.40	14.95	15.65	43.00
1985	20.51	18.92	17.05	56.48

La suma de cuadrados para subtratamientos es :

$$SC\ años = \frac{\sum X..k^2}{tr} - FC$$

$$= \frac{40.03^2 + 43.00^2 + 56.48^2}{9} - 720.8533 = 17.0791$$

Para la interacción AxR o error b :

$$SC\ AxR = \frac{\sum X.jk^2}{t} - FC - SC\ años - SC\ reps.$$

$$= \frac{11.83^2 + \ldots + 17.05^2}{3} - 720.7533 - 8.5396 - 1.3939 = 5.0080$$

La interacción Txa se obtiene igualmente del Cuadro 11.

Cuadro 11. Totales en las combinaciones ($\Sigma X_{i.k}$) de tratamiento y años que se indican.

Tratamiento	1983	1984	1985	TOTAL
V_1	14.10	15.11	23.24	52.45
V_2	13.33	14.20	20.60	48.13
V_3	12.60	13.69	12.64	38.93

$$SC\ TxA = \sum Xi^2.k - FC - SC_T - SC_A$$

$$= \frac{14.10^2 + \ldots + 12.64^2}{3} - 720.8533 = 10.5961 - 8.5396$$

$$= 10.4262$$

El error c se calcula finalmente por diferencia; y el cuadro general del análisis se presenta a continuación :

Cuadro 12. Análisis de Varianza combinao. Peso de Café por parcela

Fuente de Variación	Grados de Liberta	Suma de Cuadrados	Cuadrado medio	"F"
TOTAL	26	48.0988		
Repeticiones	2	1.3939	0.6970	
Variedades (T)	2	10.5961	5.2980	(F^1=1.94)
Error a	4	1.5298	0.3824	
Años (A)	2	17.0791	8.5396	6.82
Error b	4	5.0080	1.2520	----
T x A	4	10.4262	2.6066	10.10**
Error c	8	2.0656	0.2582	----

En el Cuadro 12 se han indicado los valores de la estadística de prueba "F", dónde para este caso particular en que los tratamientos tienen efecto fijo y los subtratamientos aleatorio, no hay de inmediato un término de error apropiado para calcular "F" de tratamientos (ver Cuadro 7). Como en la sección 4.1, se ha recurrido a la prueba ajustada de Sattertnwaitte (4) construyendo un término de error para tratamientos, dado por :

$$E' = CM \text{ error } \underline{a} + CM_{TXA} - CM \text{ error } \underline{c}$$

Ya que según cuadro 7, la prueba apropiada es :

$$"F" = \frac{\sigma e^2 + a\sigma rt^2 + r\sigma^2 ta + a\Sigma\delta i^2}{\sigma e^2 + a\sigma rt + r\sigma^2 ta}$$

Entonces: F' aprox es :

$$F' = \frac{CM\ trat}{E'} = \frac{5.2980}{0.3824+2.6066-0.2582} = \frac{5.2980}{2.7308} = 1.94$$

Los grados de libertad para este error, son aproximadamente :

$$F'\ E' = \frac{(E')^2}{\frac{(CM_{e\ (a)})^2}{f_a} + \frac{(CM\ e\ (b)^2}{fb} + \frac{(CM_{TXA})^2}{ta}}$$

$$= \frac{2.7308^2}{\frac{0.3824^2}{4} + \frac{1.2520^2}{4} + \frac{2.6066^2}{4}} = \frac{10.9232}{8.5081} = 1$$

El resultado de este ejemplo es que la mayor variación se observó entre años, lo cual es lógico cuando se trata de una plantación jóven, aunque esta variación en este ejemplo no llega al umbral de significación estadística. Entre variedades, según la prueba aproximada, no hay diferencias significativas; y

y por otra parte la interacción variedad x años es significativa, indicando que la respuesta de variedad no es estable con respecto al tiempo y en la realidad puede observarse que la variedad V_3 disminuyó su rendimiento al tercer año, al contrario de V_1 y de V_2. Se podrá deducir de este ejemplo que las pruebas de significancia efectuadas tienen precisión cuando se tienen pocas repeticiones o tratamientos, debido al bajo número de grados en los divisores correspondientes, con lo cual es poco probable detectar pequeñas diferencias reales entre tratamientos.

2.4.2.2. Repeticiones en Tiempo y Espacio.

En el caso de experimentos con plantas perennes, repetidos en varias localidades, la metodología de análisis es una extensión del caso anterior, incluyendo la dimensión adicional de espacio; para determinar los componentes de varianza en el análisis combinado.

Por ejemplo, si existe un conjunto de experimentos similares al descrito en la sección 4.2.1. conducidos en varias localidades es posible un análisis combinado con la siguiente distribución de componentes de varianza.

CUADRO 13 Análisis de varianza combinado de un grupo de experimentos con plantas perennes. Diseño unitario: bloque al azar.

Fuente de ariación	Grados de Liberta	Valor esperado del cuadrado medio Modelo mixto: trat.fijo y resto aleatorio
TOTAL	lrta - 1	
Entre experimentos o localidades	$\ell-1$	$\sigma_e^2 + tr\sigma_{la}^2 + ar\sigma_{lt}^2 + r\sigma_{atl}^2 + at\, r\sigma_l^2$
Repeticiones en localidades l (R)	$\ell(r-1)$	
Tratamientos (T)	$t-1$	$\sigma_e^2 + \ell a\, \sigma^2 rt + lr\sigma_{ta}^2 + r\sigma_{atl}^2 + \frac{alr}{t-1}\sum t^2$
LXT	$(t-1)(\ell-1)$	$\sigma_e^2 + r\,\sigma_{atl}^2 + ar\sigma_{lt}^2$
Error a	$\ell(t-1)(r-1)$	$\sigma_e^2 + \ell a\,\sigma_{rt}^2$
Años (A)	a-1	$\sigma_e^2 + \ell t\,\sigma_{ra}^2 + rt\,\sigma_{al}^2 + rtl\sigma_a^2$
AxL	$(a-1)(\ell-1)$	$\sigma_e^2 + \ell t\,\sigma_{ra}^2 + rt\,\sigma_{al}^2$
Error b combinado	$\ell(a-1)\ (r-1)$	$\sigma_e^2 + \ell t\,\sigma_{ra}^2$
TxA	(t-1 (a-1)	$\sigma_e^2 + r\,\sigma_{atl}^2 + lr\,\sigma_{ta}^2$
TxAxL	$(t-1)(a-1)(\ell-1)$	$\sigma_e^2 + r\,\sigma_{atl}^2$
Error c combinado	$\ell(t-1)(a-1)(r-1)$	σ_e^2

El análisis de varianza presentado en el Cuadro 13 se basa en un modelo lineal para una sola observación cualquiera Xijkl como el siguiente :

$$Xijkl = \mu + Li + li(\beta_j) + \tau k + L\tau ik + l_i(\beta\tau) + A_l + LA_{il}$$
$$+ li(A\beta) + a\tau_{kl} + LA\tau_{ikl} + \epsilon_{ijkl},$$

donde Li; β_j, τk y A_l son l o s efectos particulares de localidad i, repetición j, tratamiento k y del año l, a los cuales corresponde la observación Xijkl.

Se incluye en el modelo tres interacciones de dos factores y una de tres factores, entre localidades, tratamientos y años. Además, los componentes li (β_j), li ($\beta\tau$) y li (Aβ) y el error ϵijkl; los cuales son efectos "anidados" de repeticiones y las interacciones TxR, AXR; TxRxA (errores "a", "b" y "c" de los análisis de cada localidad).

Los valores esperados de los componentes de varianza, en el Cuadro 13, fueron derivados del anterior modelo, considerando que los efectos de tratamientos son fijos (τi) y los restantes efectos son aleatorios.

En el Cuadro 14 se presentan las expresiones y procedimientos para el cálculo numérico de las sumas de cuadrados de cada componente del análisis de varianza. Como podrá observarse, dicho cálculo se facilita si previamente se han hecho los análisis combinados para cada localidad, dado que los componentes de varianza de efectos "anidados" son en el análisis global, las sumas de los valores parciales o de cada localidad.

CUADRO 14. SUMAS DE CUADROS DE LOS COMPONENTES DE VARIANZA

COMPONENTES	DEFINICION O PROCEDIMIENTO DE CALCULO
SC TOTAL =	$\sum_i \sum_j \sum_l \sum_l (x_{ijkl} - \bar{x}_{....})^2 = \sum_i \sum_j \sum_k \sum_l x^2_{ijkl} - FC_g$
	donde: factor de corrección general $FCg = x^2_{..../lrta} = \frac{(\sum_i \sum_j \sum_k \sum_l Xijkl)^2}{lrta}$
Localidades: = SC (L)	$\sum_i (\bar{x}_{i...} - \bar{x}_{....})^2 = \frac{\sum_i x^2_{i....}}{rta} - FCg$
Tratamientos = (SCCT)	$\sum_k (\bar{X}_{..k.} - \bar{X}_{...})^2 = \frac{\sum_k X^2_{...k.}}{lra} - FCg$
SCTxL	$\sum_i \sum_j (\bar{x}_{i.k.} - \bar{X}_{i...} - \bar{X}_{..k.} + \bar{X}_{....})^2 = \frac{\sum_i \sum_k X^2_{i.k.}}{ra} - FCg - SC(L) - SC(T)$
Repeticiones en= localidades	Se suman los valores SC repeticiones de los análisis individulaes de cada localidad.
Error a combinado	Se suman los valores de SC error a, de cada localidad
SC (A) =	$\sum_l (\bar{X}_{...l} - \bar{X}_{...})^2 = \frac{\sum_l X^2_{...l}}{lrt} - FCg$
SC AxL	$\sum_i \sum_l (\bar{X}_{i..l} - \bar{X}_{i..} - \bar{X}_{..l} + \bar{X}_{...})^2 = \frac{\sum_i \sum_l Xi^2_{..l}}{rt} - FCg - SC(A) - SC(A)$
Error b combinado	Se suman los valores de SC error b, de cada localidad
SC T x A	$\sum_k \sum_l (\bar{X}_{..kl} - \bar{X}_{..k} - \bar{X}_{..l} + X_{..})^2 = \frac{\sum_k \sum_l X^2_{..kl}}{lr} - FCg - SC(T) - SC(A)$
Error c combinado SCTxAxL	Se suman los valores de SC error c, de cada localidad. Por diferencia, entre SCtotal y las SC previamente calculadas.
Sub-índices utilizados	Localidades: 1,2,,,i...L repeticiones:1,2,...j..r Tratamientos:1.2,..k.t.. Años: 1.2..l,.a

Para ilustrar el procedimiento de cálculo, utilizamos los resultados del cuadro 8 obtenidos en Campamento y agregamos los de otro ensayo similar, conducido en Catacamas. El conjunto de resultados en las dos localidades indicadas, se muestra en el siguiente cuadro 15.

Cuadro 15. Rendimientos experimentales (Kg café por parcela) en dos localidades; Campamento y Catacamas, cosechas 1985 y 1986

Variedad/Rep.		CAMPAMENTO				CATACAMAS			
		1983	1984	1985	Total	1983	1984	1985	Total
V1	Rep 1	4.40	4.61	9.10	18.11	6.73	3.94	11.43	22.10
	Rep 2	5.60	5.20	6.99	17.79	6.90	4.56	9.32	20.78
	Rep 3	4.10	5.30	7.15	16.55	6.47	4.69	9.47	20.63
V2	Rep 1	3.83	4.20	7.41	15.44	4.08	4.50	8.41	16.99
	Rep 2	5.00	5.25	7.01	17.26	5.25	5.52	8.20	18.97
	Rep 3	4.50	4.75	6.18	15.43	4.76	5.01	7.60	17.37
V3	Rep 1	3.60	3.59	4.00	11.19	6.15	3.79	5.62	15.56
	Rep 2	4.90	4.50	4.92	14.32	6.46	4.78	6.48	17.72
	Rep 3	4.10	5.60	3.72	13.42	5.67	5.96	5.72	17.35

Se realizó un análisis para Catacamas en forma similar al de Campamento según la metodología indicada en la sección 4.2.1.Los resultados de estos análisis son útiles en la determinación de sumas de cuadrados totales y de componentes "anidados"; y los resultados parciales se muestran adelante en el cuado 18.

Siguiendo las expresiones de cálculo del cuadro 14, se obtiene lo siguiente:

$$\text{SC Total} = \sum_i \sum_j \sum_k \sum_l X_{ijkl}^2 - FC\ g$$

$$\text{SC Total} = 4.40^2 + \ldots + 2.72^2 + 6.73^2 \ldots + 5.72^2 - FC\ g$$

$$FCg = \frac{306.98^2}{54} = 1745.1245$$

$$\text{SC Total} = 1989.3634 - 1745.1245 = 153.2389$$

Para obtener las SC de localidades, tratamientos e interacción de estos dos factores, se utilizarán totales como los que se muestran en el cuadro 16.

Cuadro 16. Totales para las combinaciones de localidad, tratamiento y año que se indican.

Variedad	CAMPAMENTO				CATACAMAS				Total
	1983	1984	1985	Subtotal	1983	1984	1985	Subtoal	
V1	14.10	15.11	23.24	52.45	20.10	13.19	30.22	63.51	115.96
V2	13.33	14.20	20.60	48.13	14.09	15.03	24.21	53.33	101.46
V3	12.60	13.69	12.64	38.93	18.28	14.53	17.82	50.63	89.56
TOTAL	40.03	43.00	56.48	139.51	52.47	42.75	72.25	167.47	306.98

$$\text{SC localidades} = \frac{\sum X_{i\ldots}^2}{rta} - FCg$$

$$= \frac{139.51^2 + 167.47^2}{27} - 1745.1245$$

$$\text{SC localidades} = 1759.6015 - 1745.1245 = 14.4771$$

$$\text{SC tratamientos} = \frac{\sum X_{..k.}^2}{lra} - FCq$$

$$= \frac{115.96^2 + 101.46^2 + 89.56^2}{18} - 1745.1245$$

SC tratamientos = 1764.5470 - 1745.1245 = 19.4226

$$SC\ TXL = \sum\sum \frac{x_{i.k}^2}{ra} - FCg - SC\ (L) - SC(T)$$

$$= \frac{52.45^2 + \ldots + 50.63^2}{9} - 1745.1245 - 14.4771 - 19.4226$$

SC TxL = 1780.4500 - 1745.1245 - 14.4771 - 19.4226 = 1.4259

SC repeticiones en localidades

SC repa (L) = 1.3939 + 0.4791 = 1.8730 (Cuadro 18)

SC error a combinado:

SC error a= 1.5298 + 1.6977 = 3.2275 (Cuadro 18)

El Cuadro 17 es un resúmen de totales que facilita los contrastes para calcular los SC de años y de años x tratamientos :

Cuadro 17: Totales para las combinaciones de año y variedad (Kg. uva)

Año	1983	1984	1986	Total
Variedad				
V1	34.20	28.30	53.46	115.96
V2	27.42	29.23	44.81	101.46
V3	20.88	28.22	30.46	89.56
Total	92.50	85.75	128.73	306.98

Cuadro 18: Componentes de varianza, en los análisis de cada localidad.

Fuente	Grados de Libertad	Campamento		Catacamas	
		SC	CM	SC	CM
Total	26	48.0988	-	90.6630	0.2396
Repeticiones	2	1.3939	0.6970	0.4791	0.2396
Tratamientos	2	10.5961	5.2980	10.2520	5.1260
Error a	4	1.5298	0.3824	1.6977	0.4244
Años	2	17.0791	8.5396	50.2210	25.1105
Error b	4	5.0080	1.2520	3.5986	0.8997
Trat.x años	4	10.4262	2.6066	22.3180	5.5795
Error c	8	2.0656	0.2582	2.0956	0.2620

$$\text{SC años} = \frac{\sum_{1} x^2 \ldots 1}{lrt} - FCg$$

$$= \frac{92.50^2+85.75^2+128.73^2}{18} - 1745.1245$$

SC años = 1804.4847 - 1745,1245 = 59.3603

$$\text{SC años x localidad} = \frac{\Sigma\Sigma x^2_{i..\ell}}{rt} - FC_g - SC(A) - SC\ (L)$$

$$SCAxL + \frac{40.03^2 + \ldots + 72.25^2}{9} - 1745.1245 - 59.3603 - 14.4771$$

SC AxL = 1826.9019 - 1745.1245 - 59.3603-14.4771 = 7.9401

Error b combinado

SC error b = 5.0080 + 1.2520 = 8.6066 (Cuadro 18)

$$SC\ tratamientos\ x\ años = \sum\sum \frac{x^2_{..tl}}{lr} - FC\ g - SC\ (T) - SC\ (A)$$

$$SC\ TxA = \frac{34.20^2 + \ldots + 30.46^2}{6} - 1745.1245 - 19.4226 - 59.3603$$

SC TxA = 1853.4069 - 1745.1245 - 19.4226-59.4603 = 29.4996

Error c combinado

SC error c = 2.0656 + 2.0956 + 2.0956 = 4.1612 (Cuadro 18)

SC TxAxL = SC total - SC de componentes anteriors

= 153.2389 - 149.9939 = 3.2450

Los cálculos anterioes conducen al cuadro general de análisis de varianza

Cuadro 19 Análisis combinado de los dos experimentos

Fuente	Grados de Libertad	Suma de Cuadrados	Cuadrado medio	"F"
Total	53	153.2389	-	
Localidad (L)	1	14.4771	14.4771	4.01
Reps. en localidades	4	1.8730	0.4683	
Tratamientos (T)	2	19.4226	9.7113	1.29
LxT	2	1.4259	0.7130	0.88
Error a combinado	8	3.2275	0.4034	-
Años (A)	2	59.3603	29.6802	3.48
AxL	2	7.9401	3.9701	3.69*
Error b combinado	8	8.6066	1.0758	-
TxA	4	29.4996	7.3749	9.09*
TxAxL	4	3.2450	0.8113	3.12*
Error c combinado	16	4.1612	0.2601	

*Significativo P= 0.05

Para concluir el análisis, que se presenta en el Cuadro 19, se puede indicar que las únicas pruebas de significancia que requieren un poco de atención, por no ser visibles de inmediato cuales son los divisores apropiados para calcular la estadística "F", son los de efectos de localidades y de tratamientos, los que por cierto son de gran importancia.

Según el modelo mixto de este ejemplo, se podrá observar en el cuadro 13, que los valores esperados de los componentes de varianza indican que para el efecto de localidades se necesita un divisor de la forma :

$$\sigma_e^2 + tr\,\sigma_{la}^2 + ar\,\sigma_{lt}^2 + r\,\sigma_{atl}^2,$$

Para la prueba F; el cual se obtiene con :

$$E' = CM_{LT} + CM_{AL} - \text{Error } b$$

$$\text{Entonces "F"} = CM_L = \frac{14.4771}{0.7130 + 3.9701 - 1.0758} = 4.01$$

El divisor utilizado tiene aproximadamente 2 grados de libertad, conforme a la expresión :

$$f_2 = \frac{(E')^2}{\frac{Eb^2}{fb} + \frac{CM_{LT}}{fet} + \frac{CM^2_{AL}}{fal}}$$

$$f_2' = \frac{(3.6073)^2}{\frac{(1.0758)^2}{8} + \frac{(0.7130)^2}{2} + \frac{(3.9701)^2}{2}} = 1.6$$

La prueba "F" para tratamientos, también requiere la elaboración de un divisor apropiado, de manera similar el caso de análisis de una sola localidad, visto en la sección 2.4.1. Este divisor es:

$$E'' = CM_{TA} + \text{Error a} - \text{Error c} = 7.5182,$$

el cual tiene aproximadamente grados de libertad dados por :

$$\sigma_2'' = \frac{(E'')^2}{\frac{Ea^2}{fa} + \frac{Eb^2}{fb} + \frac{CM_{TA}^2}{ta}} = \frac{(7.5182)^2}{\frac{(0.4034)^2}{8} + \frac{1.0758)^2}{8} + \frac{(7.3749)^2}{4}} = 4.11$$

Entonces :

$$"F" = \frac{9.7113}{7.5182} = 1.20 \quad \text{con 2,4 grados de liberta}$$

Para los restantes efectos e interacciones, las pruebas de significancia se realizan por observación directa de los divisores adecuados sugeridos por los valores esperados del cuadro 13. Asi se tiene que:

Interacción LxT "F" $= \frac{CM_{LT}}{CM_{LTA}} = \frac{0.7130}{0.8113} = 0.88$

Efectos de años "F" $= \frac{CM_A}{CM_{AL}} = \frac{29.6802}{3.9701} = 7.48$

Interacción AxL "F" $= \frac{CM_{AL}}{Eb} = \frac{3.9701}{1.0758} = 3.69$

Interacción TxA "F" $= \frac{CM_{TA}}{CM_{LTA}} = \frac{7.3749}{0.8113} = 9.09$

interacción LXTXA 'F' $= \frac{CM_{LTA}}{Ec} = \frac{0.8113}{0.2601} = 3.12$

Todos los valores de F calculados se muestran en el cuadro 19, así como la indicación de cuales pruebas alcanzaron o no el nivel de significancia estadística a 0.05 de probabilidad.

Aunque son notorias las diferencias entre localidades y entre años, estas no alcanzaron el umbral de significancia estadística pero si las interacciones AXL y AXT, lo cual implica que las inferencias basadas en los resultados de estos dos ensayos experimentales deberán restringirse a cada lugar en particular y que tampoco podrán establecerse conclusiones para un mayor período de tiempo, que el de los datos (1983-1985). Las diferencias entre tratamientos (variedades de café) no son significativas, lo que es coincidente con los resultados del análisis de cada sitio en particular, Además las variedades respondieron igual en las tres localidades, como lo indica el hecho de que la interacción TXL no sea significativa; lo que manifiesta cierta consistencia de las variedades en los dos lugares; y por otra parte, la respuesta de las variedades sí fue diferente durante los tres años del ensayo (interacción TXA es significativa). En los resultados anteriores, debe tenerse en cuenta que en este ejemplo, posiblemente no se detectaron diferencias reales entre localidades y entre años, por el bajo número de localidades y de años en el estudio, si es que se asume un efecto aleatorio de estos factores, con el propósito de hacer inferencias para una región geográfica y para varios años.

5. LITERATURA CITADA

1. BANCROFT, T.A. "Analysis and inference for incompletely specified models involving the preliminary test of significance "Biometrics 20: 427-442.1964

2. BARTLETT, M.S. "Properties of Sufficiency and Statistical Test" Citado por Le Clerg et.al. "Field Plot Technique" 2da. Ed. Burgess Publishing Co. P. 230-233. 1962.

3. COCHRAN, W.G. y Cox, G.M. "Experimental Designs" 2da. Ed. John Wiley. 611 pág. 1955

4. SATTERTHWAITE, F.E. " An approximate Distribution of Estimates of Variance Components". Citado por Le Clerg et. al: "Field Plot Technique" 2da. Ed. Burgess Publishing Co. P. 228. 1962.

5. STEEL, R.G.D. "An Analysis of Perennial Crop Data". Biometrics June: 201-212. 1955.

3. ANALISIS DE REGRESION LINEAL SIMPLE.

F. Omar Osorio.

Un moddelo de regresión define una relación de dependencia. Al estar en conocimiento de dos variables relacionadas (o concomitentes), el experimentador busca un medio para expresar las relaciones funcionales entre ambas variables. Es decir, buscará establecer una función de tipo matemático que permita expresar el tipo y características de la relación entre variables, así mismo es posible que se interese en conocer el grado de precisión que se tiene al predecir el valor de una variable teniendo el valor de la variable relacionada. Esos objetivos pueden alcanzarse a través del análisis de regresión; en este sentido puede decirse que el análisis de regresión busca la mejor relación funcional entre una variable (dependiente) y una o más variables (independientes).

3.1 MODELO DE REGRESION LINEAL SIMPLE

Considerese el modelo siguiente :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad ; \; i = 1.2,.....,n$$

donde:

Y_i = variable aleatoria observable

X_i = constante conocida

ε_i = variable aleatoria no correlacionada, con media cero y varianza común.

β_0, β_1 = constantes desconocidas, parámetros del modelo

El modelo descrito se conoce como "modelo de regresión lineal simple".

A la variable Y_i se le denomina variable dependiente pues cualquier valor de Y_i depende de la población muestreada, la variable X_i es denominada variable independiente o argumento. El propósito del análisis consiste, entonces, en buscar la relación que permita explicar la variación observada en la variable dependiente como función de la variable independiente, siempre que tal relación exista.

Los valores β_0, β_1 y ε_i, son desconocidos, en efecto ε_i, es difícil de determinar pues cambia con cada observación Y_i, sin embargo, β_0 y β_1 permanecen fijos y, aunque no podamos conocer su valor exacto, es posible obtener sus estimadores b_0 y b_1 , así el modelo de regresión lineal simple estimado puede escribirse de forma :

$$\hat{Y}_i = b_0 + b_1 X_i \quad ; \; i = 1.2, \ldots\ldots, n$$

Esta ecuación puede ser utilizada como uan ecuación predictiva, conociendo un valor de X_i podemos predecir el valor medio verdadero de Y_i para tal valor de X_i

El procedimiento de estimación de β_0 y β_1 (b_0 , b_1) que se utiliza, es el de mínimos cuadrados, este procedimiento busca encontrar los valores de la ecuación que minimicen las desviaciones de los valores observados con relación a los valores estimados por la ecuación ajustada ($Y_i - \hat{Y}$).

Si

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad ; \; i = 1.2, \ldots\ldots, n$$

La suma de cuadrados de las desviaciones con relación a la linea de regresión es :

$$S = \Sigma \varepsilon_i^2 = \Sigma \left(Y_i - \beta_0 - \beta_1 X_i \right)^2$$

Para buscar los valores de b_0 y b_1 que hagan un mínimo el valor de S, derivamos la función S primero con respecto a β_0 y luego con respecto a β_1 ,

igualando a cero ambas derivadas

$$\partial S/\partial \beta_0 = -2\,\Sigma(Y_i - \beta_0 - \beta_1 X_i)$$

$$\partial S/\partial \beta_1 = 2\,\Sigma X_i (Y_i - \beta_0 - \beta_1 X_i)$$

sustituyendo b_0 , b_1 por β_0, β_1

$$\Sigma(Y_i - b_o - b_1 X_i) = 0$$
$$\Sigma X_i (Y_i - b_o - b_1 X_i) = 0$$
$$\Sigma Y_i - nb_o - b_1 \Sigma X_i = 0$$
$$\Sigma X_i Y_i - b_o \Sigma X_i - b_1 \Sigma X_i^2 = 0$$

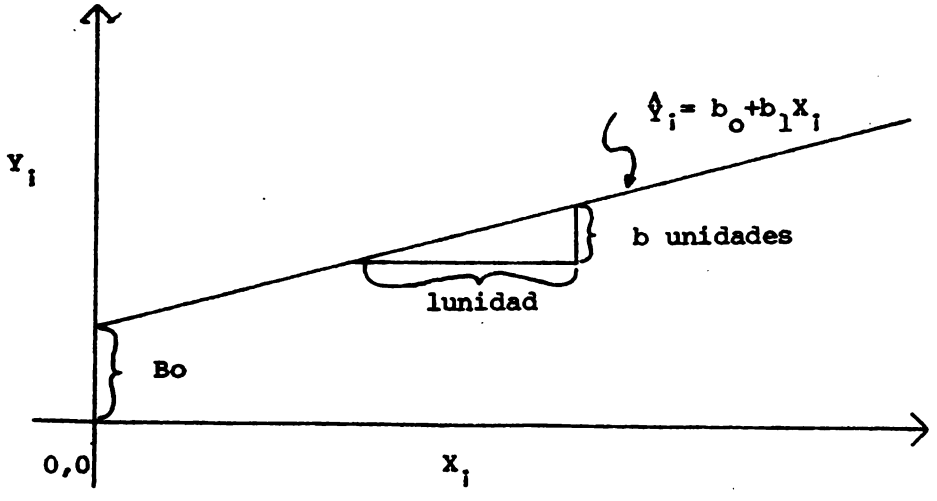
ordenando :

$$b_o n + b_1 \Sigma X_i = \Sigma Y_i$$
$$b_o \Sigma X_i + b_1 \Sigma X^2_i = \Sigma X_i Y_i$$

a estas ecuaciones se les denomina "ecuaciones normales", a prtir de ellas se puede despejar los valores de b_0 y b_1 , en este curso sólo se harán estimaciones mediante el uso del algebra de matrices, otras estimaciones más - convencionlaes se explican detalladamente en todas las referencias anotadas al final del tema

3.2 Interpretación

Representemos el modelo $\hat{Y}_i = b_0 + b_1 X_i$ por la figura siguiente :



Cualquier punto en la linea de regresión tiene una abscisa Y_i y una ordenada X_i que satisfacen la ecuación. Cuando X_i toma el valor cero (0), b_0 es el punto donde la linea de regresión Intercepta el eje de Y_i. Un cambio de 1 unidad en X resulta en un cambio de b_1 unidades en Y, y por esto se dice que b es una medida de la pendiente de la linea de regresión. En el contexto anterior, a la ecuación se le denomina ecuación de regresión, a la linea linea de regresión y a los estimadores b_0 y b_1 coeficientes de regresión.

3.2 Estimación Mediante el uso de Algebra de Matrices.

Sea el modelo de regresión lineal simple :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad ; \quad 1,2,\ldots\ldots, n$$

la expresión puede expandirse y,

$$\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_{11} + \varepsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_{21} + \varepsilon_2 \\
Y_3 &= \beta_0 + \beta_1 X_{31} + \varepsilon_1 \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
Y_n &= \beta_0 + \beta_1 X_n 1 + \varepsilon n
\end{aligned}$$

este sistema de ecuaciones puede representarse en forma matricial por:

$$Y = X\beta + \varepsilon$$

donde:

Y = un vector (n x 1) de valores observados

X = una matriz (n x2) de forma conocida

β = un vector (2 x 1) de parámetros

ε = un vector (n x 1) de errores

$$\underset{Y}{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}} = \underset{X}{\begin{bmatrix} 1 & X_{11} \\ 1 & X_{21} \\ 1 & X_{31} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_{n1} \end{bmatrix}} \underset{\beta}{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}} + \underset{\varepsilon}{\begin{bmatrix} 1 \\ 2 \\ 3 \\ \\ \\ \\ \varepsilon_N \end{bmatrix}}$$

Trabajando con el modelo muestral podemos sustituir, entonces por :

$$Y = Xb$$

El propósito es resolver el sistema de tal forma que se obtengan los valores del vector b.

Es posible demostrar que:

$$Y = Xb$$

$$X'Y = X'Xb$$

$$X'Xb = X'Y$$

$$b = (X'X)^{-1}X'Y$$

Resolveremos el sistema con un ejemplo numérico a fin de clarificar el procedimiento de cálculo.

En la Dirección Agrícola Regional del Norte, en el año 1981, se realizó un experimento en el cual se evaluó el efecto del nitrógeno sobre la producción del maíz, en la localidad de La Zorroza, Morazán, Yoro.

Rendimiento Promedio en Ton/ha.	Kg. de Nitrógeno/ha.
5.85	160
5.28	120
5.22	80
4.63	80
4.29	80
3.78	80
3.78	80
3.78	80
3.35	40
2.28	0

Matricialmente, la situación se puede represetar por :

$$\underset{Y}{\begin{bmatrix} 5.85 \\ 5.28 \\ 5.22 \\ 4.63 \\ 4.29 \\ 3.78 \\ 3.78 \\ 3.35 \\ 2.28 \\ 1.85 \end{bmatrix}} = \underset{X}{\begin{bmatrix} 1 & 160 \\ 1 & 120 \\ 1 & 80 \\ 1 & 80 \\ 1 & 80 \\ 1 & 80 \\ 1 & 80 \\ 1 & 40 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}} \underset{b}{\begin{bmatrix} bo \\ b_1 \end{bmatrix}}$$

Si $b = (X'X)^{-1} X'Y$

Lo primero que haremos será obtener la matriz X

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 160 & 120 & 80 & 80 & 80 & 80 & 80 & 40 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 160 & 120 & 80 & 10 & 80 & 80 & 80 & 40 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 160 \\ 1 & 120 \\ 1 & 80 \\ 1 & 80 \\ 1 & 80 \\ 1 & 80 \\ 1 & 80 \\ 1 & 40 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

Por lo tanto : $(X'X) = \begin{bmatrix} 10 & 720 \\ 720 & 73600 \end{bmatrix}$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 160 & 120 & 80 & 80 & 80 & 80 & 80 & 40 & 0 & 0 \end{bmatrix} \begin{bmatrix} 5.85 \\ 5.28 \\ 5.22 \\ 4.63 \\ 4.29 \\ 3.78 \\ 3.78 \\ 3.35 \\ 2.28 \\ 1.85 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 40.31 \\ 3439.6 \end{bmatrix}$$

En consecuencia :

$$b = \begin{bmatrix} 10 & 720 \\ 720 & 73600 \end{bmatrix}^{-1} \begin{bmatrix} 40.31 \\ 3439.6 \end{bmatrix}$$

Recuerdese el sistema de ecuaciones normales :

$$b_o n + b_1 \Sigma X_i = \Sigma Y_i$$
$$b_o \Sigma X_i + b_1 \Sigma X_i^2 = \Sigma X_1 Y_1$$

Puede presentarse por :

$$\underbrace{\begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X^2_i \end{bmatrix}}_{X'X} \underbrace{\begin{bmatrix} bo \\ b_1 \end{bmatrix}}_{b} \underbrace{\begin{bmatrix} \Sigma X_i \\ \Sigma X_i Y_i \end{bmatrix}}_{X'Y}$$

$$\left|X'X\right| = 10\ (73600) - 720\ (720) = 736000 - 518400 = 217600$$

$$(X'X)^{-1} = \begin{bmatrix} 10 & 720 \\ 720 & 73600 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{73600}{217600} & \frac{-720}{217600} \\ \frac{-720}{217600} & \frac{10}{217600} \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 0.3382352 & -0.0033088 \\ -0.0033088 & 0.0000459 \end{bmatrix}$$

Entonces

$$b = \begin{bmatrix} 0.3382352 & -0.0033088 \\ -0.0033088 & 0.0000459 \end{bmatrix} \begin{bmatrix} 40.31 \\ 3439.6 \end{bmatrix} = \begin{bmatrix} 2.253312 \\ 0.0244999 \end{bmatrix}$$

Si $b_1 = 2.253312$, y $b_1 = 0.0244999$, entonces

$$\hat{Y} = 2.253312 + 0.0244999\ N$$

3.4 EL ANALISIS DE VARIANZA.

El análisis de varianza para la regresión también puede realizarse utilizando el algebra de matrices, el esqueña general para tal análisis es el siguiente:

Fuente de Variación	G.L.	Suma de Cuadrados	Cuadrado Medio	F
Regresión	1	$b'X'Y' - (\Sigma Y_i)^2/n$	$b'X'Y - (\Sigma Y_i)^2/n$	$\frac{CMR}{CME}$
Error	n-2	Diferencia	SCE/n-2	

Total corregido n-1 $\quad Y'Y - (\Sigma Y_i)^2/n$

Aplicaremos el esquema al ejemplo numérco anterior :

Suma de cuadrados debido a regresión :

$$SCR = b' X'Y - (\Sigma Y_i)^2/n$$

$$SCR = \begin{bmatrix} 2.253312 & 0.0244999 \end{bmatrix} \begin{bmatrix} 40.31 \\ 3439.6 \end{bmatrix} - (40.31)^2/10$$

$$SCR = 175.10086 - 162.48961 = 12.61125$$

Suma de cuadrados totales :

$$SCT = Y'Y - (\Sigma y_i)^2/n$$

$$SCT = 177.6105 - 162.48961 = 15.12089$$

Suma de cuadrados debido al error:

$$SCE = SCT - SCR$$

$$SCE = 15.12089 - 12.61125 = 2.50964$$

Entonces :

Fuente de Variación	G.L.	Suma de Cuadrados	Cuadrado medio	F
Regresión	1	12.61125	12.61125	40.2
Error	8	2.50964	0.313705	
TOTAL	9	15.12089		

R^2 = coeficiente de determinación

R^2 = SCR/SCT

R^2 = 12.61125 / 15.12084 = 0.834

R^2 = 83.4%

CV = coeficiente de variación = $\sqrt{\text{CM Error}}$ / $(\sum Y_i/n)$

Cv (100) = 13.89%

El R^2 mide "la proporción de la variación total observada en Y_i que es expresada o explicada por el modelo de regresión", es de desear que el valor de R^2 esté lo más próximo posible a 1 o a 100% si se le expresa en porcentaje. En nuestro caso el modelo de regresión logra explicar un 83.4% de la variación total observada en la variable dependiente.

ASUNCIONES BASICAS

Hasta este momento hemos procedido sin establecer requisitos al modelo de regresión lineal. Solo se han efectuado calculos aritméticos.. Sin embargo es conveniente establecer ciertas asunciones básicas del modelo de regresión.

Sea $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$; $1 = 1.2, \ldots\ldots, n$

1) ε_i es una variable aleatoria con media cero y varianza σ^2 común

$$E(\varepsilon_i) = 0$$
$$V(\varepsilon_i) = \sigma^2$$

2) Los errores no están correlacionados cuando $i \neq j$

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

Por lo tanto

$$E(y_i) = \beta_0 + \beta_1 X_i$$
$$V(Y_i) = \sigma^2$$

3) ε_i es una variable aleatoria normalmente distribuida, con media cero y varianza común, por la asunción 1) tenemos que :

$$\varepsilon_i \sim N(0, \sigma^2)$$

bajo esta condición los errores no solo no están correlacionados, sino que necesariamente son independientes. Además de lo anterior, se supone que la variable independiente X_i es medida sin error, sus valores son fijos.

R E F E R E N C I A S

1.- DRAPER, N.R. and H. Smith. "Applied regression analysis" John Wiley, 1966.

2.- DUNN, O.J. and V.A. CLARK. "Applied statistics: Analysis of variance and regression". John Wiley,1974.

3.- NETER, J. and W. WASSERMAN. "Applied liner statistical models". Homewood, U.S.A., Irwin, 1974.

4.- STEEL, R.G.D. and TORRIE. "Principles and procedures of statistics, a biometrical approach", 2nd edition, McGraw Hill, 1980.

4. ANALISIS DE REGRESION LINEAL MULTIPLE

F. Omar Osorio

La regresión de Y_i sobre una sola variable independiente, en ocasiones es inadecuada para explicar claramente la relación de dependencia buscada. Dos o más variables independientes pueden estar disponibles para proporcionar información adicional sobre las variaciones observadas en Y_i. La metodología a utilizar es una extensión del uso del modelo de regresión lineal simple, es evidente que el proceso aritmético se incrementará al aumentar el número de variables independientes.

Trabajaremos sobre dos variables independientes, la aplicación del método a más variables independientes es obvia.

MODELO GENERAL DE REGRESION LINEAL MULTIPLE.

Sea

$$Y_i = \beta_0 + \beta_1 X_{1_i} + \beta_2 X_{2_i} + \varepsilon_i \; ; \quad i = 1,2,\ldots, n$$

donde :

X_{1_i}, X_{2_i} = constantes conocidas

β_0, β_1 = parámetros del modelo

Y_i = variable aleatoria observable

ε_i = variable aleatoria no correlacionada, con media cero y varianza común.

En este modelo se requieren las mismas asunciones básicas que las establecidas para el modelo de regresión lineal simple.

Nuevamente requerimos estimar el valor de los parámetros del modelo, para ellos utilizaremos nuevamente el método de mínimos cuadrados, sea

$$Y_i = b_0 + b_1 X_{1_i} + b_2 X_{2_i}$$

El método de mínimos cuadrados se propone minimizar la suma de cuadrados de los errores $(\Sigma \varepsilon_i^2) = Q$

$$Q = \Sigma \varepsilon_i^2 = \Sigma (Y_i - b_0 - b_1 X_{1_i} - b_2 X_{2_i})^2$$

Para minimizar Q, se deriva parcialmente la función con relación a cada coeficiente de regresión y se igualan a cero las derivadas obtenidas :

$$\partial Q / \partial b_0 = 0$$
$$\partial Q / \partial b_1 = 0$$
$$\partial Q / \partial b_2 = 0$$

Lo cual dá origen al sistema de ecuaciones normales :

$$b_0 n + b_1 \Sigma X_{1_i} + b_2 \Sigma X_{2_i} = \Sigma X_i$$
$$b_0 \Sigma X_{1_i} + b_1 \Sigma X_{1i}^2 + b_2 \Sigma X_{1_i} X_{2_i} = \Sigma X_{1_i} Y_i$$
$$b_0 \Sigma X_{2_i} + b_1 \Sigma X_{2_i} X_{1_i} + b_2 \Sigma X_{2_i}^2 = \Sigma X_{2_i} Y_i$$

Resolviendo el sistema, se obtienen los valores de b_0 , b_1 y b_2, este sistema puede resolverse mediante matrices, si :

$$X'Xb = X'Y$$
$$b = (X'X)^{-1} X'Y$$

$$\underbrace{\begin{bmatrix} n & X_{1_i} & X_{2_i} \\ X_{1_i} & X_{1_i}^2 & X_{1_i} X_{2_i} \\ X_{2_i} & X_{2_i} X_{1_i} & X_{2_i}^2 \end{bmatrix}}_{X'X} \quad \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}}_{b} \quad \underbrace{\begin{bmatrix} \Sigma Y_i \\ \Sigma X_{1_i} Y_i \\ \Sigma X_{2_i} Y_i \end{bmatrix}}_{X'Y}$$

La solución es única si $\left|X'X\right| \neq 0$

Como ejemplo, utilizaremos los resultados obtenidos en un experimento de campo realizado por la Dirección Agrícola Regional del Litoral Atlántico, con maíz, en la localidad de Brisas de San Pedro, Bajo Aguán, en el Semestre 1980-B.

Rendimiento en Ton/ha. Y	Kg. de N/ha.	N^2
4.01	0	0
3.84	0	0
4.64	40	1600
5.50	80	6400
5.75	120	14400
6.23	160	25600
5.97	80	6400
5.56	80	6400
5.78	80	6400
5.62	80	6400

En consideración a el hecho conocido de que en una cierta etapa de las funciones de producción agrícola los incrementos en los rendimientos son decrecientes, pudiendo los mismo describirse en forma gráfica a través de una parábola, parece razonable proponer como modelo de ajuste una ecuación de regresión de segundo grado del tipo

$$\hat{Y}_i = bo + b_1 N + b_2 N^2$$

Por lo tanto, en forma matricial tendríamos:

$$\underset{Y}{\begin{bmatrix} 4.01 \\ 3.84 \\ 4.64 \\ 5.50 \\ 5.75 \\ 6.23 \\ 5.97 \\ 5.56 \\ 5.78 \\ 5.62 \end{bmatrix}} = \underset{X}{\begin{array}{c} \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 40 & 1600 \\ 1 & 80 & 6400 \\ 1 & 120 & 14400 \\ 1 & 160 & 25600 \\ 1 & 80 & 6400 \\ 1 & 80 & 6400 \\ 1 & 80 & 6400 \\ 1 & 80 & 6400 \end{bmatrix} \end{array}} \begin{array}{c} N \quad N \\ \\ \end{array} \underset{b}{\begin{bmatrix} b_o \\ b1 \\ b_2 \end{bmatrix}}$$

Entonces si

$$Y = Xb$$
$$X'Y = X'Xb$$
$$X'Xb = X'Y$$
$$b = (X'X)^{-1}X'Y$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 40 & 80 & 120 & 160 & 80 & 80 & 80 & 80 \\ 0 & 0 & 160 & 6400 & 14400 & 25600 & 6400 & 6400 & 6400 & 6400 \end{bmatrix}$$

$$(X'X) = \begin{bmatrix} 10 & 720 & 73600 \\ 720 & 73600 & 8448000 \\ 73600 & 8448000 & 1070080000 \end{bmatrix} \quad ; \quad X'Y = \begin{bmatrix} 52.9 \\ 4246.8 \\ 431664 \end{bmatrix}$$

$$|X'X| = 1.5826944 \times 10^{13}$$

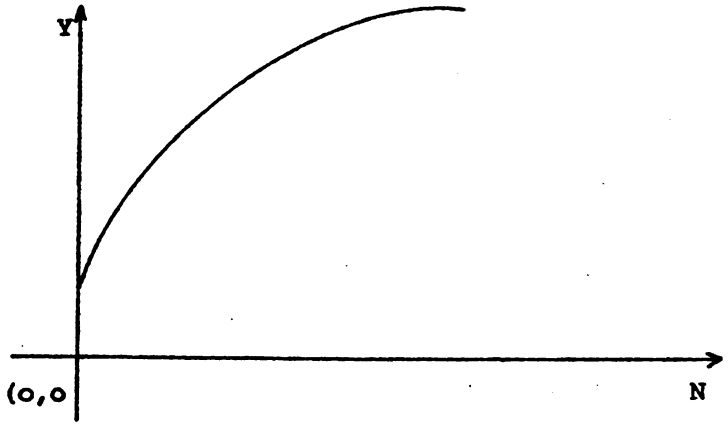
$$(X'X)^{-1} = \begin{bmatrix} 0.466873706 & -0.0093944099 & 0.0000420549 \\ -0.0093944099 & 0.0003338509 & -0.0000019895 \\ 0.0000420549 & -0.0000019895 & 0.0000000137 \end{bmatrix}$$

$$b = (X'X)^{-1}X'Y = \begin{bmatrix} 3.894451346 \\ 0.0286451863 \\ -0.0000906121 \end{bmatrix} \quad ; \quad \begin{aligned} b_o &= 3.894451346 \\ b_1 &= 0.0286451863 \\ b_2 &= 0.0000906121 \end{aligned}$$

Por lo tanto, la ecuación de regresión queda como sigue :

$$\hat{Y} = 3.89445 + 0.028645\ N - 0.0000906121N^2$$

La ecuación corresponde a una línea de regresión similar a :



Nos interesa evaluar la Ho: $B_1 = B_2 = 0$ contra $\beta_1 \neq \beta_2 \neq 0$, tal prueba de hipótesis se hace a través del análisis de varianza para la regresión , el cuadro de tal análisis para el caso de P variables independientes es:

Fuentes de Variación	G.L.	Suma de Cuadrados	Cuadrado Medio	F
Regresión	P	$b'X'Y - (\Sigma Y_i)^2/n$	$\frac{b'X'Y - (\Sigma Y_i)^2/n}{P}$	$\frac{CMR}{CME}$
Error	n-1-p	Diferencia	SCE/n-1-p	
Total Corregido	N-1	$Y'Y - (\Sigma Y_i)^2/n$		

De tal manera que:

Suma de cuadrados debido a regresión :

$SCR = b'X'Y - (\Sigma Y_i)^2/n$

$$SCR = \begin{bmatrix} 3.894451346 & 0.0286451863 & -0.0000906121 \end{bmatrix} \begin{bmatrix} 52.9 \\ 4146.8 \\ 431664 \end{bmatrix} - (52.9)^2/10$$

SCR = 285.6883532 - 279.841 = 5.8473532

Suma de cuadrados totales :

$SCT = Y'Y - (\sum Y_i)^2/n$

SCT = 286.028.-279.841 = 6.187

Sumas de cuadrados del Error :

SCE = SCT-SCR

SCE = 6.187-5.8473532 = 0.3396468

Por lo tanto el cuadro del análisis de varianza quedará como sigue :

Fuente de Varianza	G.L.	Suma de Cuadrados	Cuadrado Medio	Fc
Regresión	2	5.8473532	2.9236766	60.2559
Error	7	0.3396468	0.0485209	
Total Corregido	9	6.187		

La relación calculada de Fc se compara con una valor de Ft en tablas con, 2 y 7 grados de libertad, al valor de $\alpha = 0.01$ el valor de tales tablas es de 9.55; por cuanto Fc > Ft, se rechaza la hipótesis nula, aceptándose la H_1; por lo tanto aceptamos que los coeficientes de regresión son diferentes de cero. Otros criterios para juzgar el modelo de regresión propuesto son: el R^2.

R^2 = SCR/SCT = 5.8473532/6.187 = 0.945

R^2 (100) = 94.5%

El R^2 nos indica que la ecuación de regresión múltiple ajustada permite explicar un 94.5% de la variación total observada en la variable dependiente.

Es posible sentirse satisfecho de la ecuación de regresión ajustada, sin embargo aún es posible obtener más información sobre la misma.

Es de interés evaluar en forma individual los coeficientes de regresión b_1 y b_2, para ello se recurre a la prueba de t, de esta manera se hace una prueba de hipótesis para cada coeficiente :

Ho : $b_i = 0$ contra H_1 : $b_i \neq 0$

La prueba se hace calculando un valor de t tal que:

$$t_c = \frac{b_i - 0}{s_{bi}}$$

El valor de s_{bi} (error estandar de b_i) se calcula utilizando los elementos de la diagonal de la matríz $(X'X)^{-1}$, sea

$$(X'X)^{-1} = \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{bmatrix}$$

Error estandar de b_1

$$s_{bi} = \sqrt{\text{CM Error } (\omega_{22})}$$

$$s_{bi} = \sqrt{0.0485209 \ (0.0003338509)} = 0.0040247666$$

Error estandar de b_2

$$s_{b_2} = \sqrt{\text{CM Error } (\omega_{33})}$$

$$s_{b_2} = \sqrt{0.0485209 \ (0.0000000137)} = 0.0000257825$$

Por lo tanto :

tc para b_1 es : $Tc = \dfrac{b_1 - 0}{s_{b_1}}$

$$tc = \frac{0.0286451863 - 0}{0.0040247666} = 7.117$$

tc para b_2 es : $Tc = \dfrac{b_2 - 0}{s_{b_2}}$

$$tc = \frac{-0.0000906121-0}{0.0000257825} = -3.514$$

Ambos valores de Tc son comparados con valores de t tabulados con n-p grados de libertad, en nuestro caso 10-2 = 8 grados de libertad, t_c valor con un $\alpha = 0.01$ de 3.355, como vemos en ambos casos $|tc| > Tt$ por lo tanto se rechaza la Ho y aceptamos la H_1 ; $b_i \neq 0$ lo cual significa que ambos coeficientes son altamente significativos, en terminos de nuestro e - jemplo eso significa que tanto el efecto lineal como el efecto cuadrático del nitrógeno son altamente significativos. Todos los resultados obtenidos permiten ganar confianza en el uso de la ecuación de regresión ajustada, como modelo que permite explicar la relación funcional del nitrógeno aplicado al cultivo y la respuesta obtenida por este.

La aplicación del modelo de regresión lineal múltiple a más de dos variables (pueden ser dos o más factores productivos) solo es una ampliación de lo aquí expuesto, obviamente ello obliga el manejo de matrices de mayor dimensión, sin embargo calculadoras manuales ya poseen rutinas de inversión de matrices, en algunos casos (quizá la mayoría) se requerirá recurrir a un sistema de procesamiento de datos con la finalidad de ganar tanto tiempo como seguridad en los resultados.

R E F E R E N C I A

1.- DAPER, N.R. and H. SMITH. "Applied regression analysis"., John Wiley, 1966.

2.- DUNN, O.J. and V.A. CLARK. "Applied statistics: analysis of variance and regression". John Wiley, 1974.

3.- NETER, J. and W. WASSERMAN. "Applied linear statistical models" Homewood, U.S.A. Irwin, 1974.

4.- OSTLE, B." Estadística aplicada," Mexico, Limusa-Wiley, 1970.

5.- STEEL, R.G.D. and J.H. TORRIE. "Principles and procedures of statistic, a Biometrical approach". 2 ed Mc. Graw-Hill, 1980.

5. ANALISIS ECONOMICO DE EXPERIMENTOS CON FERTILIZANTES.

F. Omar Osorio

5.1 Conceptos sobre metodología de superficie de respuestas

La metodología de superficie de respuestas (MSR), puede conceptulizarse como un conjunto de técnicas estadístico-matemáticas, cuyo objetivo es determinar las mejores condiciones en que opera un proceso para lograr resultados óptimos, lo que a su vez permite - un mejor conocimiento de la naturaleza del sistema de operación - del proceso.

El proceso está regido por leyes que gobiernan una relación causa-efecto, en general involucra una variable de respuesta o resultado la cual se expresa como una función de factores causales denominados variables independientes.

La metodología de superficie de respuestas ha sido más utilizada en la ingeniería de procesos industriales, especialmente en el campo de la quimica, después se ha llegado a utilizar en diversos campos de la investigación científica, encontrándosele muy útiles aplicaciones en la investigación agronómica. En este campo el investigador se interesa en obtener un mejor conocimiento de las relaciones existentes entre productos e insumos, ya que este conocimiento le permite enterarse no solo del tipo de relación funcional existente entre los factores y el producto, sino porque además puede utilizarse en la finalidad de estimar el nivel de utilización de los factores

productivos con criterio tanto productivo como económico.

Para lo anterior el investigador debe recurrir a diversas técnicas que le permitan determinar las condiciones óptimas del proceso desde el punto de vista económico, en ese sentido, por lo general, el investigador agrícola está interesado en desarrollar recomendaciones cuya utilización por parte de los productores les proporcione un retorno económico (beneficio), que haga estimulante la actividad productiva.

5.2 Estimación de óptimos económicos bajo condición de capital ilimitado.

En agricultura el objetivo final del productor, generalmente, no es obtener la máxima producción posible, es posible que bajo la condición de capital ilimitado su objetivo sea maximizar sus beneficios económicos, tal situación se dará, entonces, cuando su ingreso neto se vea maximizado.

Sea

$$IN = IB - CF - CV$$

Ingreso Neto = Ingreso Bruto - Costos fijos - Costos Variables

donde:

$IB = YPy$

py = precio del producto

$CV = \Sigma X_i Px_i$

Px_i = precio del i-ésimo factor

por tanto

$$IN = YPy - CF - \Sigma X_i Px_i$$

Puesto que nuestro objetivo es maximizar la función de ingreso neto, igualamos a cero las primeras derivadas de la función con relación a los factores productivos X_i.

$$d\,IN = \partial\hat{Y}Py - 0 - \partial\Sigma X_i Px_i = 0$$

$$\partial\hat{Y}Py = \partial X_i Px_i$$

$$\partial Y/\partial X_i = Px_i/Py \qquad i = 1,2,\ldots,n$$

donde n es igual al número de factores productivos. Lo anterior significa que el ingreso neto se maximiza cuando la derivada de la respuesta $\hat{Y}$ con respecto a cada uno de los insumos X_i (producto marginal), se iguala a la relación inversa de precios Px_i/Py

Ejemplo :

Experimentos de campo realizados por la Dirección Agrícola Regional del Litoral Atlántico, con maíz, en la localidad Brisas de San Pedro en el Bajo Aguán en el semestre 1980-B, analizados adecuadamente, permitieron estimar la función de producción siguiente :

$$\hat{Y} = 3.8945 + 0.028645N - 0.00000912N^2$$

Considerando los siguientes precios supuestos :

Costo de 1 kg de N = Lps 1.28

Valor de 1 tonelada de maíz = Lps 308.oo

De acuerdo al método :

Optimo económico : $dY/dN = Pn/Py$

$dY/dN = 0.028645-0.000181224N$

$Pn/Py = 1.28/308.0 = 0.0041558$

$0.028645 - 0.000181224\,N = 0.0041558$

$-\,0.00018224\,N = 0.024489$

$N = -0.024489/-0.00018224$

$N = 134.4$ Kg. N

Utilizando esta dosis de fertilización la producción física esperada se obtiene sustituyendo el valor óptimo en la función de producción:

$Y = 3.8945 + 0.02645\ (1344) - 0.000090612\ (134.4)^2$

Y = 6. 1 Ton de maíz por hectárea.

El procedimiento es simple cuando solo se involucra un factor productivo, sin embargo, en la mayoría de las situaciónes, el experimentador obtiene funciones de producción que involucran de dos a tres factores productivos, en tales casos, durante el proceso de cálculo del óptimo económico, se obtiene un sistema de ecuaciones simultáneas con 2 ó 3 incognitas; con la finalidad de simplificar el cálculo se recurre al algebra de matrices. Ejemplificaremos el uso del algebra de matrices aplicándolo a los resultados de un experimento con frijol realizado por el PROMIF en 1978 en Danlí.

Las ecuaciones de regresión o función de producción estimada fué :

$\hat{Y} = 325.01 + 38.978\ N + 45.887\ P - 0.89625N^2 + 0.266NP$

$R^2 = 95.5\%$

Todos los coeficientes fueron altamente significativos

Costo de 1 Kg de N= Lps 1.28 = Pn

Costo de 1 Kg de P= Lps 1.05 = Pp

Precio de 1 Kg de frijol = Lps 0.40 Py

Para encontrar el nivel de uso óptimo para los factores nitrógeno y fósforo, se procede como sigue :

$$\frac{\partial Y}{\partial N} = \frac{Pn}{Py}$$

$$\frac{\partial Y}{\partial P} = \frac{Pp}{Py}$$

$$\frac{\partial Y}{\partial N} = 38.978 - 1.7925\ N + 0.2666\ P = Pn/Py = 3.2$$

$$\frac{\partial Y}{\partial P} = 45.887 - 1.20812P + 0.2666\ N = Pp/Py = 2.625$$

$$38.978 - 1.7925N + 0.2666P = 3.2$$

$$45.887 + 0.2666\ N - 1.20812P = 2.625$$

$$-1.7925\ N + 0.266\ P = -35.778$$

$$+0.2666\ N + 1.20812\ P = -43.262$$

El sistema de ecuaciones puede ordenarse en forma matricial :

$$\underset{A}{\begin{bmatrix} -1.7925 & 0.2666 \\ 0.2666 & -1.20812 \end{bmatrix}} \quad \underset{X}{\begin{bmatrix} N \\ P \end{bmatrix}} = \underset{b}{\begin{bmatrix} -35.778 \\ -43.262 \end{bmatrix}}$$

AX = b, este sistema puede resolverse mediante la regla de Cramer o invirtiendo la matriz A.

Procedermos de acuerdo a la regla de Cramer, por lo tanto :

$$|A| = -1.7925\ (-1.20812) - 0.2666\ (0.2666)$$

$$|A| = 2.1655551 - 0.0710755 = 2.0944796$$

$$N = \frac{\begin{vmatrix} -35.778 & 0.2666 \\ -43.262 & -1.20812 \end{vmatrix}}{|A|} = \frac{54.757766}{2.0944796} = 26.14$$

$$P = \frac{\begin{vmatrix} -1.7925 & -35.778 \\ 0.2666 & -43.262 \end{vmatrix}}{|A|} = \frac{87.085549}{2.0944796} = 41.58$$

De acuerdo a los cálculos los niveles óptimos son:26. Kg de nitrógeno por hectárea y 41 Kg. de fósforo por hectárea.

5.3 Estimación de óptimos económicos bajo condición de capital limitado.

En la realidad la condición de capital ilimitado se presenta en muy pocos casos, en nuestro país la generalidad de los productores agrícolas tienen una condición de recursos económicos limitados, esto es cierto aún para aquellos agricultores indiependientes y asociados que son sujetos que participan de los beneficios del sistema de crédito.

Por lo anterior se considera necesario que el investigador maneje algunas técnicas que le permitan estimar las condiciones óptimas de trabajo bajo la situación de capital limitado.

Consideramos que existe una condición de capital limitado, cuando el capital disponible para cubrir en forma óptima los costos variables sea inferior al que se requeriría si se trabajase bajo la condición de capital ilimitado. Esto significa que el primer paso será estimar el óptimo económico para la supuesta condición de capital ilimitado, si los valores óptimos encontrados suman un valor mayor que el capital disponible por el productor, pasaríamos a estimar nuevamente óptimos económicos restringiendo los mismos al capital disponible.

Con el criterio de capital limitado, los óptimos económicos de los insumos variables se obtienen en el punto en donde se iguala el valor del producto marginal de cada insumo por Lempira invertido en cada uno de ellos, bajo la restricción de que la inversión en insumos sea igual a un capital determinado.

Así la estimación de las cantidades óptimas de los insumos, es similar a la que se hace para capital ilimitado, pero incluyendo la restricción de capital limitado y utilizando multiplicadores de Lagrange, de esta manera, la función de ingreso neto sería :

$$IN = YPy - CF - \Sigma X_i Px_i + \lambda (\Sigma Px_i X_i - C)$$

donde: λ = multiplicador de Lagrange

C = capital limitado

Si en la anterior expresión, se expresa Y como su función de producción $\hat{Y}$, se deriva con respecto a cada X_i y λ, se iguala cada derivada a cero y se resuelve el sistema de ecuaciones generado, se obtienen las cantidades óptimas de cada insumo para una condición de capital limitado dado.

Se ejemplificará el uso de la ténica, aplicandola a la función de producción obtenida para el frijol por el PROMYF.

Sea

$\hat{Y} = 325.01 + 38.978N + 45.887P - 0.89625N^2 + 0.2666NP$

y Pn = 1.28

Pp = 1.05

Py = 0.40

C = 55.00 Lempiras

Al optimizar esta función de producción para condiciones de capital ilimitado se determino que para tal condición se requería invertir Lps. 77.11,considerando que solo se dispone de Lps. 55.00, se procederá a optimizar la función de producción para esta condición de

capital limitado.

Por tanto

$$\partial IN/\partial N = Py\ (\partial Y/\partial N) - Pn + \lambda Pn = 0$$

$$\partial IN/\partial P = Py\ (\partial Y/\partial P) - Pp + \lambda Pp = 0$$

$$\partial IN/\partial \lambda = PnN + PpP - C = 0$$

sustituyendo adecuadamente,

$$0.4(38.978 - 1,7925N + 0.2666P) - 1.28 + \lambda\ 1.28 = 0$$

$$0.4(45.887 + 0.2666N - 1.20812P) - 1.05 + \lambda 1.05 = 0$$

$$1.28\ N + 1.05P - 55 = 0$$

por tanto

$$15.5912 - 0.717N + 0.10664P - 1.28 + 1.28\lambda = 0$$

$$18.3548 + 0.10664N - 0.483248P - 1.05 + 1.05\lambda = 0$$

$$1.28N + 1.05P - 55 = 0$$

y,

$$-0.717N + 0.10664P + 1.28\lambda = -14.3112$$

$$0.10664N - 0.483248P + 1.05\lambda = -17.3048$$

$$1.28N + 1.05P - 0\lambda = 55$$

Este sistema de ecuaciones podemos representarlo en forma matricial:

$$\underbrace{\begin{bmatrix} -0.717 & 0.10664 & 1.28 \\ 0.10664 & -0.483248 & 1.05 \\ 1.28 & 1.05 & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} N \\ P \\ \lambda \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} -14.3112 \\ -17.4048 \\ 55.0 \end{bmatrix}}_{b}$$

Por lo tanto, de acuerdo al álgebra de matrices, Si $|A| \neq 0$, el siste-

ma tiene solución única por :

$$x = A^{-1}b$$

Procediendo correctamente se encuentra que $|A| = 1.868894343$, y que

$$x = \begin{bmatrix} 17.49674406 \\ 31.05158819 \\ -3.96669901 \end{bmatrix}$$

Por lo tanto las dósis óptimas de uso de ambos insumos bajo la condición de capital limitado (L.55.00), es de 17.5 y 31 Kg de nitrógeno y fósforo por hectárea respectivamente, el valor de multiplicador de Lagrange λ, mide la cantidad en que se ha disminuido la función objetivo (IN) al reducir el capital disponible C desde su valor óptimo ilimitado (L.77.11), hasta su condición de capital limitado de L.55.00.

Lo anterior significa que disponiendo de sólamente L. 55.00, la mejor combinación de insumos posible de uso por el productor es de 17.5 Kg. de nitrógeno y 31 Kg. de fósforo por hectárea, dicha combinación maximizará su ingreso neto para el capital disponible; cualquier otra combinación de insumos no podría arrojar igual o mayor ingreso neto.

5.4 OBSERVACIONES

El procedimiento reseñado provoca comentarios muy diversos, algunos a favor y otros contrarios al uso del mismo.

Puede decirse que en la interpretación de los óptimos económicos obtenidos, debe tenerse en cuenta la etapa de la investigación en la cual se les está obteniendo.

Los óptimos económicos obtenidos a partir de ensayos exploratorios no deberían, en la mayoría de los casos, ser utilizados para elaborar recomendaciones para los productores, más bien debe utilizarseles con la finalidad de mejorar los criterios de selección del espacio factorial en los experimentos que se planifiquen en el futuro. En este sentido no debe extrañar el que se obtengan óptimos económicos muy altos, pues esto ocurre, generalmente, como una consecuencia del diseño de los tratamientos, un espaciamiento inadecuado y niveles no convenientes pueden ser la causa de tales resultados.

En experimentos más avanzados al recomendarse a los productores el uso de determinados nivels de fertilización, independientemente del método de estimación económica utilizado, debe considerarse cuidadosamente el nivel de tecnificación del productor, sistema productivo y los recursos disponibles, es decir, no debe generalizarse un resultado aritmético ya que existen factores que de no tomarse en consideración pueden ocasionar resultados negativos.

Siempre que se desee elaborar recomendaciones para los agricultores en base a este método, debe considerarse un aspecto de importancia:

Las funciones de producción utilizadas en el método, son ecuaciones de regresión, es decir corresponden a modelos probabilísticos, en los cuales existe cierto elemento de incertidumbre (aún cuando exista un grado de ajuste muy alto) la ecuación representa el valor promedio, en este contexto debemos entender los valores óptimos cual si fueran estimaciones puntuales, lo cual permite establecer límites de confianza a tales estiamaciones. Lo anterior significa que el valor óptimo a traves de una estimación por intervalo (limites de confianza) permitirá obtener un rango de valores antre los cuales el experimentador buscará, usando varios criterios relacionados, la recomendación más adecuada al productor.

R E F E R E N C I A S

1.- ALDUNATE, P. "Análisis de productividad en parcelas CORA y fundos." Universidad Católica de Santiago de Chile, S.F.

2.- DOLL, J.P. and F. ORAZEM. " roduction economics; theory with application". Grid Ind., Columbus, Ohio, 1978.

3.- GOMEZ H., J. Et al. "Aplicaciones agronómicas de la metodología de superficie de respuestas". Agrociencia No. 32, pp 125-135, Chapingo, México, 1978.

4.- VOLKE H., V. "El uso del criterio económico de capital limitado para optimizar niveles de insumos en la agricultura de temporal-subsistencia". Agrociencia No. 36, pp 121-136, Chapingo, México 1979.

5.- VOLKE H., V y M. Villa I. "Criterios económicos para optimizar niveles de insumos variables bajo condiciones de agricultura de temporal-subsistencia". Agrociencia No. 36, pp 107-119, Chapingo, Mexico, 1979.

6.- MARTINEZ G.,A."Aspectos económicos del diseño y análisis de experimentación". Chapingo, Mexico. Colegio de Postgraduados 1972.

A P E N D I C E

ALGEBRA DE MATRICES

6. ALGEBRA DE MATRICES

F. Omar Osorio

Definición : Una matriz es un arreglo rectangular de elementos distribuidos en hileras y columnas. La dimensión de una matriz viene dada por el número de hileras y columnas que tiene.

Ejemplo :

$$\begin{array}{l} \\ \text{Hilera 1} \\ \text{Hilera 2} \\ \text{Hilera 3} \end{array} \begin{array}{c} \begin{array}{cc} \text{Columna 1} & \text{Columna 2} \end{array} \\ \begin{bmatrix} 6000 & 23 \\ 13000 & 47 \\ 11000 & 35 \end{bmatrix} \end{array}$$

Los elementos de esta matriz en particular son números que representan el ingreso (columna 1) y la edad (columna 2)de 3 personas (hileras). Así el elemento en la primera hilera y la primera columna (6000) representa el - ingreso de la primera persona. La dimensión de esta matriz es de 3X2, esto es, 3 hileras por 2 columnas.

Otros ejemplos de matrices son :

$$\begin{bmatrix} 1 & 0 \\ 5 & 10 \end{bmatrix} \qquad \begin{bmatrix} 4 & 7 & 12 & 16 \\ 3 & 15 & 9 & 8 \end{bmatrix}$$

Estas matrices tienen dimensiones 2 X 2 y 2 X 4 respectivamente, Nótese que siempre especificamos primero el número de hilera y luego el de las columnas, al dar las dimensiones de una matriz.

Tal como en el álgebra común, se puede utilizar simbolos para identificar los elementos de una matriz :

$$\begin{array}{c} \\ i = 1 \\ i = 2 \end{array} \begin{array}{c} \begin{array}{ccc} j = 1 & j = 2 & j = 3 \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \end{array}$$

Observese que el primer subindice indica el número de la hilera y el segundo el número de la columna. Así, es posible usar la notación aij para el elemento de la i-ésima hilera y j-ésima columna.

Usualmente se utilizan letras en negrilla para representar las matrices, designando los vectores por letras minúsculas, ejemplo :

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

Esta matriz puede representarse por :

$$\mathbf{A} = aij \; ; \; i = 1.2 \quad j = 1,2,3$$

Lo cual ahorra tener que escribir todos los elementos, describiéndose solamente el elemento general.

Comentarios :

1) No piense en una matriz como si fuera un número. Es un conjunto de elementos colocados en un arreglo rectangular, cuando la matriz es de dimensión 1X1, se tiene una matriz de un solo número.

2) Lo siguiente no es una matriz :

$$\begin{bmatrix} & 14 & \\ 8 & & 15 \\ & & 16 \end{bmatrix}$$

Dado que los elementos no están ordenados en hileras y columnas

MATRIZ CUADRADA.

Se dice que una matriz es cuadrada si el número de hileras es igual al número de columnas.

Ejemplo :

$$\begin{bmatrix} 4 & 7 \\ 3 & 9 \end{bmatrix} \qquad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

VECTOR.

Una matriz con una sola columna se denomina vector.

$$a = \begin{bmatrix} 4 \\ 7 \\ 10 \end{bmatrix} \qquad c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}$$

Así, el vector a, es una matriz 3 X 4, y el vector c, es una matriz 4 X 1.
Una matriz que solo tiene una hilera se denomina vector hilera:

$$b' = \begin{bmatrix} 15 & 25 & 50 \end{bmatrix} \quad d' = \begin{bmatrix} d_1 & d_2 \end{bmatrix}$$

Usamos el símbolo (') prima por razones que luego veremos. Nótese que el vector b' es una matriz 1 X 3 y que el vector d' es una matriz 1 X 2. Un subíndice es suficiente para identificar los elementos de un vector.

TRANSPUESTA DE UNA MATRIZ.

Transponer una matriz significa intercambiar sus hileras y columnas. La transpuesta de una matriz es a menudo un paso requerido en operaciones matriciales más complejas. Si una matriz A es de orden m X n, entonces la transpuesta que se denota por A' será de orden n X m.

Sea :

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 1 & 4 \end{bmatrix} \text{, entonces } A' = \begin{bmatrix} 2 & 1 \\ 0 & 1 \\ 1 & 4 \end{bmatrix}$$

Nótese que la hilera 1 de A es la columna 1 de A', y la hilera 2 de A es la columna 2 de A'.

TIPOS ESPECIALES DE MATRICES.

Ciertos tipos especiales de matrices se presentan frecuentemente en las operaciones matriciales. Consideremos las más importantes :

MATRIZ SIMETRICA.

Una matriz simétirca es una matriz cuadrada cuyos elementos que están abajo de la diagonal principal son iguales a los correspondientes elementos

que están arriba de la diagonal. La diagonal principal es aquella en donde los elementos tiene i = j.

Ejemplo:

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix} \qquad A' = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}$$

Nótese que cuando una matriz es simétrica $A = A'$

MATRIZ DIAGONAL.

Una matriz diagonal A es una matriz cuadrada cuyos elementos diferentes a la diagonal principal, son todos cero:

$$A = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \qquad B = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10 \end{bmatrix}$$

Dos tipos importantes de matrices diagonales son :

a) Matriz Idéntica : Se denota por I, es una matriz cuadrada en la que todos los elementos de la diagonal principal son unos (1) y el resto de elementos son ceros (0). Esta matriz toma el lugar de el uno en la aritmética ordinaria.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Para cualquier matriz cuadrada A, $AI = IA = A$

$$IA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & a \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Así, la matriz idéntica puede insertarse o quitarse de cualquier expresión matricial cuando quiera que sea conveniente.

b) Matriz Escalar : es una matriz cuadrada en la que los elementos de la diagonal principal son cantidades escalares idénticas, y el restro de los elementos son ceros :

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

$$\lambda I = 2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

MATRIZ NULA.

Es aquella en la que todos sus elementos son ceros, asi mismo los vectores en los que todos sus elementos son ceros se denominan vector cero,

DETERMINANTES.

Un concepto importante de una matriz cuadrada es su determinante. Se puede conceptualizar el determinante de una matriz cuadadra como un escalar asociado a la matriz de acuerdo a una regla definida:

Consideremos la matriz A (2 X 2)

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

El determinante de A se denota por A

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}\ a_{22} - a_{12}\ a_{21}$$

Numericamente

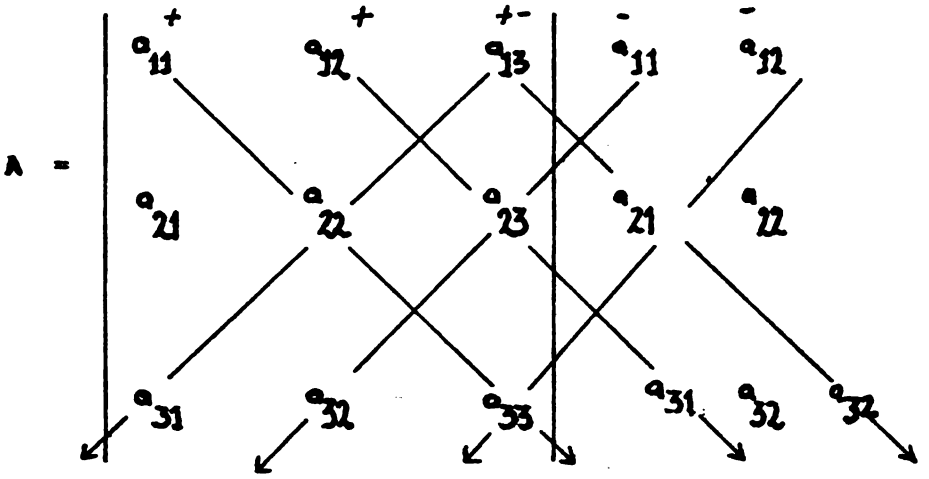
$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \text{ entonces } |A| = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} = 1(3)-2(2)=-1$$

Este procedimiento solo se aplica a matrices 2 X 2. Si se trata de una matriz más grande, se puede recurrir a la solución denominada "expansión por cofactores", su aplicación puede llegar a ser problemática en el caso de matrices muy grandes. Una solución muy simple se puede obtener con una ampliación del método utilizado en el caso de matrices 2 X 2.

Sea A una matriz 3 X 3, entonces

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\ a_{22}\ a_{33} + a_{12}\ a_{23}\ a_{31} + a_{13}\ a_{21}\ a_{32} - a_{13}\ a_{22}\ a_{31} - a_{11}\ a_{23}\ a_{32} - a_{12}\ a_{21}\ a_{33}$$

Esquemáticamente, la solución anterior se puede representar por :



En el esquema las dos columnas de la derecha se han repetido para facilitar las operaciones.

Ejemplo:

$$|A| = \begin{bmatrix} 6 & 2 & 1 \\ 3 & 0 & 1 \\ 8 & 2 & 3 \end{bmatrix} = 6(0)\ 3 + 2(1)8 + 1\ (3)\ 2 - 1(0)\ 8 - 6(1)2 - 2\ (3)\ 3 = 8$$

El determinante de una matriz será cero si 1) todos los elementos de cualquier hilera o cualquier columna son ceros, 2) cualesquera dos hileras o dos columnas son idénticas, 3) si cualquiera hilera es un multiplo de otra hilera o si una columna es un múltiplo de otra columna, y 4) una hilera o una columna puede ser expresada como una combinación lineal de dos o más hileras o columnas respectivamente.

Se dice que una matriz es <u>singular</u> o <u>degenerada</u> si su determinante es cero.

SUMA Y RESTA DE MATRICES.

Para sumar y restar matrices se requiere que tengan igual dimensión. La suma o diferencia de dos matrices es otra matriz cuyos elementos consisten de la suma o diferencia de los correspondientes elementos de las dos matrices.

Supongamos que:

$$\underset{3 \times 2}{A} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \qquad \underset{3 \times 2}{B} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}$$

Entonces

$$A + B \quad \begin{bmatrix} 1+1 & 4+2 \\ 2+2 & 5+3 \\ 3+3 & 6+4 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 4 & 8 \\ 6 & 10 \end{bmatrix}$$

De igual manera

$$A-B = \begin{bmatrix} 1-1 & 4-2 \\ 2-2 & 5-3 \\ 3-3 & 6-4 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 0 & 2 \end{bmatrix}$$

Observese que A + B = B + A como en el algebra ordinaria.

MULTIPLICACION DE MATRICES.

Para multiplicar matrices, los elementos de una hilera de la primera matriz se multiplican por los correspondientes elementos de una colum-

na de la segunda matriz - Sea

$$\underset{2 \times 2}{A} = \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \qquad \underset{2 \times 2}{B} = \begin{bmatrix} 6 & 3 \\ 5 & 4 \end{bmatrix}$$

$$AB = \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} 6 & 3 \\ 5 & 4 \end{bmatrix} = \begin{bmatrix} 2(6) + 0(5) & 2(3) + 0(4) \\ 5(6) + 1(5) & 5(3) + 1(4) \end{bmatrix}$$

$$\underset{2 \times 2}{AB} = \begin{bmatrix} 12 & 6 \\ 35 & 19 \end{bmatrix}$$

Otro ejemplo :

$$\underset{2 \times 3}{A} = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 5 & 8 \end{bmatrix} \qquad \underset{3 \times 1}{B} = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}$$

$$\underset{2 \times 1}{AB} = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 5 & 8 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 26 \\ 41 \end{bmatrix}$$

En el proceso de obtención del producto **AB**, decimos que postmultiplicamos **A** por **B**, o que premultiplicamos **B** por **A**. La razón de esta terminología se

Se debe a que las reglas de multiplicación del algebra ordinaria no se aplican a la multiplicación de matrices, por lo tanto **AB** ≠ **BA**. En efecto, puede ocurrir que el producto AB se logre definir mas no el producto BA. En general el producto AB queda definido solo cuando el número de columnas de A es igual al número de hileras de B. En los ejemplos previos :

A	**B**	**AB**	**A**	**B**		**AB**
2 X 2	2 X 2	2 X 2	2 X 3	3 X 1	=	2 X 1

dimensión del producto dimensión del producto

Nótese que el número de hileras de **A** y el número de columnas de **B** definen la dimensión del producto **AB**; además observese que en el segundo ejemplo el producto **BA** no se podría definir por cuanto el número de columnas de **B** no es igual al número de hileras de **A**.

INVERSA DE UNA MATRIZ.

En algebra ordinaria, el inverso de un número es su recíproco. Así, el inverso de 6 es 1/6. Un número multiplicado por su inverso siempre es igual a 1

$$6\ (1/6) = 1$$

En algebra matricial, la inversa de una matriz cuadrada A, es otra matriz cuadrada que produce la matriz identida cuando es multiplicada por la matriz A. La matriz inversa de **A** se denota $\mathbf{A}^{-1}$.

$$A^{-1}A = A\,A^{-1} = I$$

Se necesita de la matriz inversa debido a que podemos sumar matrices, restar matrices y multiplicar matrices, pero, no podemos dividir matrices directamente.

La función de la matriz inversa en algebra matricial es similar a la función del recíproco en artimética.

Solo las matrices cuadradas tienen inversa. Si una matriz cuadrada tiene inversa, entonces solo tiene una inversa. Si A tiene una inversa A^{-1}, premultiplicar o postmultiplicar A por su inversa producirá una matriz identidad I.

La inversa de una matriz consiste de la transpuesta de la matriz de cofactores dividida por el determinante de la matriz original. En general la inversa puede verse como :

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} Cof_{ij} \end{bmatrix}$$

Como ejemplo numérico, evaluamos la matriz 3 X 3

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & 1 \\ 2 & 3 & -1 \end{bmatrix}$$

En primer lugar obtenemos los cofactores de la primer columna :

$$Cof_{11} = -2$$

$$Cof_{21} = -5$$

$$Cof_{31} = -3$$

El determinante de A = 9, luego obtenemos los cofactores de la segunda columna:

Para obtener Cof_{12} eliminamos hilera 1 y columna 2

$$Cof_{12} = (-1)^{1+2} \begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} = -1^3 (-1-2) = -1 (-3) = 3$$

Para obtener el Cof_{22} eliminamos hilera 2 y columna 2

$$Cof_{22} = (-1)^{2+2} \begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} = (-1)^4 (-1-2) = 1 (-3) = -3$$

Para obtener el Cof_{32} eliminamos hilera 3 y columna 2

$$Cof_{32} = (-1)^{3+2} \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = (-1)^5 (1-1) = 0$$

Para obtener Cof_{13} eliminamos hilera 1 columna 3

$$Cof_{13} = (-1)^{1+3} \begin{vmatrix} 1 & -1 \\ 2 & 3 \end{vmatrix} = (-1)^4 (3+2) = 5$$

Para obtener Cof_{23} eliminamos hilera 2 columna 3

$$Cof_{23} = (-1)^{2+3} \begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} = (-1)^5 (3-4) = -1 (-1) = 1$$

Para obtener Cof_{33} eliminamos hilera 3 columna 3

$$Cof_{33} = (-1)^{3+3} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = (-1)^6 (-1-2) = 1 (-3) = -3$$

Por lo tanto la matriz de cofactores es :

$$\begin{bmatrix} -2 & 3 & 5 \\ 5 & -3 & 1 \\ 3 & 0 & -3 \end{bmatrix}$$

Y la inversa de A es

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} -2 & 3 & 5 \\ 5 & -3 & 1 \\ 3 & 0 & -3 \end{bmatrix}' = \frac{1}{9} \begin{bmatrix} -2 & 3 & 5 \\ 5 & -3 & 1 \\ 3 & 0 & -3 \end{bmatrix}'$$

$$A^{-1} = \begin{bmatrix} -2/9 & 3/9 & 5/9 \\ 5/9 & -3/9 & 1/9 \\ 3/9 & 0 & -3/9 \end{bmatrix}' = \begin{bmatrix} -2/9 & 5/9 & 3/9 \\ 3/9 & -3/9 & 0 \\ 5/9 & 1/9 & -3/9 \end{bmatrix}$$

Para probar si el calculo es correcto, se puede pre o postmultiplicar la inversa por la matriz original, de ser correcta el resultado será la matriz identidad I.

$$A^{-1}A = \begin{bmatrix} -2/9 & 5/9 & 3/9 \\ 3/9 & -3/9 & 0 \\ 5/9 & 1/9 & -3/9 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & 1 \\ 2 & 3 & -1 \end{bmatrix}$$

$$A^{-1}A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Ejercicio: encuentre la inversa de la matriz B

$$B = \begin{bmatrix} 3 & 2 & 4 \\ 4 & 1 & 6 \\ 6 & 3 & 1 \end{bmatrix}$$

SOLUCION DE SISTEMAS DE ECUACIONES

LA REGLA DE CRAMER

La regla de Cramer define un método para resolver n ecuaciones lineales simultáneas con n incógnitas.

Consideramos el siguiente sistema de ecuaciones lineales

$$a_{11} x_1 + a_{12} x_2 = b_1$$
$$a_{21} x_1 + a_{22} x_2 = b_2$$

Donde los a_{ij} son los coeficientes conocidos de las ecuaciones y la b_i con las constantes conocidas, este sistema puede representarse en forma

matricial así :

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\mathbf{AX} = \mathbf{b}$$

Siempre que $\mathbf{A} \neq 0$, el sistema tiene solución única

$$X_1 = \frac{|\mathbf{A}_1|}{|\mathbf{A}|} \qquad X_2 = \frac{|\mathbf{A}_2|}{|\mathbf{A}|}$$

Veamos un ejemplo numérico :

$$3X_1 + 2X_2 = 2$$
$$5X_1 + 5X_2 = 1$$

Que puede expresarse matricialmente :

$$\begin{bmatrix} 3 & 2 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$X_1 = \frac{|\mathbf{A}_1|}{|\mathbf{A}|}, \text{ donde } \mathbf{A}_1 = \begin{bmatrix} 2 & 2 \\ 1 & 5 \end{bmatrix}$$

Que se obtiene al intercambiar la primera columna de A por b

Que se obtiene al intercambiar la primera columna de A por b.

$$\left|A_1\right| = 8 \ , \quad \left|A\right| = 5$$

$$X_1 = \frac{8}{5}$$

$$X_2 = \frac{\left|A_2\right|}{\left|A\right|} \ , \text{ donde } A_2 = \begin{bmatrix} 3 & 2 \\ 5 & 1 \end{bmatrix}$$

$$\left|A_2\right| = -7$$

$$X_2 = \frac{-7}{5}$$

Otras soluciones involucran el cálculo de la matriz inversa, el sistema puede resolverse así :

$$\begin{aligned} Ax &= b \\ A^{-1} &= A^{-1} b \\ IX &= A^{-1} b \\ x &= A^{-1} b \end{aligned}$$

Aplicando este procedimiento a nuestro ejemplo :

$$A^{-1} = \begin{bmatrix} \frac{5}{5} & -2/5 \\ -5/5 & 3/5 \end{bmatrix}$$

$$A^{-1}b = \begin{bmatrix} 5/5 & -2/5 \\ -5/5 & 3/5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8/5 \\ -7/5 \end{bmatrix}$$

Evidentemente este procedimiento se dificulta cuando se considera la inversión de una matriz de gran tamaño.

Ejercicio:

Resolver el siguiente sistema de ecuaciones lineales :

$$\begin{aligned} 4X_1 - X_2 + 2X_3 &= 4 \\ 3X_1 + 4X_2 + X_3 &= 5 \\ -2X_1 - 2X_2 + 4X_3 &= 2 \end{aligned}$$

Resuelve tanto mediante la regla de Cramer como mediante la inversión de la matriz.

R E F E R E N C I A S

1.- AYRES, F. "Matrices, Teoría y 340 problemas resueltos" Serie compendios Schaum. McGraw-Hill, México, 1969

DRAPER, N.R. and H. SMITH. "Applied regression analysis". John Wiley, 1966.

2.- DORF, R.C. "Introducción al algebra de matrices" Mexico, Limusa-Wiley, 1972.

3.- FADDEEVA, V.N. "Computational methods of linear algebra" Dover Publications, 1959.

4.- HAVLICK, J. "Selected matrix topics for economists". Virginia Plytechnic Institute and State University Blackssburg, Virginia, 1974.

5.- NETER, J. and W. WASSERMAN. "Applied linear statistical models". Homewood, U.S.A., Irwin 1974.

INSTITUTO INTERAMERICANO DE COOPERACION PARA LA AGRICULTURA
Apartado Postal 1410, Tegucigalpa, Honduras, Tel. 31-54-52
Edificio Palmira, 2do. Piso, Frente al Hotel Honduras Maya.