

8-31



CENTRO DE DOCUMENTACION 22 FEB 1978

CURSO INTENSIVO DE ADIESTRAMIENTO EN PRODUCCION DE FRIJOL PARA
INVLTIGADORES DE AMERICA LATINA

CIAT Marzo 28 a Abril 23 de 1977

UTILIZACION DE LA ESTADISTICA Y EL DISEÑO
EXPERIMENTAL EN INVESTIGACION EN FRIJOL

Por Gastón Mendoza
/r
María Cristina Amézquita
Jorge Augusto Porras

Unidad de Biometría CIAT
Marzo de 1977

CONFERENCIA No 1

INTRODUCCION GENERAL A LA ESTADISTICA Y AL

DISEÑO EXPERIMENTAL

1 1 Introducción

Quisiera comenzar esta primera conferencia dándoles una idea bastante breve de lo que es la Unidad de Biometría dentro del CIAT y sobre cual es su papel dentro de un centro de investigación agropecuaria. A continuación definiremos el Método Científico entendido como la integración de las distintas etapas por las que pasa un investigador desde la observación crítica de un fenómeno hasta la inferencia de conclusiones respecto a tal fenómeno. Esto nos llevará a entender mejor la relación que existe entre la pregunta que se hace el investigador y el diseño experimental apropiado para poner a prueba su hipótesis. Luego hablaremos del porqué se utiliza la estadística en la investigación y finalmente introduciremos algunos conceptos y terminología básicos.

En las tres siguientes conferencias estudiaremos algunos de los diseños más utilizados en experimentación agrícola. En la quinta conferencia finalizaremos la parte teórica del curso haciendo una revisión de las técnicas de regresión, correlación y superficies de respuesta. Las conferencias 6 y 7 serán dedicadas a revisar algunas de las aplicaciones llevadas a cabo por el CIAT. En la octava conferencia describiremos los aspectos más saltantes del Sistema de Información de Irijol. Por último las conferencias 9 y 10 se utilizarán para ilustrar numéri-

camente algunos de los diseños experimentales introducidos anteriormente

1 2 Papel de la Unidad de Biometría en el CIAT

La Unidad de Biometría es un grupo central de servicio que presta asesoría en las etapas de planeación diseño procesamiento análisis e interpretación de los resultados relacionados con las distintas investigaciones y experimentos realizados por los programas del CIAT Estos servicios son sufragados con fondos de la misma Unidad y se suministran sin ningún costo a los programas de investigación y adiestramiento

Las funciones básicas de la Unidad de Biometría son

- 1 Asesoría estadística en la planeación diseño procesamiento análisis e interpretación de los experimentos
- 2 Creación y manejo de grandes sistemas de información (encuestas socio-económicas y agropecuarias bancos de germoplasma etc)
- 3 Desarrollo y proyectos de investigación cooperativos con otros programas
- 4 Evaluación de tecnología (estudio del impacto de nuevas variedades prácticas culturales etc desarrolladas por el CIAT)
- 5 Investigación sobre el desarrollo e implementación de nuevas técnicas estadísticas

6 Adiestramiento estadístico de profesionales

Las principales actividades cooperativas de la Unidad de Biometría con el Programa de Frijol durante 1976 fueron el desarrollo de un sistema de información para fitomejoramiento de frijol un estudio sobre la eficiencia relativa del diseño en látice respecto al de bloques completos al azar y procesamiento de la información de los viveros internacionales de rendimiento y de roya así como de las encuestas agro-económicas

1.3 El Método Científico

El método científico es el conjunto de las etapas lógicas que sigue un investigador para llegar a inferir algo a partir de la observación crítica de un fenómeno es decir es la aplicación objetiva de la lógica al mejor entendimiento de un fenómeno Su característica esencial es que partiendo de una observación crítica se llega a formular hipótesis que puedan probarse experimentalmente

El proceso que sigue el método científico consta de las siguientes etapas

- 1 Observación del fenómeno - Consiste en observar el fenómeno de una manera crítica sin que esto nos permita llegar a una conclusión Por ejemplo en un terreno sembrado con una misma variedad se observa que en determinadas áreas las plantas se ven raquíticas mientras que en otras se ven vigorosas

- 2 Planteamiento del problema - La observación crítica del fenómeno conduce al planteamiento operacional de un problema cuya solución debe ser la meta del investigador. En nuestro ejemplo anterior el problema podría plantearse como la respuesta a la pregunta «Es posible mejorar la producción en ese terreno?»
- 3 Establecimiento de las hipótesis - Son numerosas las hipótesis que el investigador puede plantear sobre las posibles causas del fenómeno observado. Lo importante es formular hipótesis relevantes al problema y que sean verificables experimentalmente, es decir, debe tenerse en cuenta la significancia operacional de resolver el problema. Siguiendo nuestro ejemplo, una hipótesis razonable podría ser
 H_0 La deficiencia de nitrógeno en el suelo produce falta de vigor en la planta.
- 4 Planeación del experimento - Establecida la hipótesis, el paso siguiente es la verificación objetiva de ellas a través de un experimento. En él, el investigador trata de controlar todos los factores excepto aquellos cuyo efecto desea determinar. Sin embargo, existen factores imposibles de ser controlados o que sería muy costoso controlarlos, como por ejemplo, las variables climatológicas. Estos factores no controlados constituyen el error experi-

mental. Antes de escoger un diseño experimental apropiado debe especificarse los tratamientos a ensayar, seleccionarse el material experimental, decidir a que poblaciones se espera extender los resultados del experimento y la precisión deseada. Si se deseara probar la hipótesis H_0 de nuestro ejemplo anterior, una forma de verificarla objetivamente sería ensayar distintos niveles fijos de nitrógeno y observar el comportamiento de la planta manteniendo los otros factores constantes (contenido de otros minerales en el suelo, riego, etc.).

5. Escogencia del diseño experimental - El diseño experimental es el patrón que indica la forma como se deben agrupar las unidades experimentales^{1/} y como se deben asignar los tratamientos a las unidades experimentales. Al escoger un diseño experimental se debe tratar de conciliar dos aspectos generalmente contrapuestos: sencillez y precisión. La mayor precisión se consigue seleccionando un diseño que minimize las variaciones no controladas por el investiga-

^{1/} Unidad experimental es la unidad mínima de material experimental a la cual se aplica un tratamiento dado. Por ejemplo, en experimentos de campo las unidades experimentales son generalmente las parcelas, no las plantas individuales.

dor es decir la varianza del error experimental. Además el tipo de diseño a utilizar depende de las hipótesis que se desean probar simultáneamente. Cuanto mayor sea el número de hipótesis más refinado será el diseño experimental a utilizar. Un buen diseño experimental provee la información deseada con un mínimo de esfuerzos y recursos. Luego de escoger el diseño experimental se diseñan los formatos de recolección de datos y el plan de análisis.

- 6 Ejecución del experimento - El experimento debe conducirse siguiendo estrictamente el diseño experimental y los controles culturales y estadísticos planeados. En términos generales las recomendaciones básicas para un buen manejo de experimentos agrícolas son uniformidad en la aplicación del riego, en la densidad de siembra y en la aplicación de insecticidas, fungicidas y herbicidas siempre y cuando éstos no sean los factores de interés para el investigador.
- 7 Análisis e interpretación de resultados - El análisis de los resultados que arroja un experimento tiene por objeto probar mediante métodos estadísticos la hipótesis planteada por el investigador.

8 Informe escrito - Este informe debe resumir todo aspecto de interés sobre el experimento desde su motivación hasta la interpretación de resultados. Es importante incluir todas las situaciones imprevistas que ocurrieron.

14 Utilidad de la estadística en la investigación

Existen dos tipos de experimentos: los determinísticos y los aleatorios. Un experimento determinístico es aquel cuyo resultado es para todos los efectos prácticos exacto; por ejemplo, los experimentos físicos. Un experimento aleatorio es aquel cuyo resultado no se puede predecir por estar sujeto a variaciones no controlables por el investigador; tales son los experimentos biológicos. En consecuencia, la verificación de una teoría mediante experimentos aleatorios no puede ser absoluta. El investigador sólo puede concluir que las observaciones son o no compatibles con la teoría dentro de los límites de error a los cuales las mismas observaciones están sometidas.

El papel de la Estadística consiste en proporcionar métodos que permitan distinguir entre situaciones donde las diferencias observadas entre tratamientos distintos son relativamente pequeñas y atribuibles al azar, y situaciones donde tales diferencias son relativamente grandes y son explicadas mejor por efectos diferentes de los tratamientos. En ambos casos, las conclusiones obtenidas son en un margen de confiabilidad conocido.

demostrar que $\Pr(X=x) = \binom{10}{x} \left(\frac{1}{2}\right)^{10}$ de modo que la distribución teórica de frecuencias (ajustadas a números enteros) es

x	0	1	2	3	4	5	6	7	8	9	10
Frecuencia teórica = $200 P_i(X=x)$	0	2	9	23	41	50	41	23	9	2	0

Ejemplo 2 Distribución normal con media $\mu = 5$ y varianza $\sigma^2 = 2.5$ Es una aproximación de la distribución teórica anterior con la misma tendencia central (media) e igual dispersión alrededor de la media (varianza) Su gráfico se muestra en la figura (b)

Distribución Normal Una distribución normal es caracterizada por dos parámetros μ (media) y σ^2 (varianza $\sigma =$ desviación estándar) Si X es distribuida siguiendo una distribución normal con media μ y varianza σ^2 escribiremos $X \sim N(\mu, \sigma^2)$ La distribución normal es muy usada en estadística por razones prácticas y teóricas muy manejable y extensamente tabulada muchas variables aleatorias siguen aproximadamente una distribución normal o pueden ser reducidas a normales mediante una transformación adecuada la distribución de medias muestrales de cualquier población tiende a ser normal a medida que el tamaño de la muestra aumenta

A continuación se presentan algunas de las propiedades de la distribución normal

1.3 Conceptos y Terminología Básicos

Muestra y Población Una muestra es una colección de individuos u observaciones perteneciente a una colección mayor llamada población o universo del cual deseamos información. Si el proceso de selección de los individuos es al azar se dice que la muestra es aleatoria.

Variable Aleatoria Es aquella cuyo valor no puede predecirse sino que depende del azar.

Distribución de Frecuencia Es la tabla de frecuencias obtenida agrupando los datos en clases excluyentes y exhaustivas. Su representación gráfica es llamada histograma de frecuencias. Para el caso de una variable continua si se reduce el intervalo de clase en forma indefinida se obtiene la función de distribución.

Ejemplo 1 Distribución del número de dígitos impares en cada una de 200 muestras aleatorias de 10 dígitos.

Sea X = número de dígitos impares en una muestra de 10. Entonces la distribución (observada) de frecuencias pudo ser

x	0	1	2	3	4	5	6	7	8	9	10
frecuencia observada	2	2	8	25	39	45	35	20	14	4	1

y el correspondiente histograma de frecuencias de muestra en la figura (a). La variable X tiene por distribución teórica

la binomial $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ donde $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

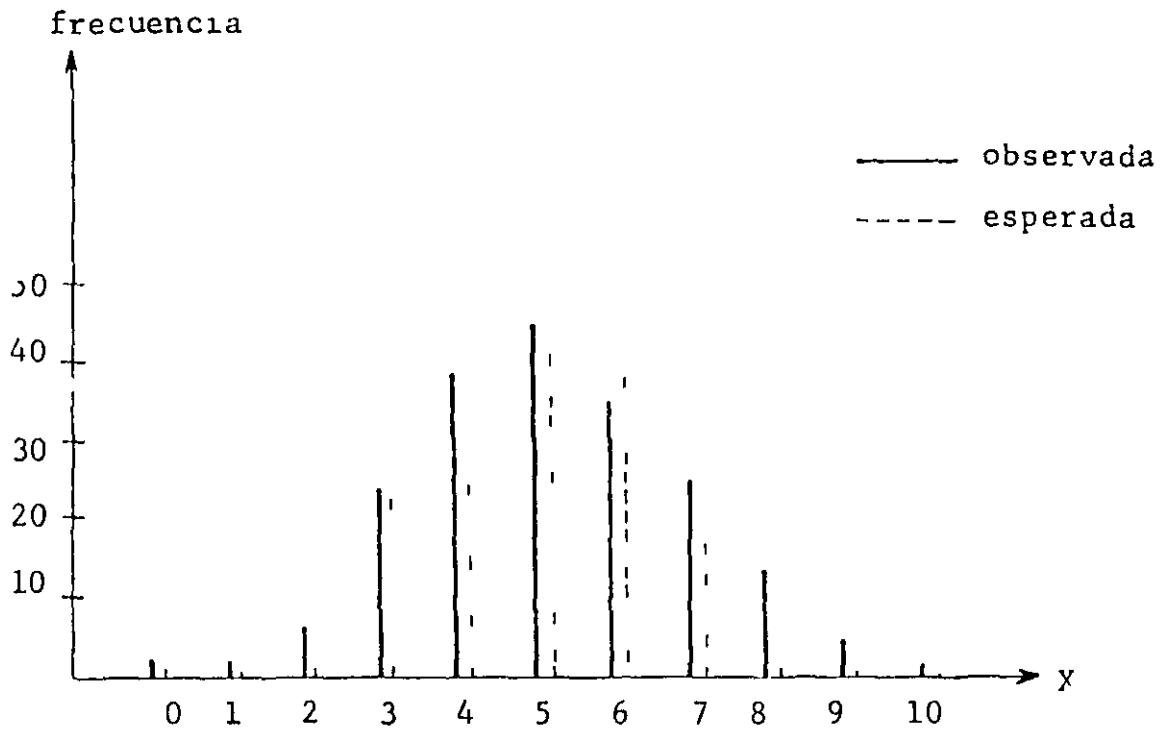


Fig (a) Distribución de frecuencias del número de dígitos impares en cada una de 200 muestras de tamaño 10

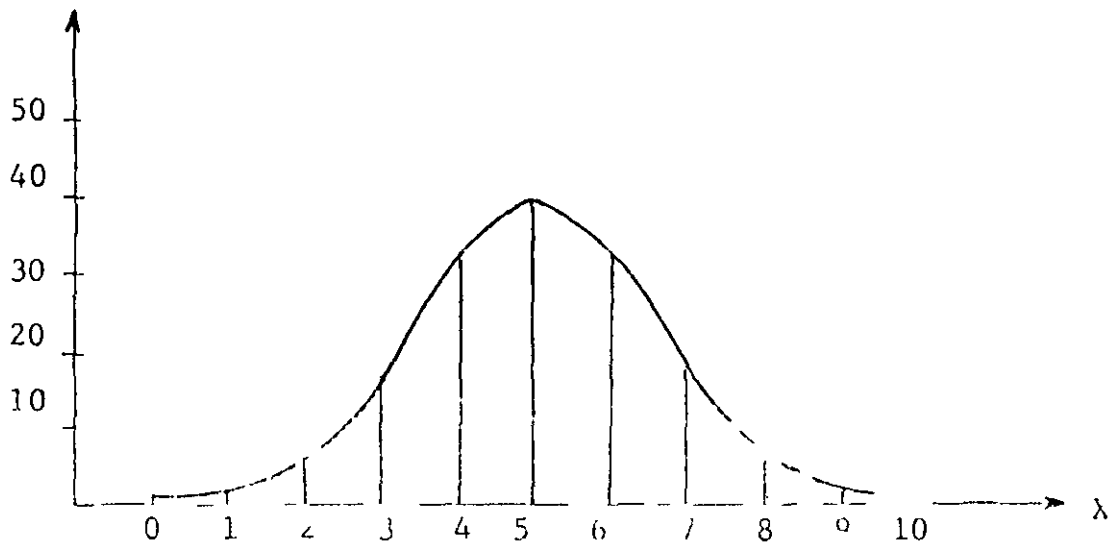


Fig (b) Gráfico de la función de distribución (o densidad) de una dist

- 1 Función de densidad de probabilidad

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/\sigma^2} \quad -\infty < x < \infty$$

- 2 Función de distribución acumulativa

$$F_X(x) = \Pr(X \leq x) = \int_{-\infty}^x f_X(x) dx$$

= area bajo la curva $f_X(x)$ desde $-\infty$ hasta x

- 3 La siguiente propiedad es válida para toda variable aleatoria

$$\int_{-\infty}^{\infty} f_X(x) dx = 1$$

- 4 Los parámetros μ y σ^2 se estiman a partir de una muestra de n observaciones por los siguientes estadísticos

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- 5 Si $X \sim N(\mu, \sigma^2)$ y σ^2 es conocido entonces

$$\Pr(\mu - \sigma < X < \mu + \sigma) = 68$$

$$\Pr(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 95$$

$$(X - \mu) / \sigma \sim N(0, 1)$$

$$\bar{x} \sim N(\mu, \sigma^2/n) \quad \text{y} \quad \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

- 6 Si $X \sim N(\mu, \sigma^2)$ pero σ^2 no se conoce entonces

$(\bar{x} - \mu) \sqrt{n} / \sigma$ sigue una distribución similar a la $N(0, 1)$ llamada la distribución t con $(n-1)$ grados de libertad

Prueba de Hipótesis

Hipótesis nula (H_0) Es aquella que se acepta actualmente como cierta de modo que sólo es rechazada si la evidencia experimental en su contra es muy grande

Hipótesis alternativa (H_1) Es aquella que se desearía probar al rechazar la hipótesis nula

Ejemplo 1 Prueba de os colas $H_0 \mu_1 = \mu_2$ Vs $H_1 \mu_1 \neq \mu_2$

Ejemplo 2 Prueba de una cola $H_0 \mu_1 < \mu_2$ Vs $H_1 \mu_1 > \mu_2$

Error de tipo I Es aquel que se comete cuando se rechaza una hipótesis nula verdadera

Error de tipo II Es aquel que se comete cuando se acepta una hipótesis alternativa falsa

Los procedimientos clásicos de prueba de hipótesis minimizan las probabilidades de error de tipo II (denominadas 'potencia de la prueba') para un nivel predeterminado de probabilidad de error de tipo I (denominada nivel de significancia) y del tamaño de la muestra

Nivel de Significancia $\alpha = \text{Pr}(\text{rechazar } H_0 | H_0 \text{ es verdadera})$

Nivel de confianza $1 - \alpha = \text{Pr}(\text{aceptar } H_0 | H_0 \text{ es verdadera})$

DISEÑOS MAS UTILIZADOS EN EXPERIMENTACION CON FRIJOL

2 1 Qué es el diseño experimental

Por diseño experimental se entiende el conjunto de reglas que indican cómo asignar los tratamientos a las unidades experimentales. Un buen diseño permite efectuar comparaciones válidas entre tratamientos y controlar la principal fuente de variación que presentan los experimentos de campo, la heterogeneidad del suelo. Un buen diseño debe incluir tres aspectos importantes: Aleatorización en la aplicación de los tratamientos, un adecuado número de replicaciones y un control máximo del error experimental.

2 2 Escogencia del diseño

El mejor tipo de diseño para un experimento dado depende de la magnitud de la heterogeneidad del suelo en el área experimental, del tipo y número de tratamientos que se deseen probar y del grado de precisión deseado.

2 3 Diseños más utilizados en experimentación con frijol

Los diseños más comunmente usados en experimentos de campo en frijol son

- Completamente al azar (para uno o varios factores)
- Bloques completos al azar (para uno o varios factores)
- Parcelas divididas
- Jatices
- Diseños sistematicos

Vamos a describir brevemente como y cuando usar cada uno de estos diseños. Presentamos un resumen de los cálculos necesarios para el análisis estadístico y algunos ejemplos ilustrando su uso.

2.4 Diseño Completamente al Azar

- Se usa cuando las unidades experimentales son homogéneas
- Con él se puede probar cualquier número de tratamientos (ya sean niveles de un solo factor o combinaciones de niveles de varios factores)
- Los tratamientos se aplican a las unidades experimentales al azar
- Cualquier número de repeticiones es posible

2.4.1 Análisis de Varianza

Modelo Matemático

$$Y_{ij} = \mu + \tau_i + e_{ij} \quad \begin{matrix} i = 1, 2, \dots, t \\ j = 1, 2, \dots, r \end{matrix}$$

↙ Error experimental en la celda (i, j)

↖ efecto del tratamiento i

↖ media global

↖ característica bajo estudio observada en la parcela j y donde se aplicó el tratamiento i

$$\text{supuestos: } e_{ij} \sim \text{NID}(0, \sigma^2) \quad \sum_{i=1}^t \tau_i = 0$$

$$S_1 \bar{Y} = \left(\sum_{i=1}^t \sum_{j=1}^r Y_{1j} \right) / (rt)$$

$$\bar{Y}_1 = \left(\sum_{j=1}^r Y_{1j} \right) / r$$

entonces \bar{Y} es un estimador de μ

\bar{Y}_1 es un estimador de $\mu + \tau_1$

Además, la suma de cuadrados de las desviaciones respecto a \bar{Y} denominada suma total de cuadrados puede descomponerse de la siguiente manera

$$\sum \sum (Y_{1j} - \bar{Y})^2 = \sum (\bar{Y}_1 - \bar{Y})^2 + \sum \sum (Y_{1j} - \bar{Y}_1)^2$$

La primera suma del segundo miembro es un indicador de las diferencias entre medias de tratamientos y la segunda es un indicador de la variabilidad de las observaciones respecto a la correspondiente media de tratamiento. Por tal motivo ellas se denominan suma de cuadrados debido a tratamientos y suma de cuadrados del error respectivamente. Para hacer comparables estos indicadores se introducen los denominados grados de libertad. El cociente de una suma de cuadrados por su correspondiente número de grados de libertad es denominado cuadrado medio del efecto bajo consideración. Los grados de libertad asociados con SCT, SCTR y SCE son respectivamente $rt-1$, $t-1$ y $(r-1)t$.

Consideremos la hipótesis nula $H_0: \tau_1 = 0$ $i=1, \dots, t$,

Vs la hipótesis alternativa

$H_1: \tau_1 \neq 0$ para al menos

Si la hipótesis H_0 es verdadera es decir no hay diferencias entre medias de tratamientos entonces $CMTR$ y CME tienden a ser similares y en consecuencia el cociente $CMTR/CME$ tiende a ser aproximadamente igual a uno. Si por el contrario H_1 es verdadera entonces $CMTR$ tiende a ser mayor que CME y en consecuencia el cociente $CMTR/CME$ tiende a ser mayor que uno. Por lo tanto valores de $CMTR/CME$ cercanos a uno soportan H_0 y valores mucho más grandes que uno soportan H_1 . Es interesante señalar que CME es un estimador de la varianza σ^2 la cual existe debido a los factores aleatorios no controlables por el investigador. Queda por decidir cuan grande debe ser $CMTR/CME$ para poder concluir con cierto grado de confianza que las diferencias observadas entre medias de tratamientos son debidas a diferencias reales entre los tratamientos y no al azar. Para ello es necesario escoger el nivel de confianza $1-\alpha$ y utilizar el hecho de que bajo H_0 el cociente $CMTR/CME$ sigue una distribución llamada como la distribución F con $t-1$ y $(r-1)t$ grados de libertad. Resumiendo la hipótesis H_0 es rechazada al nivel de significancia α si y sólo si

$$F_{\text{observado}} = \frac{CMTR}{CME} > F_{t-1, (r-1)t}(\alpha) = \alpha\text{-percentil superior de la distribución } F_{t-1, (r-1)t}$$

Todo el procedimiento anterior puede condensarse en la llamada tabla de ANOVA

<u>Fuente de Variación</u>	<u>g. de l.</u>	<u>S. C.</u>	<u>c. m.</u>	<u>F. observ.</u>
Tritamiento	$t - 1$	$\frac{1}{r} \sum Y_{1j}^2 - \frac{1}{rt} Y^2$	CMTR	$\frac{CMTR}{CME}$
Error	$t(t-1)$	$SCT - SCT_k$	CMF	
Total (corregido por la media global)	$tr - 1$	$\sum \sum Y_{1j}^2 - \frac{1}{rt} Y^2$		

En la tabla anterior las fórmulas para las sumas de cuadrados son apropiadas para el cálculo por calculadora

Y_{1j} denota $\sum_j Y_{1j}$ e Y denota $\sum_{1j} Y_{1j}$

2.4.2 Ejemplo Numérico

Se desea comparar el rendimiento de $t=5$ variedades de frijol. Por experiencias anteriores se conoce que el terreno disponible es homogéneo. Además siguiendo las recomendaciones sobre tamaño de parcela se disponen de 30 parcelas. Entonces podemos utilizar $r = 6$ repeticiones para cada variedad (cabe señalar que el número de repeticiones a utilizar es generalmente determinado por la precisión deseada y no por el área disponible). El siguiente paso es asignar las variedades a las parcelas completamente al azar. Supongamos que los siguientes fueron los rendimientos observados en gramos por m^2

<u>Variedad</u>							<u>Y_{1j}</u>	<u>\bar{Y}_{1j}</u>
1	88	129	117	312	220	99	965	161
2	235	263	216	156	244	233	1347	224
3	412	225	218	463	156	226	1700	283
4	284	404	164	445	388	436	2201	367
5	674	332	595	498	571	366	3036	506
							$Y = 9249$	308.3

completamente al azar se requiere que la variación entre bloques sea lo más grande posible y que la variación dentro del bloque sea mínima. Además para que las pruebas de significancia sean válidas es necesario que no exista interacción tratamiento x bloque

3.1.1 Análisis de Variancia

Modelo Matemático

$$Y_{1j} = \mu + \tau_1 + \beta_j + e_{1j} \quad \begin{matrix} i = 1, 2, \dots, t \\ j = 1, 2, \dots, b \end{matrix}$$

e_{1j} - error experimental en la celda (i, j)
 β_j - efecto del bloque j
 τ_1 - efecto del tratamiento i
 μ - media global
 Y_{1j} - respuesta observada en la celda (i, j)

Supuestos $e_{1j} \sim \text{NID}(0, \sigma^2)$ $\sum_{i=1}^t \tau_i = 0 = \sum_{j=1}^b \beta_j$

Si $\bar{Y} = \left(\sum_{i=1}^t \sum_{j=1}^b Y_{1j} \right) / (bt)$

$$\bar{Y}_1 = \left(\sum_{j=1}^b Y_{1j} \right) / b \quad \bar{Y}_j = \left(\sum_{i=1}^t Y_{1j} \right) / t$$

entonces \bar{Y} es un estimador de μ

\bar{Y}_1 es un estimador de $\mu + \tau_1$

\bar{Y}_j es un estimador de $\mu + \beta_j$

La suma total de cuadrados conseguida por la media puede descomponerse de la siguiente manera

$$\sum_{i=1}^t (Y_{ij} - \bar{Y}_j)^2 + \sum_{j=1}^b (Y_{1j} - \bar{Y}_j)^2 + \sum_{j=1}^b (Y_{2j} - \bar{Y}_j)^2 + \dots + \sum_{j=1}^b (Y_{tj} - \bar{Y}_j)^2$$

Los cuadrados del miembro son de
 el tipo de bloques. Los grados de libertad asociados
 son $SC_{T} = SC_{Tb} = SC_{T(t-1)} = (t-1)(b-1)$ y $(b-1)(t-1)$
 respectivamente. Como el caso del diseño completamente al azar,
 el error es una proporción constante número de g.l.
 nominalmente cuadrados m. de tal efecto. Así mismo CME es
 estimador de σ^2 . Sin embargo el σ^2 del modelo de B C A no
 es el mismo que el del diseño completamente al azar. De hecho
 el bloqueo efectivo $\sigma^2_{\text{bloqueo}} \ll \sigma^2$ completamente al azar
 lo que es justamente el objetivo del bloqueo: aumentar la precisión
 de las comparaciones extrayendo de la variabilidad del diseño
 completamente al azar aquella variabilidad debida a diferencias
 entre bloques.

Para el diseño de BCA es posible probar independien-
temente los siguientes pares de hipótesis:

$$H_0: \tau_1 = 0 \quad i = 1, \dots, t \quad \text{Vs} \quad (I)$$

$$H_1: \tau_1 \neq 0 \quad \text{para al menos un } i \quad \text{y}$$

$$H_0: \beta_j = 0 \quad j = 1, \dots, b \quad \text{Vs} \quad (II)$$

$$H_1: \beta_j \neq 0 \quad \text{para al menos un } j$$

Los tests son similares a los expuestos para el caso
 del diseño completamente al azar. El test F para los

pares de hipótesis anteriores son

(i) Rechazar la hipótesis $\tau_1 = 0$ $i = 1, \dots, t$ al nivel

de significancia α sss $\frac{CMTR}{CME} > F_{t-1, (b-1)(t-1)}(\alpha)$

(ii) Rechazar la hipótesis $\beta_j = 0$ $j = 1, \dots, b$ al nivel

de significancia α sss $\frac{CMBL}{CME} > F_{b-1, (b-1)(t-1)}(\alpha)$

La Tabla del ANOVA es

<u>Fuente de Variación</u>	<u>g. d. l.</u>	<u>s. c.</u>	<u>c. m.</u>	$F_{\text{observados}}$
Tratamientos	$t - 1$	$\frac{1}{b} \sum_1^t Y_{1j}^2 - \frac{1}{bt} Y^2$	CMTR	CMTR/CME
Bloques	$b - 1$	$\frac{1}{t} \sum_j Y_j^2 - \frac{1}{bt} Y^2$	CMBL	CMBL/CME
<u>Error</u>	<u>$(t-1)(b-1)$</u>	<u>$SC1 - SCTR - SCBL$</u>	<u>CML</u>	
Total (corregido por la media global)	$tb - 1$	$\sum \sum Y_{1j}^2 - \frac{1}{bt} Y^2$		

3.1.2 Ejemplo Numerico

Supongamos que se desea comparar el rendimiento de $t = 5$ variedades de frijol pero el terreno no es homogéneo. Además supongamos que es posible agrupar las 30 parcelas disponibles en $b = 6$ bloques de 5 parcelas cada uno de tal manera que parcelas de un mismo bloque son mas o menos igualmente fértiles. El siguiente paso es asignar las variedades al azar a las parcelas de cada bloque. Debe usarse una aleatorización diferente para cada bloque. Supongamos que los rendimientos en toneladas por m^2 fueron

Variedad	I	II	III	IV	V	VI	Y_1	\bar{Y}_1
1	88	129	117	312	220	99	965	161
2	235	263	216	156	244	233	1347	224
3	412	225	218	463	156	236	1700	283
4	284	484	164	445	388	436	2201	367
5	674	332	595	498	571	366	3036	506
Y_j	1693	1433	1310	1874	1579	1360	$\Sigma Y = 9249$	$\bar{Y} = 308.3$
Y_j	338.6	286.6	262.0	374.8	315.8	272.0		

$$SCTR = (965^2 + 1347^2 + 1700^2 + 2201^2 + 3036^2) / 6 - 9249^2 / 30 = 431421.8$$

$$SCBL = (1693^2 + 1433^2 + 1310^2 + 1874^2 + 1579^2 + 1360^2) / 5 - 9249^2 / 30 = 46644.3$$

$$SCT = 88^2 + 129^2 + 117^2 + 312^2 + 220^2 + 99^2 + 235^2 + 263^2 + 216^2 + 156^2 + 244^2 + 233^2 + 412^2 + 225^2 + 218^2 + 463^2 + 156^2 + 236^2 + 284^2 + 484^2 + 164^2 + 445^2 + 388^2 + 436^2 + 674^2 + 332^2 + 595^2 + 498^2 + 571^2 + 366^2 - 9249^2 / 30 = 716036.3$$

$$SCL = 716036.3 - 431421.8 - 46644.3 = 237970.2$$

La tabla del ANOVA es

<u>ΓV</u>	<u>g l</u>	<u>S C</u>	<u>c m</u>	$F_{\text{observados}}$
Variedades	4	431421.8	107855.5	9.065
Bloques	5	46644.3	9328.9	0.784
<u>Error</u>	<u>20</u>	<u>237970.2</u>	<u>11898.5</u>	
Total	29	716036.2		

Si $\alpha = 0.1$ la hipótesis nula de (I) es aceptada puesto que $9.065 < 14.02 = F_{4, 20}(0.1)$ similarmente la hipótesis nula de (II) es aceptada puesto que

$$0.784 < 9.55 = F_{5, 20}(0.1)$$

Es claro entonces que en este caso el bloque no fue efectivo

3 2 Diseño en Parcelas Divididas y Subdivididas

Se utilizan cuando la madurez de los niveles de un factor o problemas de manejo del experimento requieren el uso de unidades grandes mientras que los niveles de otros factores si pueden asignarse a unidades más pequeñas. Por ejemplo en experimentos en donde el factor riego es uno de los que se desea medir es aconsejable separar las parcelas que reciben un determinado nivel de riego. Casos similares son los experimentos de fertilizantes e insecticidas.

El diseño en parcelas divididas se usa cuando se desea estudiar dos factores uno de los cuales requiere unidades grandes y el otro puede asignarse a unidades mas pequeñas. Los niveles del primer factor se asignan al azar a las parcelas grandes. Los niveles del segundo factor se asignan al azar a las sub-parcelas de cada parcela grande. Cada parcela grande contiene tantas sub-parcelas como niveles del segundo factor existan. Las comparaciones entre los niveles del primer factor son menos precisas que aquellas entre los niveles del segundo factor y aquellas entre interacciones del primer y segundo factor. Para poder hacer estas comparaciones es necesario utilizar por lo menos dos replicaciones.

El diseño en parcelas sub-divididas se usa cuando se de-

sea estudiar tres factores uno de los cuales requiere unidades grandes y los otros dos pueden asignarse a unidades más pequeñas. Los niveles del primer factor se asignan al azar a las parcelas grandes. Los niveles del segundo factor se asignan al azar a las sub-parcelas de cada parcela grande y los del tercer factor se asignan al azar a las sub-sub-parcelas de cada sub-parcela. Cada parcela grande contiene tantas sub-parcelas como niveles del segundo factor existan. Similarmente cada sub-parcela contiene tantas sub-sub-parcelas como niveles del tercer factor existan. Las comparaciones entre los niveles del primer factor son las menos precisas. Las comparaciones entre los niveles del segundo factor y de las interacciones del primer y segundo factor son de precisión intermedia. Finalmente las comparaciones entre los niveles del tercer factor y de las interacciones que lo contienen son las más precisas. Como en el caso de parcelas divididas para poder hacer comparaciones válidas se requiere al menos dos replicaciones.

3 2 1 Ejemplo Numérico Parcelas sub-sub-divididas

Se desea analizar el efecto del empoasca sobre el rendimiento de cuatro variedades de frijol con y sin la aplicación de insecticida con y sin riego. Es decir

Factor A	Riego a dos niveles	Con riego (a_1)
		Sin riego (a_0)

Factor B	Insecticida a dos niveles	Con insecticida (b_1) Sin insecticida (b_0)
Factor C	Variedad a cuatro niveles	Variedad 1 Variedad 2 Variedad 3 Variedad 4

Con el objeto de mantener separadas las parcelas que reciben riego de las que no lo reciben se dividió el terreno en dos parcelas grandes y se les asigno al azar los dos niveles del factor riego. Cada parcela grande se dividió en dos sub-parcelas a las que se asignó al azar los niveles del factor insecticida. Finalmente cada sub-parcela se dividió en cuatro sub sub-parcelas a las que se asigno al azar las cuatro variedades. Se utilizarón dos replicaciones en el experimento.

La disposición de los factores sobre el terreno quedó como muestra la gráfica siguiente

REPLICACION I

Parcela Grande 1
(con riego)

Parcela Grande 2
(Sin riego)

Con Insecticida	Sin Insecticida
V ₂ (8)	V ₁ (4)
V ₄ (5)	V (6)
V ₁ (6)	V ₂ (6)
V ₃ (7)	V ₄ (3)

Con Insecticida	Sin Insecticida
V ₂ (5)	V ₃ (6)
V ₄ (3)	V ₁ (3)
V (5)	V ₂ (5)
V ₁ (3)	V ₄ (2)

REPLICACION II

Parcela Grande 1
(Sin riego)

Parcela Grande 2
(Con riego)

Sin Insecticida	Con Insecticida	Con Insecticida	Sin Insecticida
V ₂ (8)	V ₃ (6)	V ₃ (7)	V ₁ (4)
V ₁ (4)	V ₂ (4)	V ₁ (6)	V ₃ (8)
V ₃ (5)	V ₁ (5)	V ₄ (4)	V ₂ (4)
V ₄ (4)	V ₄ (4)	V ₂ (7)	V ₄ (5)

Los números en parentesis son los rendimientos en Kgs

Para el calculo de las sumas de cuadrados del ANOVA se requieren las siguientes tablas

AxB

	b ₀	b ₁	
a ₀	35	35	70
a ₁	40	50	90
	75	85	160 = Y

AxC

	V ₁	V ₂	V ₃	V ₄	
a ₀	15	20	22	13	70
a ₁	20	25	28	17	90
	35	45	50	30	160

BxC

	V ₁	V ₂	V ₃	V ₄	
b ₀	15	21	25	14	75
b ₁	20	24	25	16	85
	35	45	50	30	160

AxBxC

a₀

	V ₁	V ₂	V ₃	V ₄	
b ₀	7	11	11	6	35
b ₁	8	9	11	7	35
	15	20	22	13	70

a₁

	V ₁	V ₂	V ₃	V ₄	
b ₀	8	10	14	8	40
b ₁	12	15	14	9	50
	20	25	28	17	90

Ax Rep

	I	II	
a ₀	32	38	70
a ₁	45	45	90
	77	83	160

B x Rep

	I	II	
b ₀	35	40	75
b ₁	42	43	85
	77	83	160

AxBxRep

I

	b ₀	b ₁	
a ₀	16	16	32
a ₁	19	26	45
	35	42	77

II

	b ₀	b ₁	
a ₀	19	19	38
a ₁	21	24	45
	40	43	83

El término de corrección es

$$TC = \frac{Y^2}{\# \text{ total d parcelas}} = \frac{160^2}{2 \times (2 \times 2 \times 4)} = 800$$

$$SCT = 8^2 + 5^2 + \dots + 4^2 + 5^2 - TC = 70\ 000$$

$$SC \text{ Rep} = (77^2 + 83^2) / 16 - TC = 1\ 125$$

$$SCA = (70^2 + 90^2) / 16 \quad C = 12\ 500$$

$$SCB = (75^2 + 85^2) / 16 - TC = 3\ 125$$

$$SCAB = (35^2 + 35^2 + 40^2 + 50^2) / 8 - SCA - SCB - TC = 3\ 125$$

$$SCC = (35^2 + 45^2 + 50^2 + 30^2) / 8 - TC = 31\ 25$$

$$SCAC = (15^2 + 20^2 + 22^2 + 13^2 + 20^2 + 25^2 + 28^2 + 17^2) / 4 - SCA - SCC - TC = 0\ 250$$

$$SCBC = (15^2 + 21^2 + 25^2 + 14^2 + 20^2 + 24^2 + 25^2 + 16^2) / 4 - SCB - SCC - TC = 1\ 625$$

$$SCABC = (7^2 + 11^2 + \dots + 14^2 + 9^2) / 2 - SCA - SCB - SCC - SCAB - SCAC - SCBC - TC = 4\ 125$$

$$SC(AxRep) = (32^2 + 38^2 + 45^2 + 45^2) / 8 - SCA - SCRep - TC = 1\ 125 = \text{Error (a)}$$

$$SC(BxRep) = (35^2 + 40^2 + 42^2 + 43^2) / 8 - SCB - SCRep - TC = 0\ 5$$

$$SC(AxBxRep) = (16^2 + 16^2 + 19^2 + 26^2 + 19^2 + 19^2 + 21^2 + 24^2) / 4 - SCA - SCB - SCRep -$$

$$SCAB - SC(AxRep) - SC(BxRep) - TC = 0\ 5$$

$$\text{Error (b)} = SC(BxRep) = SC(AxBxRep) = 1\ 000$$

<u>Fuente de Variación</u>	<u>g de l</u>	<u>s c</u>	<u>c m</u>	<u>F</u>
Replicación	$r-1=1$	1 125	1 125	1 00
Riego = A	$a-1=1$	12 500	12 500	11 11
Error (a)	$(r-1)(a-1)=1$	1 125	<u>1 125</u>	
Insecticida = B	$b-1=1$	3 125	3 125	6 25
AxB	$(a-1)(b-1)=1$	3 125	3 125	6 25
Error (b)	$a(b-1)(r-1)=2$	1 000	<u>0 500</u>	
Variedad = C	$c-1=3$	31 250	10 417	11 63
AxC	$(a-1)(c-1)=3$	0 250	0 083	0 09
BxC	$(b-1)(c-1)=3$	1 625	0 542	0 61
AxBxC	$(a-1)(b-1)(c-1)=3$	4 125	1 375	1 54
<u>Error (c)</u>	<u>$ab(c-1)=12$</u>	<u>10 750^{1/}</u>	<u>0 896</u>	
Total	$abc - 1 = 31$	70 000		

Los efectos de replicación y riego se prueban con el Error (a)

$$\frac{CM_{Rep}}{CM_{Error(a)}} = 1.00 \div 161 = F_{1, 1}(0.5) \quad \text{se acepta la hipótesis de iguales efectos de todas las repeticiones}$$

$$\frac{CMA}{CM_{Error(a)}} = 11.11 \div 161 = F_{1, 1}(0.5) \quad \text{se acepta la hipótesis de iguales efectos con y sin riego}$$

Los efectos de insecticida y de la interacción insecticida x riego se prueban con el Error(b)

^{1/}

El error (c) es obtenido por diferencia

$$\frac{CMB}{CME_{\text{error}(b)}} = 6.25 \uparrow 18.51 = F_{1, 2}(05) \quad \text{Se acepta la hipótesis de iguales efectos con y sin insecticida}$$

$$\frac{CMAB}{CME_{\text{error}(b)}} = 6.25 \uparrow 18.51 = F_{1, 2}(05) \quad \text{Se acepta la hipótesis de que no hay interacción AxB}$$

Los efectos de variedad y las interacciones que lo contienen se prueban en el Error(c)

$$\frac{CNC}{CME_{\text{error}(c)}} = 11.63 > 8.74 = F_{3, 12}(05) \quad \text{Se rechaza la hipótesis de iguales efectos varietales}$$

Finalmente los bajos valores de F para AxC BxC y AxBxC conducen a aceptar las hipótesis de que no hay tales interacciones

3 2 2 Análisis de Varianza para un Diseño en Parcelas Divididas

<u>Fuente de Variación</u>	<u>g l</u>	<u>S C</u>
Replicación (Pep)	r-1	$(R_1^2 + \dots + R_r^2) / (ab) - TC$
Parcela Principal (A)	a-1	$(A_0^2 + \dots + A_{a-1}^2) / (rb) - TC$
Error (a)=RepxA	(r-1)(a-1)	$\frac{\overline{A_0 R_1^2} + \overline{A_0 R_2^2} + \dots + \overline{A_{a-1} R_r^2}}{SC_{Rep} - SCA - TC} / b -$
Sub-parcela (B)	b-1	$(B_0^2 + \dots + B_{b-1}^2) / (rb) - TC$
AxB	(a-1)(b-1)	$\frac{\overline{A_0 B_0^2} + \overline{A_0 B_1^2} + \dots + \overline{A_{a-1} B_{b-1}^2}}{SCA - SCB - TC} / r -$
Error (b)=RepxB+RepxAxB	a(r-1)(b-1)	Por diferencia

Total (corregido por la media global) $abr - 1 \quad \sum \sum Y_{1jk}^2 - TC$

donde

$R_k = \text{total para la replicación } k \quad k=1, 2, \dots, r$

$$\frac{CM_B}{CME_{\text{error}(b)}} = 6.25 \dagger 18.51 = F_{1, 2}(05) \quad \text{Se acepta la hipótesis de iguales efectos con y sin insecticida}$$

$$\frac{CM_{AB}}{CME_{\text{error}(b)}} = 6.25 \dagger 18.51 = F_{1, 2}(05) \quad \text{Se acepta la hipótesis de que no hay interacción AxB}$$

Los efectos de variedad y las interacciones que lo contienen se prueban en el Lrro(c)

$$\frac{CM_C}{CME_{\text{error}(c)}} = 11.63 > 8.74 = F_{3, 12}(05) \quad \text{Se rechaza la hipótesis de iguales efectos varietales}$$

Finalmente los bajos valores de F para AxC BxC y AxBxC conducen a aceptar las hipótesis de que no hay tales interacciones

3 2 2 Análisis de Varianza para un Diseño en Parcelas Divididas

<u>Fuente de Variación</u>	<u>g l</u>	<u>S C</u>
Replicación (Pep)	r-1	$(R_1^2 + \dots + R_r^2) / (ab) - TC$
Parcela Principal (A)	a-1	$(A_0^2 + \dots + A_{a-1}^2) / (rb) - TC$
Error (a)=RepxA	(r-1)(a-1)	$\frac{\overline{A_0 R_1^2} + \overline{A_0 R_2^2} + \dots + \overline{A_{a-1} R_r^2}}{SC_{Rep} - SCA - TC} / b -$
Sub-parcela (B)	b-1	$(B_0^2 + \dots + B_{b-1}^2) / (rb) - TC$
AxB	(a-1)(b-1)	$\frac{\overline{A_0 B_0^2} + \overline{A_0 B_1^2} + \dots + \overline{A_{a-1} B_{b-1}^2}}{SCA - SCB - TC} / r -$
Error(b)=RepxB+RepxAxB	a(r-1)(b-1)	Por diferencia
Total (corregido por la media global)	abr - 1	$\sum \sum Y_{1jk}^2 - TC$

donde

$$R_k = \text{total para la replicación } k \quad k=1, 2, \dots, r$$

A_i = total para el nivel i -ésimo del factor A $i = 0 \ 1 \ \dots \ a-1$

B_j = total para el nivel j -ésimo del factor B $j = 0 \ 1 \ \dots \ b-1$

$\overline{A_i B_j}$ = total para la combinación $(a_i \ b_j)$ $i = 0 \ 1 \ \dots \ a-1 \ j = 0 \ 1 \ \dots \ b-1$

$\overline{A_i R_k}$ = total para la combinación $(a_i \ r_k)$ $i = 0 \ 1 \ \dots \ a-1 \ k = 1 \ 2 \ \dots \ r$

TC = $Y^2 / (abr)$ = término de corrección

- (i) La hipótesis de que las medias de todos los niveles del factor A son iguales se rechaza sss $CMA / CME_{\text{error}}(a) > F_{a-1 \ (r-1) \ (a-1) \ (\alpha)}$
- (ii) La hipótesis de que las medias de todos los niveles del factor B son iguales se rechaza sss $CMB / CME_{\text{error}}(b) > F_{b-1 \ a(r-1) \ (b-1) \ (\alpha)}$
- (iii) La hipótesis de que las medias de todas las celdas $(a_i \ b_j)$ son iguales se rechaza sss $CMAB / CME_{\text{error}}(b) > F_{(a-1) \ (b-1) \ a(r-1) \ (b-1) \ (\alpha)}$

3 2 3 Ejemplo Numérico Parcelas Divididas

Un experimento realizado en la Universidad de Wisconsin comparó los rendimientos de cuatro lotes de avena ($a=4$) para tres tratamientos químicos de semilla y un control ($b=4$). Los lotes de semilla fueron asignados al azar a las parcelas grandes dentro de cada replicación.

Los tratamientos de semilla fueron asignados al azar a las sub-parcelas dentro de cada parcela grande. Los rendimientos en bushels por acre son dados en la tabla siguiente.

Lote de Semilla (A)	Replicación	Tratamiento (B)				Totales
		Control (a ₀)	Ceresan M (a ₁)	Panogen (a ₂)	Agrox (a ₃)	
Vicland (1)	1	42 9	53 8	49 5	44 4	190 6
	2	41 6	58 5	53 8	41 8	195 7
	3	28 9	43 9	40 7	28 3	141 8
	4	30 8	46 3	39 4	34 7	151 2
Totales		144 2	202 5	183 4	149 2	679 3
Vicland (2)	1	53 3	57 6	59 8	64 1	234 8
	2	69 6	69 6	65 8	57 4	262 4
	3	45 4	42 4	41 4	44 1	173 3
	4	35 1	51 9	45 4	51 6	184 0
Totales		203 4	221 5	212 4	217 2	854 5
Clinton	1	62 3	63 4	64 5	63 6	253 8
	2	58 5	50 4	46 1	56 1	211 1
	3	44 6	45 0	62 6	52 7	204 9
	4	50 3	46 7	50 3	51 8	199 1
Totales		215 7	205 5	223 5	224 2	868 9
Branch	1	75 4	70 3	68 8	71 6	286 1
	2	65 6	67 3	65 3	69 4	267 6
	3	54 0	57 6	45 6	56 6	213 8
	4	52 7	58 5	51 0	47 4	209 6
Totales		247 7	253 7	230 7	245 0	977 1
Totales de Tratamientos		811 0	883 2	850 0	835 6	3,379 8

Replicación	Totales
1	965 3
2	936 8
3	733 8
4	743 9

- 1
- 1 Cochran William G y Cox Gertrude M Diseños Experimentales Editorial F Trillas México 1974
 - 2 Ching Chun Li Introducción a la Estadística Experimental Ediciones Omega S A Barcelona
 - 3 Davies Owen L (editor) The Design and Analysis of Industrial Experiments Segunda edición Oliver and Boyd London y Hiner Publishing Co New York 1967
 - 4 Draper Norman y Smith Harvey Applied Regression Analysis John Wiley and Sons New York 1966
 - 5 Federer W T Experimental Design Macmillan New York 1955
 - 6 Kempthorne Oscar The Design and Analysis of Experiments R E Krieger Publishing Co Huntington N Y 1973
 - 7 Sredecor G W Métodos Estadísticos Aplicados a la Investigación Agrícola y Biológica Compañía Editorial Continental México
 - 8 Steel Robert G D y Torrie James H Principles and Procedures of Statistics McGraw-Hill Book Company Inc New York 1960

DISEÑOS MAS UTILIZADOS EN EXPERIMENTACION CON FRIJOL (Cont)

4 1 Diseños en Látices

Pertenecen a los diseños en bloques incompletos en los que cada bloque sólo contiene algunos de los tratamientos de este modo el efecto de heterogeneidad de las unidades experimentales se reduce en un grado mayor que con bloques completos al azar Se agrupan en

1 Látices balanceados - El numero de tratamientos debe ser un cuadrado exacto digamos k^2 el numero de unidades por bloque es k y el número de repeticiones es $(k+1)$ Se caracterizan porque todos los pares de tratamiento se comparan (aproximadamente) con la misma precisión por grandes que sean las diferencias entre bloques

2 Látices parcialmente balanceados - Son idénticos a los látices balanceados excepto que tienen menos repeticiones El diseño con 2 repeticiones se llama látice simple y aquel con 3 repeticiones se llama látice triple Cuando la variación entre bloques es grande algunos pares de tratamientos se comparan más precisamente que otros

El análisis estadístico del diseño en latices se basa en el método de máxima verosimilitud que es equivalente a la minimización de una suma de cuadrados ponderada Además no sólo se usa la información intrabloques sino que se recupera información interbloque

4 2 Eficiencia relativa de látices respecto a bloques completos al azar - Una característica de los látices es que pueden ser analizados como bloques completos al azar considerando las repeticiones como los bloques completos. Esta característica permite determinar la eficiencia relativa del diseño en látice con respecto a de bloques completos al azar definida como el cociente del cuadrado medio del error en BCA entre el cuadrado medio del error en látice. Por ejemplo si la eficiencia relativa es 125/ esto significa que 4 repeticiones de un diseño en látice proporcionan tanta precisión como 5 repeticiones en BCA. Si la eficiencia es inferior a 105/ se recomienda usar el análisis en BCA.

4 3 Arreglo del material experimental - Los siguientes criterios son de utilidad

1 Las unidades experimentales dentro del mismo bloque deben ser homogéneas. Para ello hacer los bloques incompletos tan cuadrados como sea posible.

2 Los bloques incompletos dentro de una misma repetición deben ser tan similares entre sí como lo permite el material disponible.

3 Si se hace el análisis del experimento como látice es más importante tener los bloques incompletos homogéneos que tienen homogéneas las repeticiones.

4 Si es muy probable que haya muchas observaciones faltantes

puede ser necesario recurrir al análisis en BCA En tal caso es deseable tener repeticiones homogéneas

4 4 Aleatorización - La aleatorización consiste de tres pasos

1 Aleatoricense los bloques separada e independientemente dentro de cada repeticion

2 Aleatoricense los tratamientos separada e independientemente dentro de cada bloque

3 Designarse al azar los tratamientos a los números de tratamiento

4 5 Tabla del ANOVA para latice balanceado kxk

<u>Fuente de Variación</u>	<u>g l</u>	<u>S C</u>	<u>c m</u>
Repeticiones	k		
Tratamientos	$k^2 - 1$		
Bloques dentro de Rep ajustados por tratamiento	$k^2 - 1$		
<u>Error intrabloque</u>	<u>$(k-1)(k^2-1)$</u>		
Total	$k^3 + k^2 - 1$		

4 6 Tabla del ANOVA para látice parcialmente balanceado

<u>Fuente de Variación</u>	<u>g l</u>	<u>S C</u>	<u>c m</u>
Repeticiones	$r-1$		
Tratamientos (sin ajustes)	k^2-1		
Bloques ajustados por tratamiento	$r(k-1)$		
<u>Error intrabloque</u>	<u>$(k-1)(rk-k-1)$</u>		
Total	$rk^2 - 1$		

CONFERENCIA No 5

REGRESION Y CORRELACION

5 0 Introducción

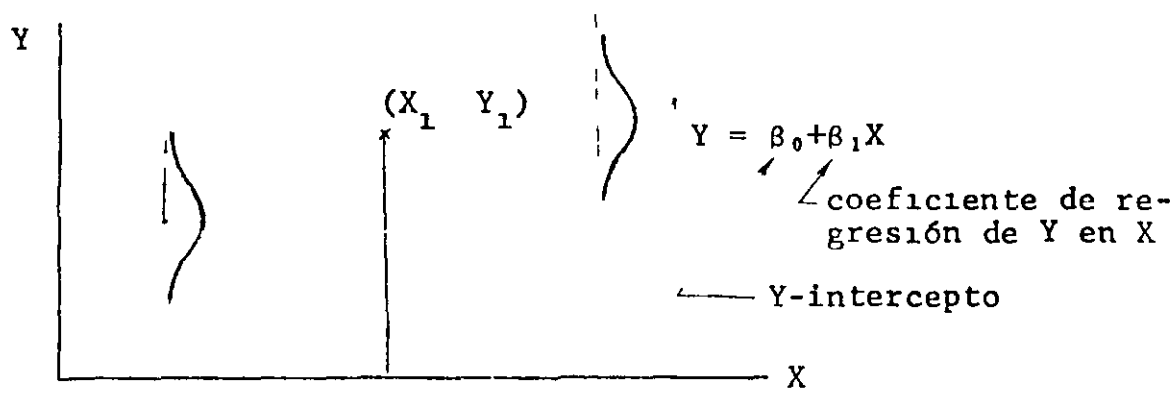
Son técnicas estadísticas que permiten analizar la relación existente entre k variables continuas independientes X_1, \dots, X_k y una dependiente Y a partir de n conjuntos de datos de la forma (X_1, \dots, X_k, Y) correspondientes a una misma unidad experimental. Se exige que las X 's sean estadísticamente independientes pero pueden ser estructuralmente dependientes en el sentido de que la función de respuesta para un factor depende de los niveles de los otros factores.

5 1 Regresion Lineal Simple

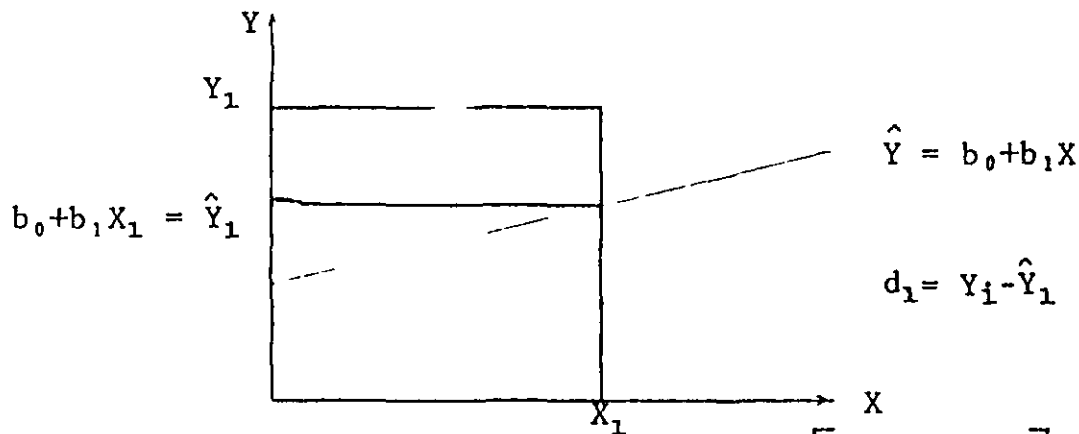
Es el caso más simple de regresión en el cual la relación funcional entre X e Y se asume que es lineal de acuerdo al modelo

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad e_i \sim \text{NID}(0, \sigma^2)$$

El objetivo es estimar los parámetros β_0 y β_1 a partir de n pares (X_i, Y_i) observados



El criterio usado para determinar estimadores de β_0 y β_1 es minimizar la suma de los cuadrados de las desviaciones observadas con respecto a la línea de regresión ajustada a los datos Este método de ajuste es llamado de mínimos cuadrados



$$b_0 = \bar{Y} - b_1 \bar{X}$$

$$b_1 = \frac{\sum (X_1 - \bar{X})(Y_1 - \bar{Y})}{\sum (X_1 - \bar{X})^2}$$

$$\text{Var}(b_0) = \sigma^2 \left[\frac{\sum X_1^2}{n \sum (X_1 - \bar{X})^2} \right]$$

$$\text{Var}(b_1) = \frac{\sigma^2}{\sum (X_1 - \bar{X})^2}$$

$$\text{Var}(\hat{Y}_1) = \sigma^2 \left[\frac{1}{n} + \frac{(X_1 - \bar{X})^2}{\sum (X_1 - \bar{X})^2} \right]$$

$$\text{Var}(b_0 + b_1 X_*) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_* - \bar{X})^2}{\sum (X_1 - \bar{X})^2} \right]$$

ANOVA

<u>Fuente de Variación</u>	<u>g</u>	<u>S C</u>	<u>C M</u>
Reducción debida a la Regresión (Regresión)	1	$b_1 \left\{ \sum X_1 Y_1 - \frac{(\sum X_1)(\sum Y_1)}{n} \right\}$	CMR
Alrededor de la línea ajustada (error)	<u>n-2</u>	<u>Por diferencia</u>	CME
Total corregido	n-1	$\sum Y_1^2 - (\sum Y_1)^2/n$	

CME es un estimador de σ^2

La prueba de la hipótesis nula $H_0 \quad \beta_1 = 0$ Vs

$H_1 \quad \beta_1 \neq 0$ al nivel α

se hace con la siguiente regla de decisión

Rechazar H_0 sss $\frac{CMR}{CML} > F_{1 \ n-2}(\alpha)$

donde $F_{1 \ n-2}(\alpha)$ es el α - percentil superior de la distribución de F con 1 y (n-2) grados de libertad

Coefficiente de Determinación R^2

$R^2 = \frac{SCR}{SCT} =$ proporción de la variación total explicada por la regresión lineal de Y en X

Valores de R^2 cercanos a 0 indican un ajuste pobre valores cercanos a 1 indican un buen ajuste

Coefficiente de Correlación Lineal entre X e Y

$r_{xy} = \pm \sqrt{R^2}$ donde el signo de r es igual al de b_1

$$= \sqrt{\frac{\sum (X_1 - \bar{X})^2}{\sum (Y_1 - \bar{Y})^2}} b_1$$

En general $-1 \leq r_{xy} < 1$

Valores de $|r_{xy}|$ cercanos a cero indican una pobre correlación lineal entre X e Y mientras que valores de $|r_{xy}|$ cercanos a uno indican una alta correlación lineal entre X e Y

r_{xy} mide el grado de asociación lineal entre X e Y mientras b_1 (ambas consideradas como aleatorias) mide el cambio en Y que

puede predecirse cuando X es aumentado en una unidad

5.2 Regresión Lineal Múltiple

Si se tiene más de una variable independiente el modelo es $Y_u = \beta_0 + \beta_1 X_{1u} + \beta_2 X_{2u} + \dots + \beta_k X_{ku} + e_u$ $e_u \sim \text{NID}(0, \sigma^2)$

La ecuación estimada es

$$\hat{Y} = b_0 + b_1 X_1 + \dots + b_k X_k$$

donde b_0, b_1, \dots, b_k son estimados por el método de mínimos cuadrados y pueden obtenerse del computador usando cualquiera de los muchos programas de regresión múltiple existentes. Así mismo la Tabla ANOVA resumen se obtiene del computador y tiene la forma siguiente

<u>Fuente de Variación</u>	<u>g l</u>	<u>S C</u>	<u>C M</u>
R	k	SCR	CMR
<u>E</u>	<u>n-k-1</u>	<u>SCE</u>	<u>CME</u>
T	n - 1	SCT	

Como en el caso de regresión simple CMR es un estimador de σ^2 y

$$H_0: \beta_k = 0 \quad \forall k \quad \forall s$$

$$H_1: \beta_k \neq 0 \text{ para alguna } k$$

se prueba con la regla de decisión

$$\text{Rechazar } H_0 \text{ sss } \frac{\text{CMR}}{\text{CME}} > F_{k, n-1-k}(\alpha)$$

donde $F_{k, n-1-k}(\alpha)$ es el α -percentil superior de la distribución de F con k y (n-1-k) grados de libertad

El Coeficiente de determinación $R^2 = SCR/SCT$ tiene la misma interpretación que para el caso $k = 1$ pero r no tiene interpretación

5 3 Usos de Regresión

- 1 Predecir Y para valores dados de las X 's
- 2 Examinar los efectos de las X s sobre Y (modelos de causalidad s entre las X 's e Y hay una relación de causa a efecto)
- 3 Determinar la forma de la curva de regresión (Regresión Polinomial y Regresión no-lineal)
- 4 Ajustar Y por efectos no controlados cuantificados por las X s (ANCOVA)

5 4 Superficies de Respuesta

Una superficie de respuesta es la función que expresa la respuesta η en términos de los niveles de variables que "explican o predicen η

$$\eta = \emptyset (x_1 \quad x_k)$$

Los X s son estadísticamente independientes pero en general son estructuralmente dependientes en el sentido de que la función de respuesta para un factor depende de los niveles de los otros factores

Generalmente \emptyset es desconocida y se le debe aproximar dentro de la región experimental por un polinomio de las X 's de bajo grado Además de usar las superficies de respuesta para predecir η se le usa para encontrar la dependencia estructural

que puede conducir a un mejor entendimiento del mecanismo básico que produce la respuesta o encontrar la combinación de niveles de las X s que optimiza η

La forma más sencilla de ajustar una superficie de respuesta es ajustando un modelo de regresión lineal múltiple

$$Y_u = \beta_0 + \beta_1 X_{1u} + \beta_k X_{ku} + e_u$$

Si las X s están bajo nuestro control en el sentido de que podemos variarlos a nuestra voluntad, los diseños más simples son los de primer orden es decir aquellos que ajustan un modelo lineal en las X s. Entre estos modelos los más usados son los factoriales 2^k donde cada una de las k variables es probada a 2 niveles

Diseños de Segundo Orden Usan el modelo cuadrático

$$Y_u = \beta_0 + \sum \beta_i X_{iu} + \sum_{i < j} \beta_{ij} X_{iu} X_{ju}$$

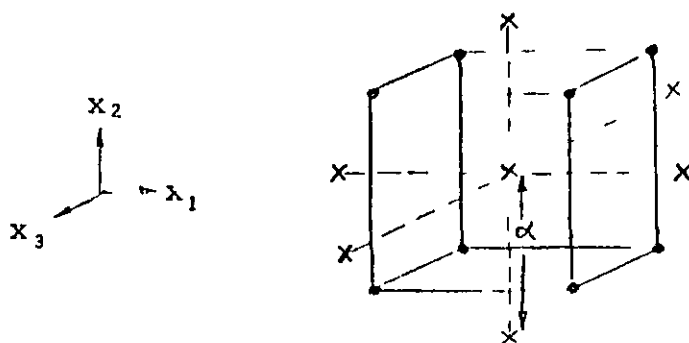
Se requiere que cada X_i deba tomar cuando menos tres diferentes niveles. Sin embargo los factoriales 3^k no son apropiados por su gran magnitud. Por ello se usan los llamados

Diseños Compuestos Centrales Se construyen adicionando al factorial 2^k las siguientes $(2k+1)$ combinaciones

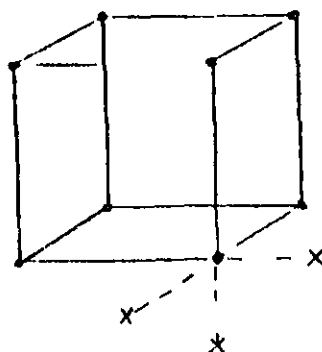
$$\begin{array}{ccccc} (0 & 0 & 0) & (-\alpha & 0 & 0) & (\alpha & 0 & 0) \\ & & & (0 & -\alpha & 0) & (0, & \alpha & 0) \\ & & & (0 & 0 & -\alpha) & (0 & 0 & \alpha) \end{array}$$

El diseño central compuesto puede comenzar con un factorial ex-

plorativo 2^k al cual se le ajusta una superficie de respuesta lineal. Si la 'falta de ajuste' del modelo lineal es evidente se agregan los puntos antes mencionados escogiendo α apropiadamente (haciendo que los β 's sean ortogonales minimizando el sesgo si la verdadera superficie no es cuadrática o exigiéndole que sea rotacional)



Diseño Central Compuesto



Diseño Compuesto No central

Ilustración de los dos tipos de diseño compuesto para tres factores

• = Puntos de un factorial inicial 2^3

x = Puntos adicionales agregados para el diseño compuesto

Si el análisis preliminar con 2^k puntos sugiere que el punto de máxima respuesta esta más cercano a alguna de las otras combinaciones de factores que al centro, es aconsejable usar un diseño compuesto no central

Uno de los criterios a satisfacer por parte de un buen diseño es el de rotabilidad que consiste en que el error estándar de la respuesta estimada debe ser el mismo para todos los puntos que están a una misma distancia del centro de la región. Los diseños rotacionales más usados son los llamados diseños rotacionales centrales compuestos que consisten de 3 tipos de puntos

- a) Puntos del factorial 2^k
- b) $(2k+1)$ puntos adicionales en estrella para formar el compuesto [$c=2^{k/4}$ para factorial completo]
- c) n_1 puntos en el centro n_1 se elige de modo que la precisión de estimación sea mas o menos constante dentro de una 'esfera' de radio 1

Para el análisis estadístico de estos diseños y los métodos para determinar la combinación óptima de los niveles de los factores ver Cochran Capitulo 8-A

Ejemplo de Regresión Lineal Simple

Datos observados de caída de lluvia (X_1) y Rendimiento de trigo (Y_1) en una zona durante 10 años

X_1 (mm)	Y_1 (Kgr /Ha)
230	2600
210	2500
280	2900
270	2700
230	2700
280	3200
270	3300
220	2800
260	3000
250	3300

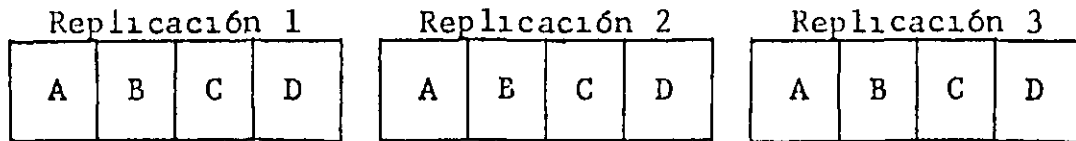
DISEÑOS SISTEMÁTICOS

Diseños Sistemáticos

Los diseños experimentales pueden dividirse en dos grupos aleatorios y sistemáticos. Los diseños que hasta ahora hemos tratado son aleatorios y se caracterizan porque la asignación de los tratamientos a las unidades experimentales se hace en una forma perfectamente al azar.

En contraposición con ellos se encuentran los diseños sistemáticos en los cuales la asignación de tratamientos a las unidades experimentales se efectúa en forma ordenada o sistemática. El objetivo de este tipo de diseños es permitirle al investigador observar una respuesta continua al tratamiento. Por ejemplo si se desea estudiar la respuesta de una variedad de frijol al nitrógeno se puede diseñar un experimento que consista en administrar distintas dosis de N al suelo en forma creciente y medir el rendimiento de la (o las) plantas que reciban el respectivo tratamiento.

Antes del desarrollo del diseño experimental moderno esto es antes de que Fisher introdujera el principio de aleatorización en la asignación de tratamientos a las parcelas experimentales un ordenamiento sistemático de los tratamientos en cada replicación parecía muy natural. Uno de los tipos más comunes de arreglo sistemático es aquel en el cual el ordenamiento de los tratamientos es exactamente el mismo en cualquiera de las repeticiones como se aprecia en la gráfica



Muchos otros diseños sistemáticos han sido desarrollados sin embargo todos presentan relativamente las mismas desventajas con respecto a los diseños aleatorios y son

- 1 Las diferencias detectadas entre tratamientos pueden contener un error sistemático debido a la correlación entre parcelas adyacentes
- 2 No son eficientes cuando el área experimental es muy heterogénea pues no permiten un estimativo válido de la varianza

Las ventajas son

- 1 Simplicidad
- 2 Permiten un ordenamiento de los tratamientos Por ejemplo, las variedades pueden ordenarse según su madurez los fertilizantes en orden de su eficacia etc
- 3 La respuesta al tratamiento se puede apreciar en forma continua

Como ejemplo de algunos diseños sistemáticos mencionaremos los utilizados para experimentación en yuca en el CIAT

- 1 Superficies de respuesta
- 2 Diseño de abanico
- 3 Diseño en surcos paralelos

Superficie de Respuesta

Cuando se desea estudiar el efecto de uno o más factores

- x_1 x_2 x_3 ... x_n que representan variables continuas como tiempo cantidad de nitrógeno temperatura etc es natural pensar en los rendimientos o respuesta y como una función de los niveles de estas variables Esto es

$$y = f(x_1, x_2, x_3, \dots, x_n) + \epsilon$$

donde ϵ representa el error experimental

La función f se denomina "superficie de respuesta"

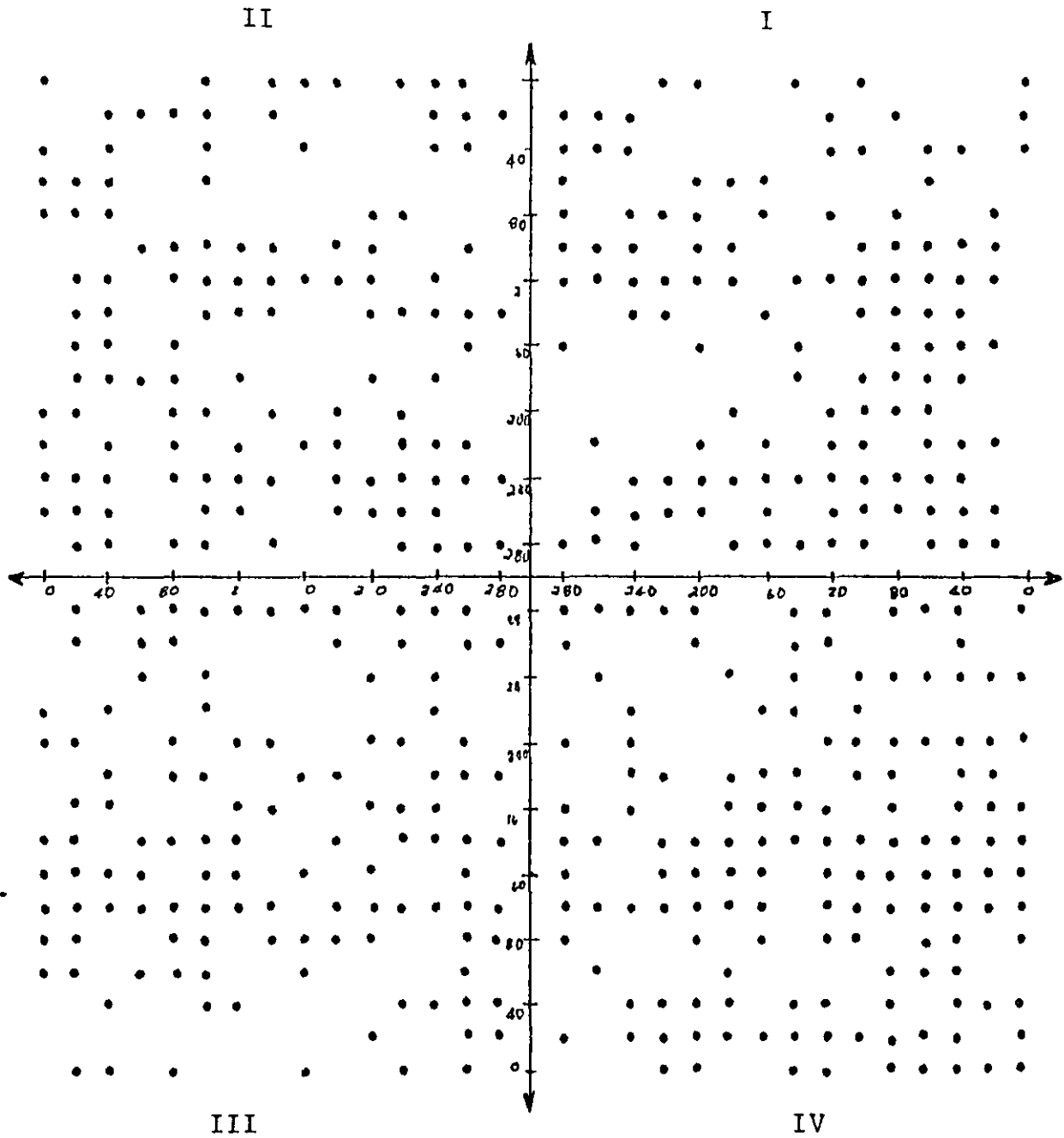
Un conocimiento de f da un resumen completo de los resultados del experimento y permite predecir las respuestas para determinados valores de los factores x_1

Ejemplo 'Efecto de N y K sobre el rendimiento de la planta de

Yuca Se desea medir el efecto de 16 niveles de nitrógeno 0 20 40 60 80 300 gr/planta y 16 niveles de potasio 0 20 40 60 300 gr/planta sobre el rendimiento de una variedad de yuca (medida en peso fresco de raíces) Las observaciones se hacen sobre plantas individuales

Se sembraron las plantas a una distancia de 80cm y se aplicaron los niveles de N y K en la forma que muestra la gráfica de tal manera que cada planta estuviera expuesta a una determinada combinación de NxK Cada cuadrante corresponde a una replicación El número de tratamientos por replicación que corresponde al número de plantas es de $16 \times 16 = 256$

DISÑO SISILMÁICO N x K EN YUCA EN 4 REPLICACIONES



Cada cuadrante representa una replicación completa del diseño con 256 plantas por replicación. Cada punto es una planta individual y recibe una de las 256 combinaciones de Nitrogeno por Potasio.

La respuesta de la yuca al N y al K se puede expresar mediante la siguiente superficie de respuesta

$$Y_{1j} = a_0 + a_1 N_1 + a_2 K_j + a_3 N_1 K_j + a_4 N_1^2 + a_5 K_j^2 + \epsilon_{1j}$$

Rendimiento de la planta con nivel 1 de N y nivel j de K

que mide el efecto tanto lineal como cuadrático de N y de K y el de la interacción NxK y corresponde a un modelo de regresión cuadrática

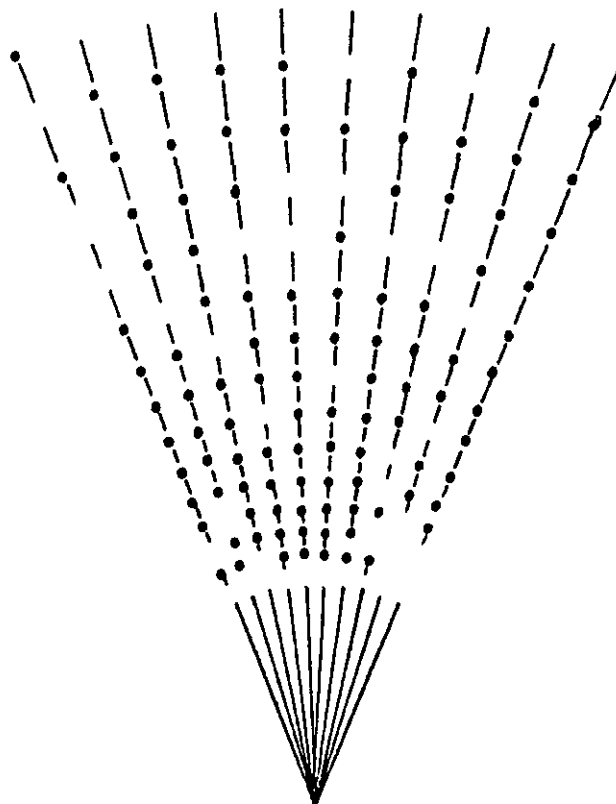
Diseños en Abanico y Surcos Paralelos^{1/}

Estos dos diseños se usan básicamente para medir el rendimiento de distintas variedades bajo un amplio rango de densidades de población. El número de plantas por unidad de área varía sistemáticamente de una parcela a otra pero el arreglo de las plantas se mantiene constante. Cualquier rango de densidades puede ser probado.

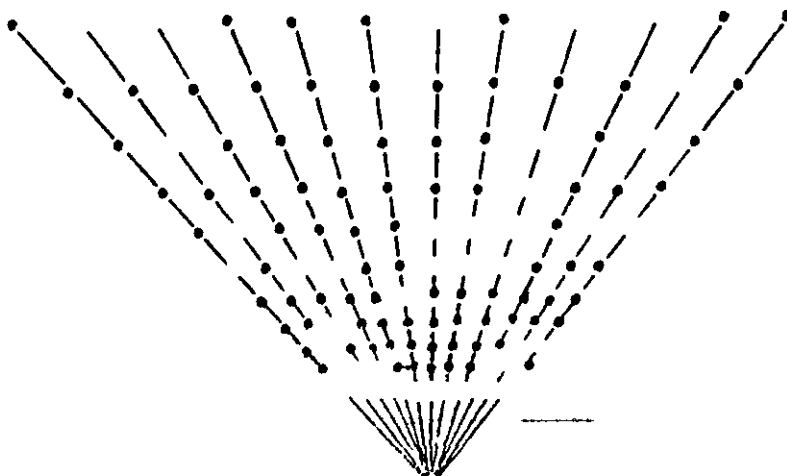
En las gráficas que aparecen a continuación se puede apreciar la disposición de las plantas en el campo bajo el diseño en Abanico y de Surcos Paralelos respectivamente para una sola variedad.

^{1/}

DISPOSICION DE LAS PLANTAS EN UN DISEÑO DE ABANICO PARA
PROBAR 14 DENSIDADES



DISPOSICION DE LAS PLANTAS EN UN DISEÑO DE SURCOS PARALELOS
PARA PROBAR 10 DENSIDADES



In el diseño de abanico las plantas se siembran en filas que irradian de un punto central de tal manera que la distancia entre plantas a lo largo de un radio sea aproximadamente igual a la distancia entre los radios en ese punto. Cada arco corresponde a un distinto nivel de densidad de población. Cuando se desea probar más de una variedad se repite este arreglo en otra sección circular manteniendo entre dos variedades contiguas plantas de borde o un espaciamento adecuado a lo largo de los radios laterales.

Para medir la respuesta del rendimiento a las distintas densidades de población se puede ajustar una función,

$$P_{1j} = f(D_j)$$

Rendimiento de la 1-ésima planta sembrada bajo densidad j
 Nivel j-ésimo de densidad

que puede ser o no lineal y encontrar cuál es la densidad que produce el máximo rendimiento.

En el diseño de Surcos Paralelos cada fila corresponde a un distinto nivel de densidad de población. El número de plantas por fila se mantiene constante pero la distancia entre filas varía de forma sistemática.

La forma de análisis es similar a la utilizada en el caso del diseño de abanico.

ESTUDIO DEL DISEÑO EXPERIMENTAL SELECCIONADO PARA EL PROYECTO
VIVERO INTERNACIONAL DE RENDIMIENTO DE FRIJOL LÁTICE (5x5)^{1/}

I INTRODUCCION

El propósito de este trabajo es estudiar la eficiencia y precisión del diseño de Látice (5x5) bajo distinto número de repeticiones y diferentes tamaños de parcela experimental (número de surcos por parcela) en un ensayo de rendimiento de 25 variedades de frijol negro. Como medida de eficiencia utilizaremos la Eficiencia Relativa (ER) del diseño en Látices con respecto al diseño en Bloques Completos al Azar (BCA) considerando que el Látice es más eficiente cuando su ER excede a 105/

Como índice de precisión utilizaremos el Coeficiente de Variación (CV) y la Diferencia Mínima Significativa (DMS), siendo esta última un índice de la magnitud de las diferencias entre tratamientos que pueden ser detectadas por el diseño y el primero una medida de la variabilidad no controlada.

De esta manera estaremos en capacidad de predecir qué CV y qué magnitud de DMS puede esperarse al utilizar como diseño experimental un Látice de (5x5) con un número dado en repeticiones y un cierto tamaño de parcela experimental. Se estudiará el Látice con 2, 3, 4, 5 y 6 repeticiones y con parcelas de 1, 2, 4 y 6 surcos de 3m de largo siendo el área de parcela respectivamente igual a 1,5, 2,0, 6,0, y 9,0 m².

Este trabajo fué motivado por el Proyecto Vivero Interna-

^{1/}

Basado en trabajo presentado en la XVIII Reunión Anual del PCCICA Panamá, 11 de Julio de 1971.

cional de Rendimiento en Frijol' (VIRF) que tiene a su cargo el Programa de Agronomía de Frijol del CIAT desde mediados de 1976 y cuyo objetivo es estudiar el comportamiento de 25 variedades en un amplio rango de ambientes. Para el primer año de realización del VIRF se instalaron 90 ensayos en 30 países y se espera un número igual o mayor de ensayos por país para el segundo año (1977-1978) disponerse de mayor cantidad de material a probar. El diseño experimental único utilizado en cada localidad ha sido el de Látices de (5x5) con 4 repeticiones, con parcelas de 12m² de área total y 6m² de área útil. Esta dimensión para el área útil de parcela obedece a los resultados de un trabajo previamente realizado en la Unidad de Biometría del CIAT (5). El área que cubre cada ensayo es de 1 800m² aproximadamente.

Dado que estudios realizados por la Unidad de Biometría del CIAT (4) muestran que el diseño en Látices es entre 2 y 84% más eficiente que el diseño en BCA según las condiciones de heterogeneidad del suelo que las diferencias entre variedades que se desearía detectar como significativas en el VIRF oscilan entre 350 y 400 Kg/Ha y tomando en cuenta que la mayor restricción con la que tropiezan los investigadores involucrados en el VIRF es la escasez de terreno hemos considerado de suma importancia realizar este estudio que permite encontrar una combinación de tamaño de parcela y número de repeticiones que se ajuste a las necesidades futuras del Vivero.

II MATERIALES Y METODOS

2 1 Materiales y Métodos de campo

El ensayo se realizó en la granja experimental del CIAI-Palmira (Colombia) con temperatura promedio de 24 C altura sobre el nivel del mar de 1000m y precipitación promedio anual de 1000mm

Se probaron 25 variedades de color negro y hábitos de crecimiento II y III (ver tabla 1) En ensayo fué sembrado bajo un diseño de Látice (5x5) balanceado (con 6 repeticiones) que cubrió un área total aproximada de 3000m²

Se consideró como unidad experimental una parcela de 12m² de área total constituida por 6 surcos de 4m de largo (3 camas con 2 surcos por cama y 1m de distancia entre camas) (ver figura 1)

La cosecha se realizó por surcos denominados A B C D E y F dejando bordes de 0.5m en las cabeceras de tal manera que el area útil por parcela fué de 9m²

2 2 Metodología de Análisis

Se evaluó el diseño de Látice (5x5) bajo las siguientes combinaciones de tamaño de parcela y número de repeticiones

Tamaño de parcela

- De 1 surco (Surco C)
Area de parcela = 1.5m²
- De 2 surcos (Surcos C y D)
Area de parcela = 3m²
- De 4 surcos (Surcos B C, D y E)
Area de parcela = 6m²
- De 6 surcos (Surcos A B, C D E y F)
Area de parcela = 9m²

TABLE 1 VARIETIES UTILIZADAS

Numero	Clave	Nombre	Habito de Crecimiento
1	P9	Pecho Amarillo	II
2	P14	PI 310 909	II
3	P199	PI 196 927	II
4	P209	PI 201 333	II
5	P225	P 207 130	II
6	P226	PI 207 198	II
7	P320	PI 310 686	II
8	P322	PI 310 724	III
9	P337	PI 310 797	II
10	P349	PI 311 930	II
11	P422	Compuesto Negro	II
12	P437	Frijol Negro	II
13	P443	Frijol Negro C	II
14	P459 B	Jamapa	II
15	P459 C	Jamapa	II
16	P481	N 257	II
17	P491	Puebla 41 1	III
18	P509	San Pedro P 72	II
19	P529	Trujillo 7	III
20	P566	Porrillo Sintético	II
21	P579	II 313 868	II
22	P667	C - 166 -N	II
23	P668	C - 168 -N	II
24	P675	ICA - Pijao	II
25	P700	NEGRO CIP 72	II

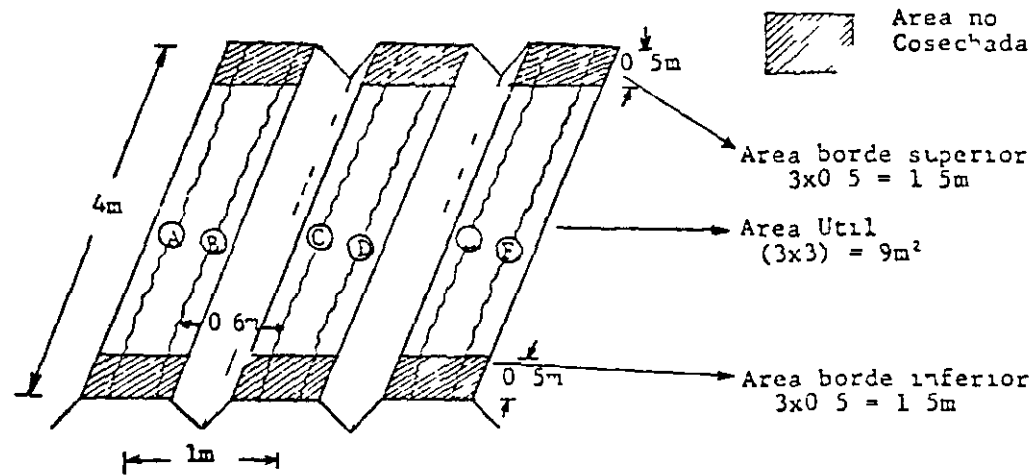


FIGURA 1 ESQUEMA DE LA PARCELA EXPERIMENTAL

y número de repeticiones	2	(3 Grupos	I y II	III y IV	^{1/}
	3	(2 Grupos	I	II y III	IV V y VI)
	4	(3 Grupos	I	II	III IV III IV
			V	VI	I II V VI)
	5	(1 Grupo	I	II	III IV y V)
	6	(1 Grupo	I	II	III IV V y VI)

Los grupos de repeticiones fueron escogidos de tal manera que cada repetición apareciera el mismo numero de veces La única excepción es el caso de 5 repeticiones donde no aparece la VI Como parcela de 1 surco se eligió la parcela constituida por el surco C (ver fig 1) por ser éste junto con el D los más representativos de parcelas de 1 surco sujetos a competencia uniforme El mismo criterio se siguió en la elección de parcelas de dos y cuatro surcos

En cada caso se efectuaron análisis de varianza para rendimiento (en Kg/Ha) según la disposición de fuentes de variación y grados de libertad correspondientes al diseño de Lá-tices (5x5), así

Fuente de Variación	g l	CM
Repeticiones	r-1	CMR
Bloque incompleto dentro de repetición (ajustado)	4r	CMB
Variedades (sin ajustar)	24	CMV
Error intra-bloque	20r-24	CME
Total	25r-1	
Error correspondiente al BCA	24(r-1)	CME _{BCA}

1/

I II III IV V y VI corresponden al número de la repetición

Se calculó el CV la DMS y la ER del látice con respecto al BCA para cada combinación

donde

$$CV = \frac{S}{\bar{X}} \times 100 \quad \text{siendo } S = \sqrt{CM_E} \text{ y } \bar{X} = \text{media general}$$

$$DMS = t_f \sqrt{\frac{2CM_L}{r}} \quad \text{siendo } f = \text{grados de libertad del error} \\ (20r-24)$$

n = número de observaciones por tratamiento

$$ER_L \text{ BCA} = \frac{CM_{LBCA}}{CM_{E'}} \quad \text{siendo } CM_{E'} = \text{cuadrado medio del error efectivo en Látices } 1/$$

III RESULTADOS Y DISCUSION

La Tabla 2 muestra los valores de CV DMS y ER_L BCA para las distintas combinaciones de número de repeticiones y tamaño de parcela (numero de surcos por parcela) Los datos en ella consignados corresponden a promedios entre los distintos grupos con igual número de repeticiones Así por ejemplo el CV que corresponde a 2 repeticiones con parcelas de 1 surco (24 33/) fué obtenido como el promedio de los CV respectivamente hallados para parcelas de 1 surco en los 3 grupos de 2 repeticiones (I y

$$\frac{1/}{CM_{L'}} \text{ para Látices balanceados} = CM_L (1+k\mu) \text{ con } k = \sqrt{\frac{\text{No de tratamientos}}{r-1}}$$

$$\mu = \frac{CM_B - CM_E}{k^2 CM_B}$$

$$CM_E \text{ para Látices no balanceados} = CM_L \left(\frac{1+k\mu}{k+1} \right) \quad \mu' = \frac{CM_B - CM_E}{k(r-1)CM_B}$$

II III y IV V y VI)

Las figuras 2 y 3 presentan la relación entre el CV y la DMS respectivamente con el número de surcos por parcela para 2 3 4 5 y 6 repeticiones del Látice (5x5) Los datos utilizados para estas gráficas corresponden a los promedios presentados en la Tabla 2

Como se aprecia en la Figura 2 el CV decrece a medida que aumenta el tamaño de la parcela para cualquier número de repeticiones Los valores del CV son siempre superiores a 20/ con parcelas de 1 surco (su valor oscila entre 21 y 29/) pero menores ó iguales a 12/ con parcelas de 2 4 y 6 surcos El número de repeticiones no tiene ningún efecto en lo que respecta al CV

La Figura 3 muestra las diferencias en rendimiento (DMS en Kg/Ha) que el Látice (5x5) puede detectar al usar distintos tamaños de parcela y 2 3 4 5 ó 6 repeticiones En ella se observa que la DMS sí es sensible al número de repeticiones que se utilicen Con 2 la diferencia en rendimiento detectada en promedio es de 604 7 Kg/Ha a medida que se aumenta el número de repeticiones el diseño es capaz de detectar diferencias menores cuando el Látice (5x5) es balanceado es decir cuando utiliza 6 repeticiones puede detectar en promedio diferencias de 343 8 Kg/Ha (ver Tabla 2)

La DMS es también muy sensible al tamaño de parcela El Látice detecta diferencias más pequeñas a medida que se incre-

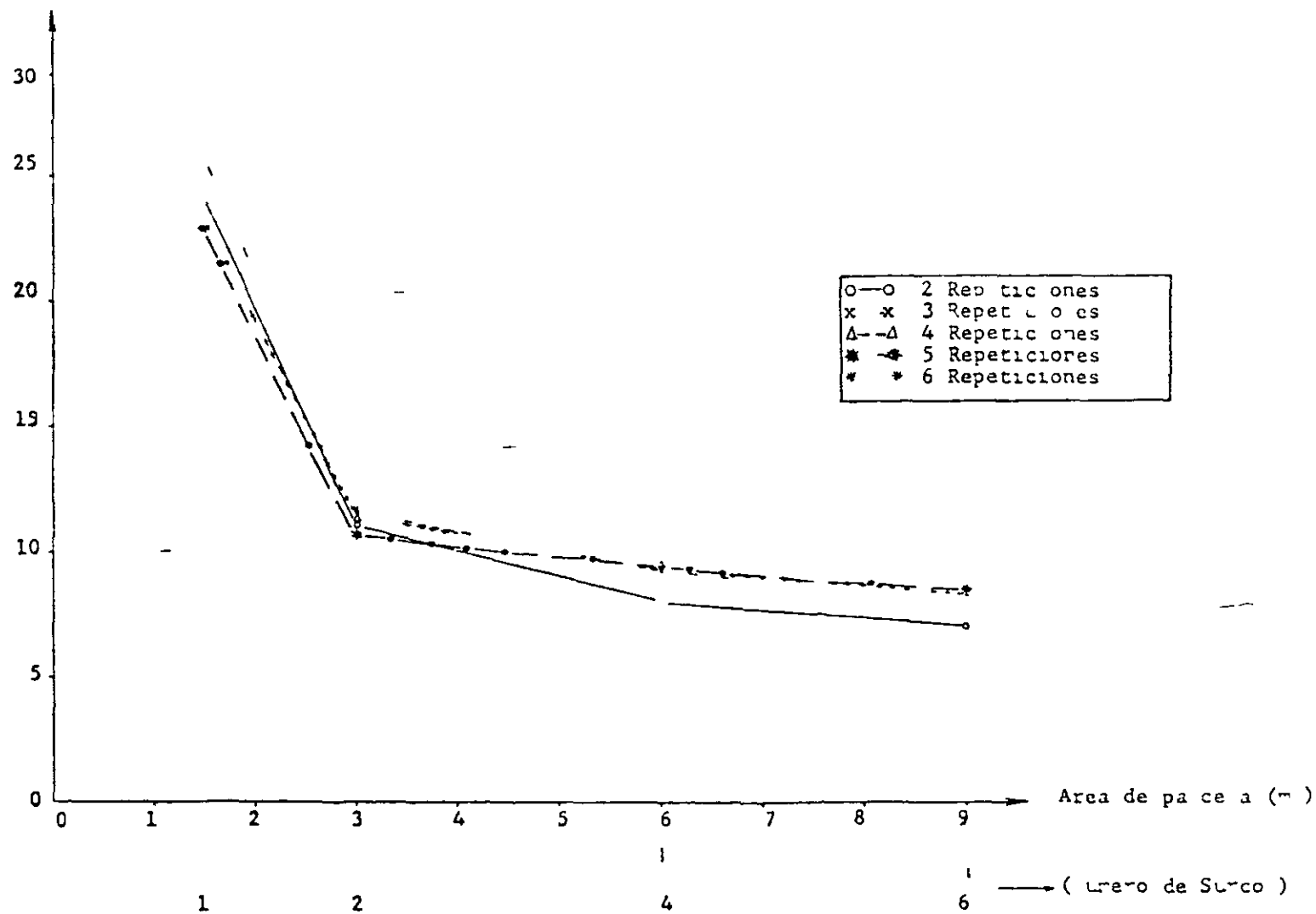


FIGURA 2 GV EN FUNCION DEL AREA DE PARCELA (No. de Surcos) A UN DETERMINADO NUMERO DE REPETICIONES EN UN LAT. C. (5x5)

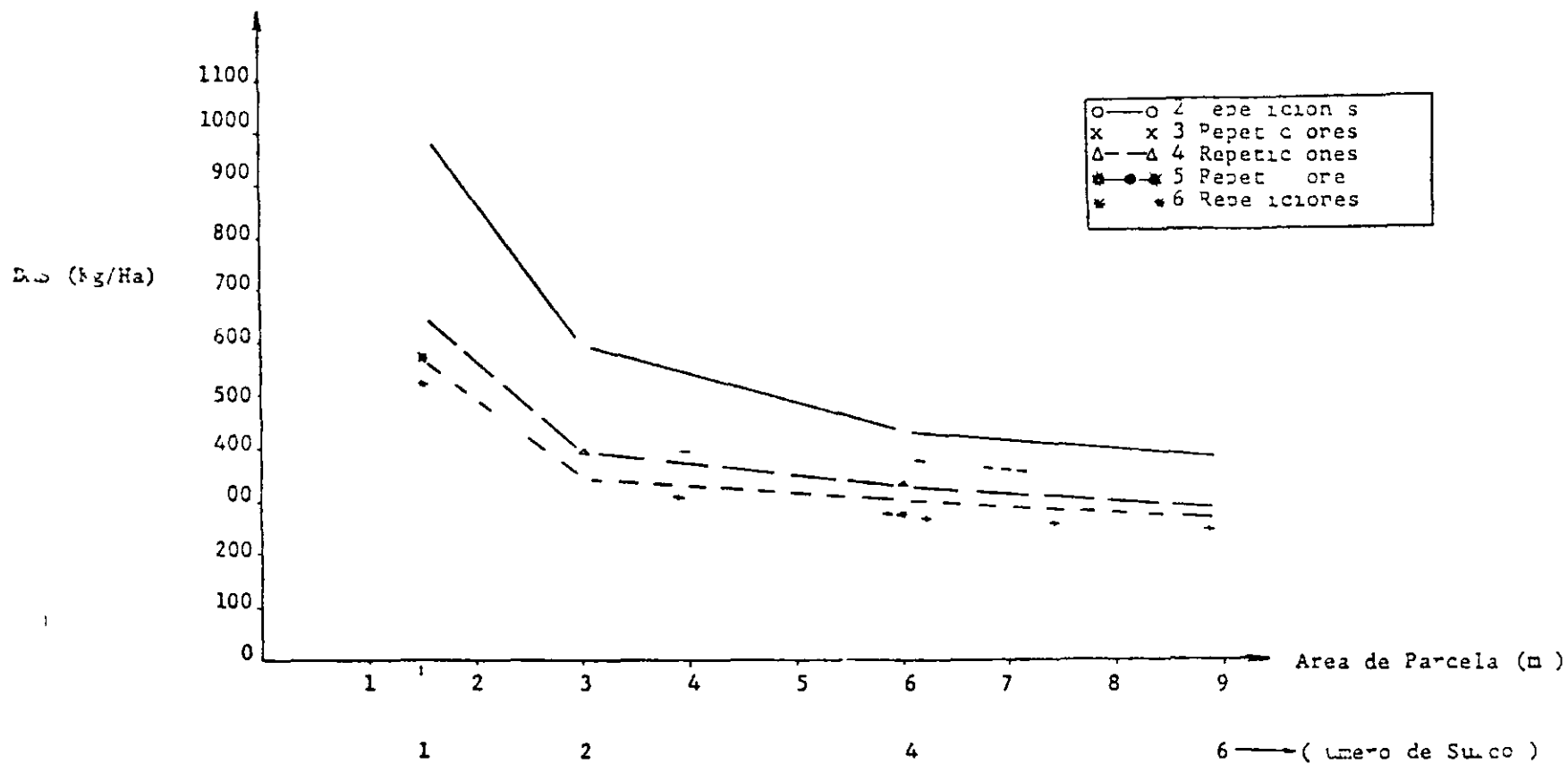


FIGURA 3 DMS EN FUNCION DEL AREA DE PARCELA (o de Surcos) PARA DISTINTO NUMERO DE REPETICIONES EN UN ATICE (5x5)

menta el número de surcos por parcela (Ver Tabla 2 y Figura 3)

En general podemos afirmar que el número de repeticiones que se utilicen afecta a la DMS mientras que el tamaño de parcela experimental afecta tanto a la DMS como al CV

Sin embargo en cuanto a la ER del Látice con respecto al BCA no se observa ninguna tendencia definida al variar el número de repeticiones o el tamaño de la parcela experimental. Lo único que podemos afirmar es que bajo todas las condiciones el diseño en Látice supera o es igualmente eficiente que el diseño en BCA

BIBLIOGRAFIA

- 1 Cochran y Cox G Diseños experimentales Editorial Tri-
llas 1965 p 416-469
- 2 Hatheway W H and Williams E J Efficient estimation of
the relationship between plot size and the variabili-
ty of crop yields Biometrics June 1958 p 207-222
- 3 LeClere E Leonard, W and Clark A Field plot techni-
que Burgess Publishing Company 1962 373 p
- 4 Muñoz J E y Amézquita M C Eficiencia relativa del Dise-
ño en Látices con respecto al Diseño en Bloques al
Azar XXII Reunión PCCMCA San José Costa Rica
1976
- 5 Muñoz J E Salazar L C López Y Determinación del ta-
maño forma y número de repeticiones más adecuadas
en ensayos de rendimiento en Fríjol (Phaseolus vulga-
ris L) y comparación de dos métodos para estimar su
rendimiento comercial 1975

CONFERENCIA No 8

SISTEMA DE INFORMACION DE FITOMEJORAMIENTO DE FRIJOL SIFFRI

El banco de germoplasma para frijol contiene la mayoría del material genético disponible y está constituido por las semillas provenientes de colecciones introducciones y selecciones Dentro de este grupo iste un subconjunto que contiene los materiales usados como padres en las cruzas o que potencialmente podrían usarse como padres La selección que llega al banco de germoplasma lo hace despues de haber pasado en las generaciones tempranas y avanzadas las evaluaciones que le catalogan como promisoria

El sistema de información tiene como su propósito básico mantener e integrar los datos producidos por las diferentes disciplinas e informar sobre el flujo de materiales a través del banco y sobre las acciones que generan ese flujo

Hasta ahora el sistema ha satisfecho algunas necesidades que se consideran básicas tales como

- 1 Conocer todas las cruzas y selecciones que vienen de un padre
- 2 Conocer los padres ó cruzas con combinaciones de características o criterios de selección
- 3 Reconstruir la genealogia de cruza a padres
- 4 Informes sobre datos faltantes
- 5 Elaboración de etiquetas para las cruzas que van a usarse en el libro de campo

6 Análisis estadístico

El sistema de información consta de un grupo de archivos y programas interrelacionados tal que satisfacen las consultas sobre material genético

Los archivos (ó relaciones) son

Banco de Geoplasma

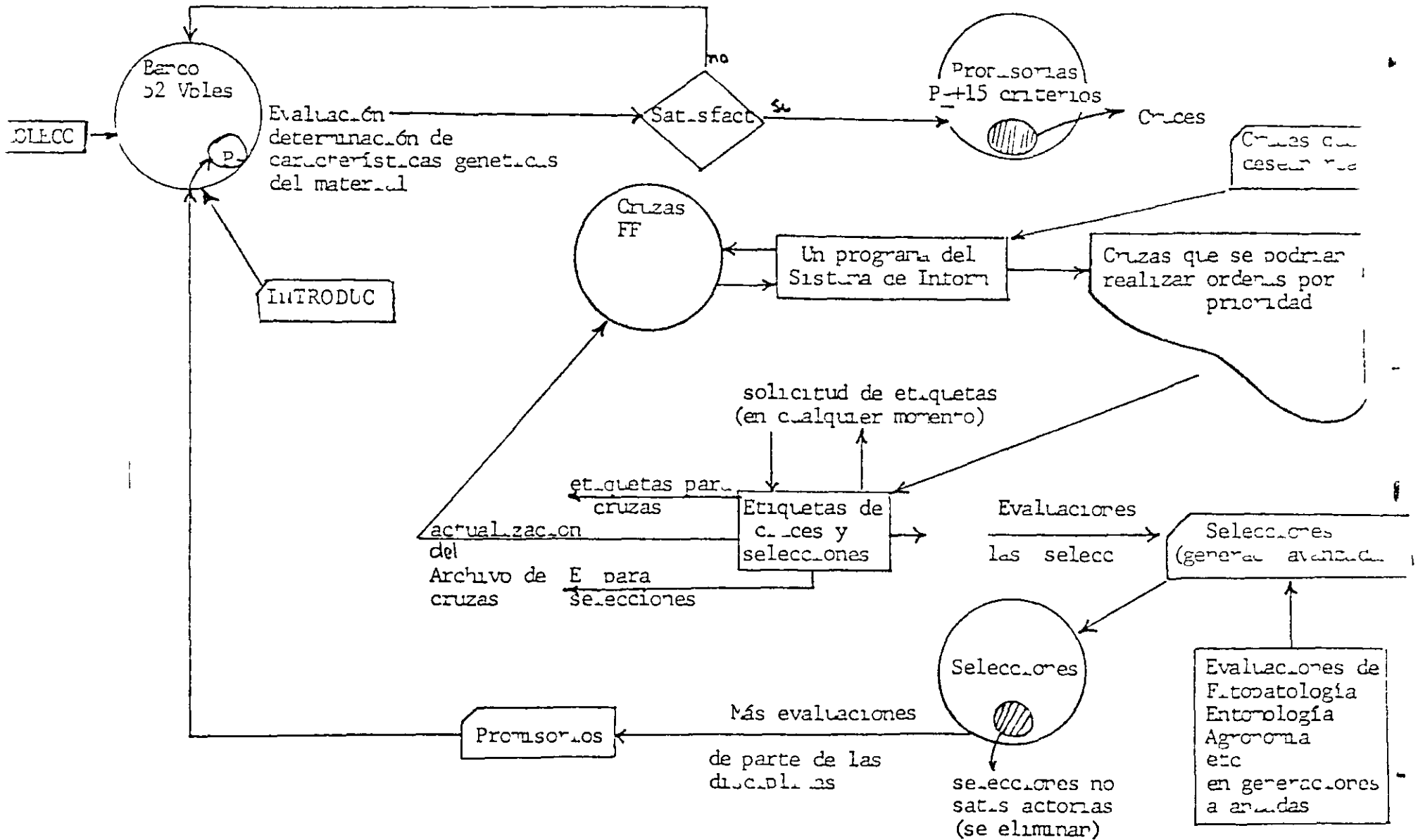
Promisorios

Cruzas

Selecciones (por formarse)

Un diagrama detallado del sistema se halla a continuación es nuestro interés ampliar la cobertura hasta recolectar todos los datos que se produzcan en la experimentación con frijol

Diagrama detallado del sistema



BIBLIOGRAFIA

- 1 Cochran William G y Cox Gertrude M Diseños Experimentales Editorial F Trillas México 1974
- 2 Ching Chun I. Introducción a la Estadística Experimental Ediciones Omega S A Barcelona
- 3 Davies Owen L (editor) The Design and Analysis of Industrial Experiments Segunda edición Oliver and Boyd London y Hafner Publishing Co New York 1967
- 4 Draper Norman v Smith Harvey Applied Regression Analysis John Wiley and Sons New York 1966
- 5 Federer W T Experimental Design Macmillan New York 1955
- 6 Kempthorne Oscar The Design and Analysis of Experiments R E Krieger Publishing Co Huntington N Y 1973
- 7 Sredecor G W Métodos Estadísticos Aplicados a la Investigación Agrícola y Biológica Compañía Editorial Continental México
- 8 Steel Robert G D y Torrie James H Principles and Procedures of Statistics McGraw-Hill Book Company Inc New York 1960