MANUAL ESTADISTICO PARA LA EXPERIMENTACION EN ARROZ

Jaime Eduardo Muñoz.

MANUAL ESTADISTICO PARA LA EXPERIMENTACION EN ARROZ

Por:

Jaime Eduardo Muñoz

ESTADISTICA

PROPOSITO:

"PROPORCIONAR UNA BASE OBJETIVA PARA EL ANALISIS DE PROBLEMAS EN LOS CUALES LOS DATOS SE APARTAN DE LAS LEYES DE LA CAUSALIDAD EXACTA, LO CUAL ES ESPECIALMENTE CIERTO PARA LAS CIENCIAS AGRICOLAS Y BIOLOGICAS"

FINNEY

Introducción:

Con estos apuntes se pretende dar una visión general sobre la estadística y el uso de los Diseños experimentales en el cultivo de arroz. Inicialmente se da algunos elementos de estadística descriptiva y estadística inferencial y por último se describen con ejemplos los diseños completamente al azar, bloques al azar, con un solo factor y varios factores y finalmente el diseño en parcelas divididas.

Este pequeño manual se pudo realizar por la colaboración brindada por los miembros de la Unidad de Servicios de Datos del CIAT, por el Programa de Arroz, y las enseñanzas del profesor Jorge A. Escobar.

INDICE DE CONTENIDO

1. <u>INTRODUCCION A LA ESTADISTICA</u>:

Para una mejor comprensión de los temas que se desarrollarán, se darán algunos conceptos básicos:

1.1. <u>Estadística</u>:

"Es una ciencia pura y aplicada, que crea, desarrolla y aplica técnicas, que permitan colectar, analizar, interpretar datos, y tomar decisiones en casos de incertidumbre". La estadística <u>descriptiva</u>, como su nombre lo indica, solo sirve para describir poblaciones o muestras y la estadística <u>inferencial</u> permite tomar decisiones en casos de incertidumbre con una probabilidad de error.

1.2. <u>Variable</u>:

Es una característica que varia, por ej.: Producción de arroz, color de hojas, número de macollas por planta. Las variables pueden ser <u>cualitativas</u> o <u>cuantitativas</u>, las primeras se refieren principalmente a cualidades como color, sabor, etc. y las segundas son aquellas que se pueden medir o cuantificar, como rendimiento o altura de planta. Las variables también se pueden clasificar en: <u>Continuas</u> o <u>discretas</u>.

Contínuas son aquellas que pueden tomar cualquier valor en un rango determinado, por ejemplo, la producción de arroz a nivel experimental puede variar entre 1000 y 10000 kilos/ha., las variables discretas, solo pueden tomar un número entero de valores por ej.: el número de macollas por sitio puede ser 15, 20, o el número de plantas de arroz por parcela puede ser 150, 160, etc.

Otra clasificación de las variables es: Variables dependientes y variables independientes; estas últimas toman uno u otro valor dependiendo de los valores que tome la variable independiente, así por ejemplo. La producción de arroz (variable dependiente) depende de la cantidad de nitrógeno que se haya aplicado al suelo.

1.3. <u>Población física</u>:

Conjunto de individuos con características comunes, por ej.: Las plantas de arroz de la variedad CICA6.

1.4. <u>Población estadística</u>:

Conjunto de observaciones o mediciones de una población física, por ej.: Las alturas de las plantas de arroz. De una población física se pueden generar varias poblaciones estadísticas, así de las plantas de arroz, se pueden tener además de las alturas, los rendimientos de grano, las áreas foliares y otras.

1.5. <u>Muestra</u>:

Parte de una población física o estadística. <u>Razones por las cuales se forman muestras</u>.

a) Porque es imposible o resulta antieconómico tomar a toda la población

b) Porque las muestras se pueden destruir, así por ej., si se va a determinar el contenido de Nitrógeno en las hojas de una variedad de arroz, hay que destruir las hojas.

c) Porque con muestras se pueden obtener buenos resultados.

1.6. Usos de la estadística en la investigación biológica:

Los fenómenos biológicos tienen la característica de ser variables y no se puede predecir con exactitud el resultado de un evento dado, en forma individual así, no se puede saber si el sexo de un niño proximo a nacer va a ser masculino o femenino, pero si se puede decir que: "en una gran cantidad de nacimientos, aproximadamente el 50% de ellos tendrá el sexo masculino" es decir que la probabilidad de que nazca niño o niña es 0.5 o 50%.

De igual manera al sembrar una variedad de arroz en un lote, no se puede determinar la producción, pero si puede establecerse el rango dentro del cual estaran las producciones, haciendo uso de registros anteriores, por ej,: Se espera una producción entre 4 y 5 ton/ha en 9 de 10 casos, dado que entre estos dos valores se ha encontrado el 90% de las producciones del lote.

Por esta razón es aplicable el siguiente concepto: "EL PROPOSITO DE LA CIENCIA ESTADISTICA ES PROPORCIONAR UNA BASE OBJETIVA PARA EL ANALISIS DE PROBLEMAS EN LOS CUALES LOS DATOS SE APARTAN DE LAS LEYES DE LA CAUSALIDAD EXACTA, LO CUAL ES ESPECIALMENTE CIERTO PARA LAS CIENCIAS AGRICOLAS Y BIOLOGICAS"[1/] Ante la variabilidad de los fenómenos, la estadística proporciona herramientas muy útiles, para poder tomar decisiones con niveles altos de confiabilidad, por ej.: Se puede llegar a decidir si un sistema de siembra (transplante) en arroz, supera o no a los sistemas tradicionales, con un margen de probabilidad dado.

2. CARACTERIZACION DE UNA POBLACION:

Cuando se obtienen datos de una muestra o de una población física, es muy díficil sacar conclusiones, si el número de observaciones es muy grande, pudiéndose aplicar en este caso la frase "Los árboles no dejan apreciar el bosque", es decir que las observaciones miradas en forma individual dan muy poca información sobre el comportamiento de la población. Para caracterizar una población son de mucha utilidad las tablas de frecuencia, las medidas de tendencia central y las medidas de dispersión.

1/ Finney, J.D. Introducción a la Ciencia estadística en la Agricultura.

2.1. Tablas de frecuencia:

Permiten agrupar las N observaciones de una muestra de acuerdo a su magnitud, en varias clases o grupos. Los pasos a seguir en la construcción de una tabla de frecuencia son:

1. Seleccionar el número de clases (K); para esto puede utilizarse una regla práctica: $[5 \leq K \leq 15]$, o una fórmula matemática, en la cual el número de clases a escoger depende del número de observaciones así:

$$\boxed{K = 1+3.3 \ \text{Log}_{10} \ N}$$ (fórmula de STURGES)

Número de observaciones

Logaritmo en base 10

Número de clases

En general el número de clases debe ser tal; que permita una clara visión de la naturaleza o tipo de distribución. Debe tenerse en cuenta que la selección del número de clases es arbitraria y los criterios mencionados no deben ser tomados como una "Ley". Para ilustrar con un ejemplo, se tomaron los datos del "Vivero internacional de rendimiento de arroz para America Latina - Variedades tempranas del año 1977", en el cual se obtuvieron 725 datos de producciones de arroz expresadas en toneladas/ha. Haciendo uso del criterio práctico se tomaron 12 clases (ver tabla 1). Nótese que la diferencia entre los limites de clase es igual para todas las clases 0.899 Ton/ha, lo cual es deseable cuando se construye una tabla de frecuencia.

2. La Marca de Clase (MC_i) se obtiene al promediar los límites de clase, y su función es: Representar a todas las observaciones que caen dentro de una clase i.

3. La frecuencia absoluta (f_i), es el número de observaciones que quedan dentro de una clase.

4. La Frecuencia Relativa (FR_i) es el % de individuos que quedan ubicados en una clase i, se halla así:

$$\boxed{FR_i = \frac{Fi}{N} \times 100}$$

Por ejemplo la frecuencia relativa para la clase 5 es:

$$FR_5 = \frac{95}{725} \times 100 = 13.10\%$$

TABLA 1: Tabla de frecuencia, de la producción en toneladas por hectarea del "Vivero internacional de rendimiento de arroz para America Latina" - Variedades Tempranas.

Clase No.	Limites de Clase	Marca de clase	Frecuencia Absoluta	Frecuencia Relativa(%)	Frecuencia Acumulada
1	0.001 - 0.900	0.4555	40	5.52	5.52
2	0.901 - 1.800	1.3505	18	2.48	8.00
3	1.801 - 2.700	2.2505	33	4.55	12.55
4	2.701 - 3.600	3.1505	86	11.86	22.41
5	3.601 - 4.500	4.0505	95	13.10	37.51
6	4.501 - 5.400	4.9505	119	16.41	53.92
7	5.401 - 6.300	5.8505	125	17.24	71.16
8	6.301 - 7.200	6.7505	99	13.66	84.82
9	7.201 - 8.100	7.6505	56	7.72	92.54
10	8.101 - 9.000	8.5505	31	4.28	96.82
11	9.001 - 9.900	9.4505	15	2.07	98.84
12	9.901 - 10.800	10.3505	8	1.11	100.00
			725	100.00	

5. La frecuencia acumulada, (FA_i), representa las frecuencias relativas acumuladas, e indica el porcentaje de veces que la variable toma valores que no exceden al limite superior de la clase a la cual corresponde. Por ej.: la FA_7 con un valor de 71.16 significa que el 76.16% de los rendimientos fueron menores que 6.3 ton/ha.

2.2. Representaciones gráficas de tablas de frecuencia:

Para visualizar con claridad el tipo de distribución, son muy útiles los histogramas y los polígonos de frecuencias. Los primeros son representaciones graficas construídas a partir de rectangulos, cuya base es el intervalo de clase, y cuya altura es la frecuencia absoluta o relativa de la clase. Los polígonos de frecuencia se construyen uniendo los puntos medios de las Marcas de clase a la altura de su frecuencia (ver figura 1). En esta figura se observa que la mayoría de las producciones se encuentran agrupadas en las clases centrales, así: el 60.41% de las producciones estuvieron entre 3.6 y 7.2 ton/ha., y solo el 8% fueron menores de 1.8 ton/ha y el 3.18% mayores de 9.0 ton/ha. Esta forma de distribución se conoce como Distribución Normal, y es de común ocurrencia en los fenómenos biológicos.

repite cinco veces.

Media ponderada. Además de las observaciones, tiene en cuenta un factor de ponderación, es decir que no todas las observaciones tienen igual importancia. Así por ej. si se tienen los siguientes registros de producción y áreas sembradas en arroz:

Producción Ton/ha (X_i)	Hectáreas (f_i)
5.0	10
4.0	15
8.0	5

La media ponderada $\overline{X}_p$ es:

$$\boxed{\overline{X}_p = \frac{\sum_{i=1}^{n} X_i f_i}{\sum_{i=1}^{n} f_i}}$$

Siendo f_i, el factor de ponderación, en este caso el número de Has. o sea:

$\overline{X}_p = [5.0(10) + 4.0(15) + 8.0(50)]/10 + 15 + 5$

$\overline{X}_p = 5.0$ ton/ha.

En las tablas de frecuencias, se halla la media, ponderando las marcas de clase por la frecuencia absoluta. Por ej.: para hallar el promedio de las producciones del VIRAL-Tempranas (ver tabla 1) se

- procede así:

$$\boxed{\overline{X}_p = \frac{\sum_{i=1}^{12} MC_i \times FA_i}{\sum_{i=1}^{12} FA_i}}$$

$$\therefore \overline{X}_p = \frac{0.4555+40+\ldots+10.3505\times 8}{40+18+\ldots+8}$$

$$\therefore \overline{X}_p = 5.07 \text{ ton/ha.}$$

2.4. <u>Medidas de dispersión o variabilidad</u>:

Las medidas más conocidas son:

- Rango (R)
- Varianza (S^2)
- Desviación estándar (S)
- Coeficiente de Variación (CV)
- Varianza ponderada

El <u>Rango</u> (R) es la diferencia entre el mayor y el menor valor de un conjunto de datos, así el Rango de las observaciones de la Pág. 10 es: 7.0-1.0=6. Tiene las mismas unidades que las observaciones, en este caso Ton/ha.

La <u>Varianza</u> (S^2) se halla así:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \rightarrow \text{Fórmula general}$$

ó

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1} \rightarrow \text{Fórmula de máquina}$$

Para los datos de la Pág. 10 la varianza es:

$$S^2 = \frac{3.0^2+4.0^2+\ldots+4.0^2-11(4.0)^2}{10}$$

$$\therefore \boxed{S^2 = 2.05\ (\text{ton/ha})^2}$$

La <u>desviación estándar</u> es la raíz cuadrada de la varianza, su fórmula es:

$$S = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}}$$

Para el ejemplo: $S = \sqrt{2.05}$

$$\boxed{S = 1.43\ \text{Ton/ha}}$$

El <u>coeficiente de Variación</u> (CV) es una medida de dispersión que relaciona la desviación estándar con la media así:

$$\boxed{CV = \frac{S}{\bar{X}} \times 100}$$

$$\therefore CV = \frac{1.43\ \text{Ton/ha}}{4.0\ \text{Ton/ha}} \times 100$$

$$\therefore \boxed{CV = 35.79\%}$$

Este parámetro no tiene unidades y permite:

- Comparar la variabilidad de características iguales en poblaciones diferentes. Por ej,: Rendimiento de grano de arroz vs. rendimiento de grano de fríjol.

- Comparar la variabilidad de características diferentes en poblaciones iguales, asi por ej.: el rendimiento de grano en arroz vs. altura de plantas.

En las tablas de frecuencia se puede calcular la varianza así:

$$S = \frac{\sum_{i=1}^{n} f_i d_i^2}{\sum_{i=1}^{n} f_i - 1}$$

Siendo f_i = frecuencia absoluta

d_i = diferencia entre la marca de clase y la media $(MC_i - \overline{X})$

Para la tabla de frecuencia construída, la varianza es:

$$S^2 = \frac{40(0.4555-5.07)^2+\ldots+8(10.3505-5.07)^2}{725 - 1}$$

$$S^2 = \frac{3395.50}{724} = 4.69 \text{ Ton/ha}^2$$

y la desviación $S = \sqrt{4.69} = 2.17$ Ton/ha

La <u>Varianza ponderada</u>, se utiliza cuando se quiere estimar una varianza general a partir de varianzas que provienen de muestras con diferente número de observaciones.

La fórmula es: $$S^2_p = \frac{\sum_{i=1}^{n} S^2_i GL_i}{\sum_{i=1}^{n} GL_i}$$

Siendo S^2_i = Varianza de la muestra i

GL_i = Grados de libertad asociados a la varianza i (n_i-1)

En el siguiente ejemplo, en el cual se tienen las varianzas de la producción de una variedad en (ton/ha)^2, provenientes de experimentos con diferente número de repeticiones, así:

S^2_i	No. de repeticiones	Grados de libertad
0.36	2	1
0.40	4	3
0.30	3	2

La varianza ponderada S^2_p sería:

S^2_p = (0.36(1)+0.40(3)+0.30(2))/(1+3+2)

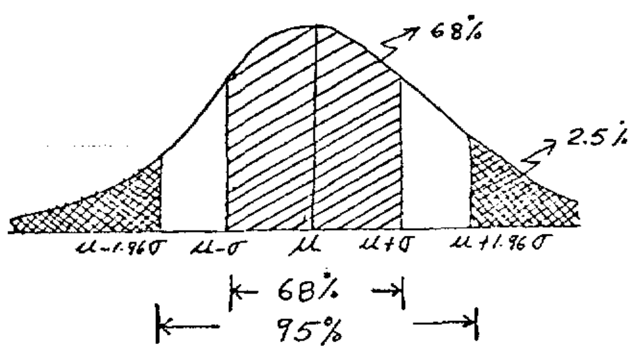
S^2_p = 0.36 (Ton/ha)^2

Nota: Los grados de libertad, son iguales al número de datos, menos el número de parámetros estimados, para el caso de la varianza, solo se ha estimado la media.

2.5. Distribución Normal:

Los fenómenos biológicos en su gran mayoría tienen una distribución normal, la cual tiene como características principales:

- Simetría. Con respecto a un eje que pase por la media.
- La media, la mediana y la moda coinciden
- La media de la población (μ) mas o menos una desviación estándar, contiene el 68.27% de todos los elementos, y la media mas o menos 1.96σ contiene el 95% de los elementos.



- La mayoría de los elementos quedan localizados en la parte central.

2.6. Teorema del Límite Central:

"Las medias de las muestras tomadas de una población normal, se distribuyen normalmente. Podiamos ir más allá y decir, que conforme crece el tamaño de la muestra, las medias muestrales tienden, a distribuirse normalmente aun cuando la población madre sea anormal, lo cual es muy conveniente en aplicaciones prácticas, en que la forma exacta de las poblaciones muestreadas generalmente se desconoce"[1]

"Este teorema del Límite Central es el más importante en estadística visto desde el punto de vista teórico y práctico y en todas las matemáticas es uno de los más importantes"[2]

Puede resumirse este teorema así: "Si se seleccionan muestras al azar de una población con media y varianza σ^2, las medias muestrales se distribuiran con una media muestral $\sigma^2_{\bar{x}}$ igual a $\frac{\sigma^2}{n}$, siendo n el tamaño de la muestra".

3. <u>INFERENCIA ESTADISTICA</u>:

Al realizarse un análisis estadístico, el objetivo que se persigue, es obtener información de la población a través de una muestra, siendo generalmente esta un experimento, lo que se conoce con el nombre de "<u>Inferencia estadística</u>".

3.1. <u>Tipos de errores al generalizar</u>:

Cuando se quiere probar una nueva variedad, con una variedad tradicional la pregunta que surge es:

1/ Snedecor y Cochran. Métodos estadísticos

2/ Mood A. Introduction to theory of statistics.

¿La nueva variedad rinde igual a la variedad tradicional? o expresándolo en diferente forma: $\mu_1=\mu_2$, a esta hipótesis se conoce como Hipótesis nula o H_0. Al probar este tipo de hipótesis de nulidad pueden surgir dos tipos de errores:

- Rechazar la hipótesis, dado de que esta es cierta lo cual se conoce como error TIPO I, es decir, se concluye que existe diferencia real entre las medias, cuando la diferencia se debe al azar.
- Y segundo, aceptar una hipótesis, dado de que esta es falsa, lo cual se conoce como error TIPO II.

Estos tipos de errores se pueden resumir en la siguiente tabla:

HIPOTESIS NULA	CIERTA	FALSA
Acepto	Conclusión correcta	Error Tipo II
Rechazo	Error Tipo I	Conclusión Correcta

A la probabilidad de cometer ERROR TIPO I, α; se le conoce como NIVEL DE SIGNIFICANCIA, y β es la probabilidad de cometer ERROR TIPO II. Se llama POTENCIA DE LA PRUEBA a $(1-\beta)$. Un diseño experimental adecuado y el incremento del número de repeticiones disminuyen la probabilidad de cometer ambos tipos de errores.

3.2. <u>Prueba de "T"</u>:

Esta prueba sirve para comparar dos promedios, y se requiere de una tabla para entrar a aceptar o rechazar una hipótesis nula. Dependiendo de la relación existente entre los pares de datos se usa la prueba de "T" para datos pareados, o para datos no pareados.

Ejercicios:

1. En la Tabla 1, se calcula:

   - Qué % de los Rendimientos fueron menores de 7.2 ton/ha
   - Qué % de los Rendimientos fueron mayores de 9.9 ton/ha.

2. En las siguientes tablas de frecuencias incompletas correspondientes al "Vivero de resistencia al añublo de la vaina" VIAVAL y al "Vivero internacional de rendimiento para América Latina variedades precoces" VIRAL-P.

Clase No.	Límites de Clase Ton/ha	VIAVAL	VIRAL-P
1	0.001 - 0.900	17	6
2	0.901 - 1.800	50	9
3	1.801 - 2.700	51	32
4	2.701 - 3.600	60	83
5	3.601 - 4.500	79	78
6	4.501 - 5.400	54	101
7	5.401 - 6.300	20	95
8	6.301 - 7.200	35	79
9	7.201 - 8.100	21	45
10	8.101 - 9.000	10	21
11	9.001 - 9.900		10
12	9.901 -10.800		5

Calcule: a) Promedios,

b) Desviaciones estándar

c) Discuta porque pueden presentarse diferentes promedios y desviaciones estándar. Desde el punto de vista biológico.

d) Construya la tabla de frecuencia y Realice los respectivos histogramas y polígonos de frecuencia. Discuta el tipo de distribución.

e) Qué porcentaje de las observaciones fueron mayores a 4.501 ton/ha.

- Para el VIAVAL
- Para el VIRAL-P

f) Qué ventajas puede tener para el investigador o para el profesional, conocer las tablas de frecuencia?

3. Si una población estadística de alturas de plantas de arroz, se distribuye normalmente con media igual a 80cm. y desviación estándar igual a 6cm., entre que valores estarían:

a) El 68.27 o (68%) de las alturas?

b) El 95% de las alturas.

4. Si en el vivero "VIRAL TEMPRANAS-1977" en el experimento realizado en Nataima el promedio de altura de planta fué 107 cm. y la desviación estándar 5 cm. calcule:

a) El coeficiente de variación

b) La varianza

c) A partir de que altura está el 2.5% de las plantas más bajas?

d) A partir de que altura está el 2.5% de las plantas mas altas?

5. En el mismo experimento anterior el promedio de rendimiento fué 7.54 ton/ha, y la desviación estándar 1.26 ton/ha. Indique, haciendo uso del coeficiente de variación que característica es más variable. (Producción o altura).

4. CORRELACION Y REGRESION:*

Introducción

En esta parte se pretende dar una idea general sobre la correlación y la regresión, sin entrar a detalles matemáticos; este tema sería tratado con mayor profundidad al final de este curso.

4.1. Correlación:

Es el grado de asociación que existe entre dos variables.

Existe correlación positiva cuando al incrementarse la magnitud de una variable, la otra también se incrementa, así por ejemplo al incrementarse hasta cierto límite la cantidad de Nitrógeno, la producción de arroz aumenta, o al aumentarse el número de desyerbas la producción aumenta.

Existe correlación negativa, cuando al aumentarse la magnitud de una variable, la magnitud de la otra disminuye, así por ejemplo, al aumentar el peso de malezas por unidad de área, disminuye el rendimiento de grano.

Cuando se asocian dos variables, generalmente a una se le denomina variable independiente y a la otra variable dependiente; los valores que toma esta última dependen de los valores que ha tomado la variable dependiente, así por ejemplo, la producción de

---

* Capítulo tomado de MUÑOZ, J.E. "Curso de Arroz para profesionales de America Latina". Primer curso de 1979.

arroz depende de la cantidad de Nitrógeno que se le agregue al suelo y no al contrario. Asi también la producción depende de la cantidad de malezas presentes.

4.1.1. Como cuantificar el grado de asociación:

Al realizar un diagrama de dispersión, en el cual se ubica en el eje X la variable independiente y en el eje Y la variable dependiente, y localizar los respectivos pares de datos, se puede apreciar si existe o no correlación y si esta es positiva o negativa; pero no se puede determinar ningun valor.

Para cuantificar el grado de asociación se usa el coeficiente de correlación = r, cuyos valores están entre menos uno y más uno así:

$[-1 \leq r \leq 1]$

En la medida en que el coeficiente de correlación se acerque a uno, en valor absoluto, la asociación tiende a ser perfecta, cuando el γ es positivo hay correlación positiva y cuando r es negativo hay correlación negativa.

La fórmula para calcular r es:

$$r = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2] x [\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2]}}$$

La hipótesis nula que se establece es:

H_0: $r=0$ o sea que no hay asociación entre las variables.

Regla de Decisión. Acepto H_0: si $r_c < |rt|$

Siendo r_c = Coeficiente de correlación calculado.

r_t = Coeficiente de correlación de las tablas.

Las probabilidades de Error Tipo I se fijan de antemano, para este ejemplo usaremos 5% o 0.05, los grados de libertad con los cuales se busca en la tabla son: n-2; siendo n el número de pares de datos. Al buscar en la tabla, el sitio de intersección es:

GL	$\alpha \rightarrow 0.05$
.	
.	
.	
5	$\boxed{0.7545}$

O sea que se acepte H_0: Si $r_c < 0.7545$

De acuerdo a la fórmula de r y basado en la tabla se tiene que:

$$r_c = \frac{\Sigma X_i Y_i - \overline{nXY}}{[\Sigma X_i^2 - n\overline{X}^2][\Sigma Y_i^2 - n\overline{Y}^2]}$$

$$\therefore \quad r_c = \frac{3462.5 - 7(75)(5.84)}{[56875 - 7(75)^2][248.11 - 7(5.84)^2]}$$

$$\therefore \quad r_c = 0.98$$

De acuerdo a la regla de Decisión, rechazo H_0, y concluyo que existe relación o asociación entre el Nitrógeno aplicado y la producción de arroz.

4.2. Regresión:

"Regresión es la cantidad de cambio de una variable asociada a un cambio unico en otra variable".

En esta parte solo hablaremos de las "Rectas de Regresión", es decir sólo se tomaran ejemplos en los cuales, el fenómeno bajo consideración es explicado por una linea recta; cuando se trate de "relaciones no explicables por lineas rectas, el lector puede acudir a textos de estadística*.

Si se sabe que la relación entre dos variables es lineal, la pregunta que surge es: "Cuál es la línea recta que más se ajusta a los datos".

Para comprender mejor este tema, construyamos un diagrama de dispersión de la tabla 2.1., tomando como variable independiente (X) la cantidad de Nitrógeno aplicada y como variable dependiente el rendimiento de arroz (Y) en toneladas por hectárea, ver figura 2.1.

---

* Steel y Torrie. Principles and procedures of statistics. 1960. Snedecor. Métodos estadísticos. 1970.

TABLA 2.1.: Rendimiento de grano (Ton/ha) $\underline{Y}$ y la cantidad de Nitrógeno aplicada en Kg/ha. (Resultados obtenidos en la variedad de arroz CICA-4)

	X_i	Y_i	X_i^2	Y_i^2	X_iY_i
	0	4.0	00000	16.00	0.0
	25	4.6	625 625	21.16	115.0
	50	5.4	2500	29.16	270.0
	75	6.0	5625	36.00	450.0
	100	6.7	10000	44.89	670.0
	125	6.9	15625	47.61	862.5
	150	7.3	22500	53.29	1095.0
$\sum_{i=1}^{7}$	525	40.9	56875	248.11	3462.5

$\overline{X} = 75$ $\overline{Y} = 5.84$

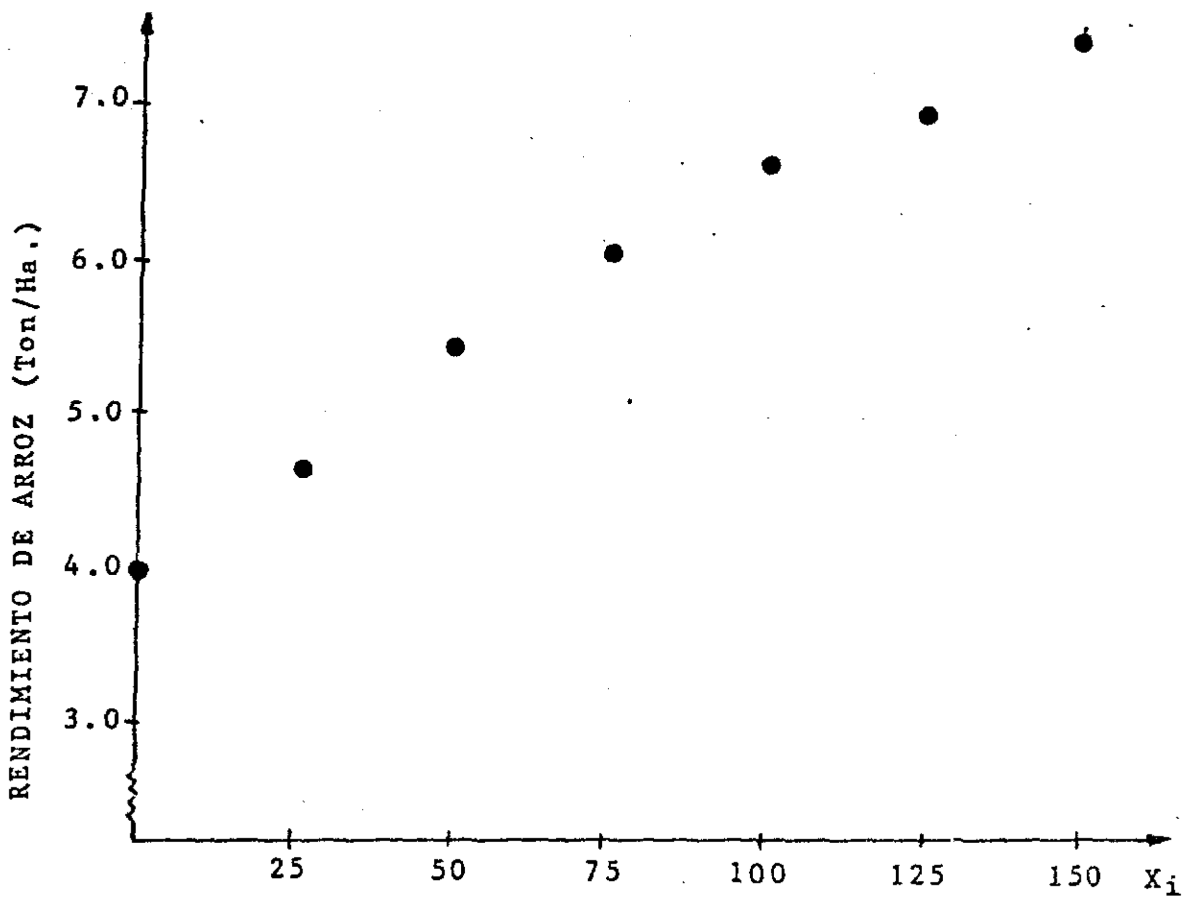


FIG. 2.1. Diagrama de dispersión de rendimiento de arroz en Ton/ha y Nitrógeno aplicado en Kg/ha.

En la figura 2.1., para hallar la recta de mejor ajuste, se parte de la base de que a pesar de no aplicar Nitrógeno al suelo, se va a obtener producción de arroz, esto es lógico, ya que hay producción a pesar de que este elemento mayor, no se agregue al suelo. Cuando se espera que a un valor de cero en la variable independiente X, corresponda un valor diferente de cero, en la variable dependiente Y, se usa un MODELO LINEAL, de la forma:

$$\hat{Y}_i = \alpha + \beta X_i$$

β: Coeficiente de regresión o pendiente

α: Intercepto

$\hat{Y}_i$: Valor estimado de Y

Para estimar los parámetros α y β se usan las fórmulas obtenidas por el "Método de los Mínimos Cuadrados".

Siendo:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} \qquad y \qquad \hat{\alpha} = \overline{Y} - \beta\overline{X}$$

La interpretación de $\hat{\beta}$ es: En cuantas unidades cambia la variable dependiente al cambiar en una unidad la variable independiente; y la de α o intercepto es:

A un valor de cero de la variable independiente; cuál es el valor que toma la variable dependiente. Asi para el ejemplo de la tabla 2.1.

$$\hat{\beta} = \frac{3462.50 - 7(75)(5.84)}{56875 - 7(75)^2}$$

$$\therefore \hat{\beta} = 0.023 \ \frac{\text{Ton/ha de grano}}{\text{Kg de N.}}$$

y $\hat{\alpha} = 5.84 - 0.023(75)$

$\alpha = 4.12$ Ton/ha de grano

Lo que indican estos parámetros para el ejemplo es:

$\hat{\beta} \rightarrow 0.023$: Al aumentar en un Kg. la aplicación de Nitrógeno se aumenta la producción en 0.023 Ton/ha o 23 Kg/ha.

$\alpha \rightarrow 4.12$ Ton/ha indica que cuando no se aplica fertilizante al suelo se obtiene una producción de 4.12 Ton/ha.

El modelo de la recta de mejor ajuste quedaría.

$$\boxed{\hat{Y} = 4.12 + 0.023X_i}$$

Para ubicar "La Recta" en el Diagrama de dispersión se le dan valores a la variable X_i así:

Para $X = 0 \rightarrow \hat{Y} = 4.12$

$X = 100 \rightarrow \hat{Y} = 4.12 + 0.023\ (1.00) = 6.42$

Con lo cual podemos localizar los respectivos pares ordenados (X,Y) en la figura 2.2. y trazar la linea recta de mejor ajuste.

En base a la ecuación, podríamos predecir, por ejemplo, cuál sería la producción de arroz, cuando se aplican 80 Kg de N. por hectarea, así:

$$\hat{Y} = 4.12 + 0.023(80)$$

$$\therefore \hat{Y} = 5.96 \text{ Ton/ha.}$$

En este ejemplo se asumió que la relación entre el nivel de Nitrógeno aplicado y la producción es explicada por una linea recta, pero hay que tener presente, que para este ejemplo, esto es cierto con niveles de Nitrógeno entre 0 y 150 Kg/ha, lo cual no significa que al aumentar indefinidamente los niveles de Nitrógeno, implique un aumento indefinido en la producción. Normalmente esta relación se explica mejor con modelos no lineales, en los cuales los rendimientos son finalmente decrecientes.

Cuando se espera que a un valor de cero en la variable independiente, corresponda un valor de cero en la variable dependiente, debe seleccionarse el modelo de la forma:

$$\boxed{\hat{Y}_i = \beta X_i}$$

En donde β es el coeficiente de regresión o pendiente y se estima en base al método de los mínimos

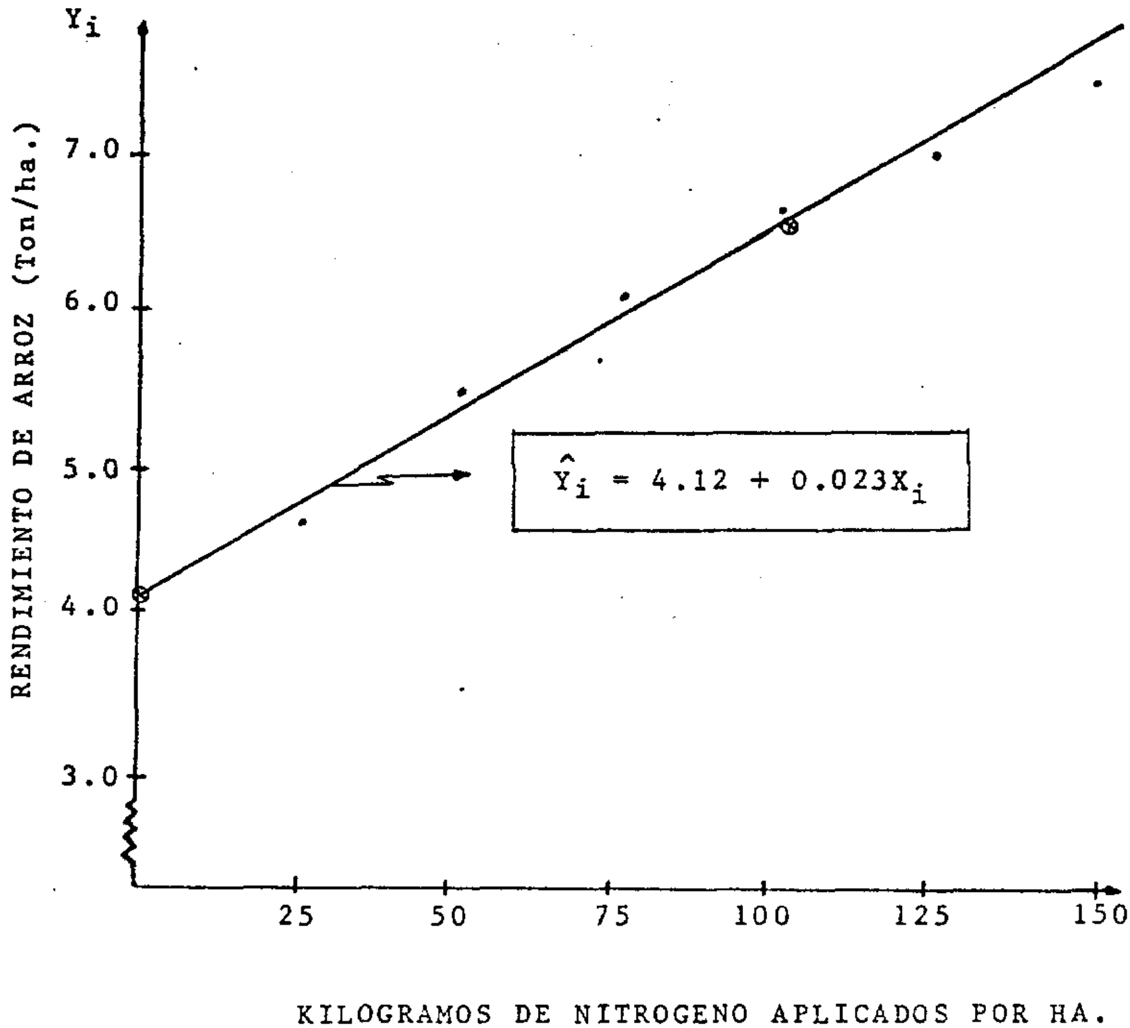
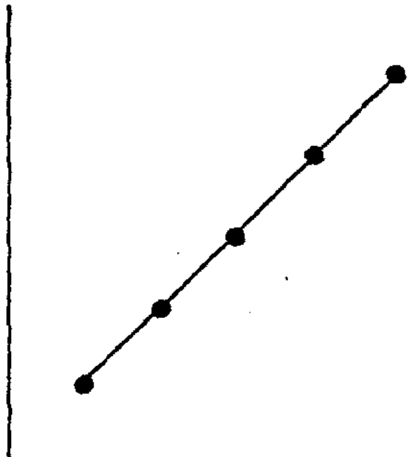


FIG. 2.2.: Diagrama de dispersión y ecuación de mejor ajuste de rendimiento de arroz en Ton/ha., en función de los Kg. de Nitrógeno por Ha., aplicados al suelo.

cuadrados en la siguiente forma:

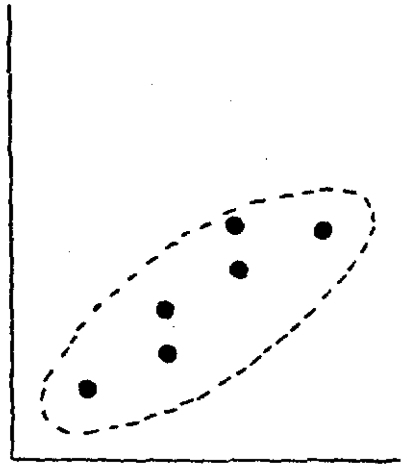
$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

Un ejemplo donde se podría usar este modelo, es: "Se quiere hallar la relación entre el Producto del largo de la hoja por su ancho, con el área foliar". Porque: Cuando "No hay largo o ancho", la hoja no existe, o sea cuando $X = 0$ $Y = 0$.



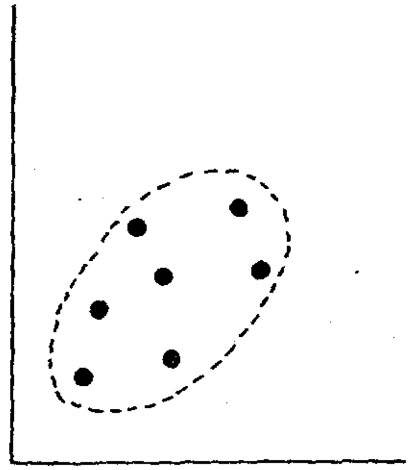
POSITIVO PERFECTO

Y=1.0



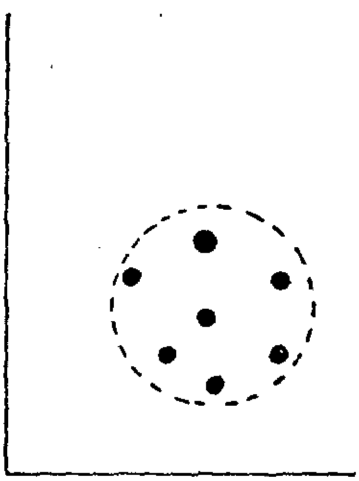
POSITIVO ALTO

Y = 0.8



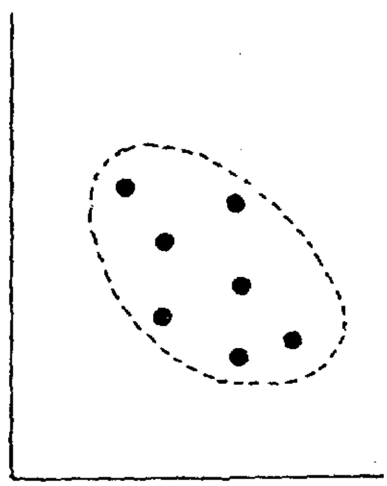
POSITIVO BAJO

Y = 0.1



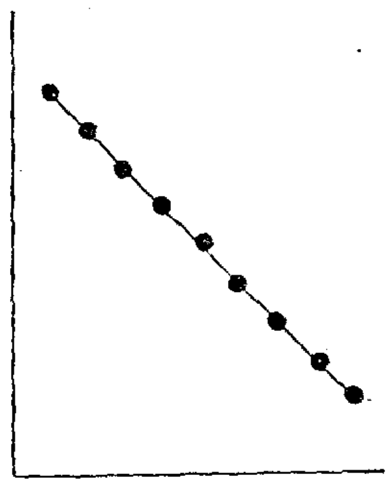
CERO

Y = 0



NEGATIVO MEDIO

Y = 0.5



NEGATIVO PERFECTO

R = -1

Resumen de fórmulas:

Coeficiente de Correlación:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{(\Sigma X_i^2 - n\overline{X}^2)(\Sigma Y_i - n\overline{Y}^2)}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

Cuando
$\downarrow$
$Y_i = \beta X_i + \varepsilon_i$

$$\hat{\beta} \quad \frac{\Sigma X_i Y_i - n\overline{X}Y}{\Sigma X_i^2 - n\overline{X}^2}$$

Cuando
$\downarrow$
$Y_i = \alpha + \beta X_i + \varepsilon_i$

$$\alpha = Y - \beta\overline{X}$$

Ejercicios:

1. Tabla 2: "Efecto de la competencia de malezas en arroz". Rendimientos de arroz en Kg/parcela y peso de malezas en Gr/m^2

Y_i Rendimiento	X_i P. Seco malezas	X_iY_i
3.75	4.3	16.13
3.67	3.6	13.21
3.15	41.6	131.04
2.53	200.0	506.00
2.58	336.0	866.88
0.10	1116.0	111.60
2.73	325.0	887.25
0.04	946.0	37.84
3.19	3.3	10.53
2.30	604.0	1389.20

a) Construya un diagrama de dispersión, e indique cual fué el criterio para seleccionar la variable independiente y la dependiente.

b) Cuál de los dos modelos que se dan a continuación seleccionaria y porqué?

$\hat{Y} = \beta X_i$

$\hat{Y} = \alpha + \beta X_i$

c) Calcule el ó los parámetros del modelo

d) Cuando se tiene un peso seco de malezas de 500 gr., que rendimiento por parcela espera?

e) Calcule el coeficiente de correlación, e indique si es o no significativo (α = 0.05 ó 5%)

2. En estudios de nutrición en cerdos se ha encontrado que existe relación entre el peso vivo de los lechones y la cantidad de alimento requerido para aumentar 1Kg de peso.

Se tuvieron machos castrados de la misma raza durante 147 días, en corrales individuales, con la misma alimentación desde el destete hasta aproximadamente 100 Kg de peso. Se controlo peso cada 21 días, los resultados promedio de cada pasaje fueron los siguientes:

Tiempo de pesaje	Peso vivo Kg.	Kg. de alimento por 1 Kg. de aumento
Al iniciar el experimento	17.3	-
A los 21 días	26.4	2.45
A los 42 días	36.4	2.67
A los 63 días	48.1	2.89
A los 84 días	63.2	3.15
A los 105 días	76.2	3.44
A los 126 días	88.1	3.77
A los 147 días	100.3	4.35

1. Seleccione la variable independiente y la dependiente, indique cual fué el criterio de selección.
2. Realice un diagrama de dispersión, y discuta desde el punto de vista biológico la relación entre las

dos variables.

3. Calcule el coeficiente de correlación, lance su hipótesis y establezca la regla de decisión y en base a ella concluya si existe o no asociación entre las dos variables.

4. En base al método de los mínimos cuadrados estime, los parámetros α y β. Usando un modelo de regresión con intercepto, recuerde que estudia el fenómeno para valores de peso vivo entre 26.4 y 100.3 Kg.

5. Desde el punto de vista biológico, de el significado de $\hat{\beta}$, e indique que unidades tiene.

Ejercicios:

En estudios agronómicos muchos estudios fisiológicos se basan en el área foliar, la cual permite estudiar su fotosíntesis, tolerancia a duplicación, intensidad de transpiración, etc.

Por estudios anteriores se sabe que existe una relación en maiz entre el producto del largo máximo por el ancho máximo y el área foliar. Para los hibridos de maiz ICA H-207 e ICA H-210 se tomaron al azar 10 hojas obteniendose los siguientes resultados:

ICA H-210		ICA H-207	
LxA(cm^2)	A P(cm^2)	LxA(cm^2)	A P(cm^2)
678.9	520.5	937.2	669.9
1020.8	800.2	927.0	649.6
1253.5	1030.1	990.8	768.4
970.2	655.8	1166.7	912.6
1064.6	866.4	747.3	556.7
1060.2	831.3	833.1	488.7
1312.6	1097.1	999.7	744.1
1111.5	761.4	1180.8	894.4
704.4	785.1	902.0	671.5
1430.0	1051.0	1326.0	1129.0

LA = Producto del largo máximo por el ancho máximo en cm^2

AP = Area planimetrica, la cual se considera la mejor aproximación del área foliar verdadera.

Si se quiere en base a los datos anteriores encontrar un factor que permita hallar el área foliar en base al producto del largo por el ancho máximo, realice:

1. Un diagrama de dispersión. Seleccione con cuidado la variable independiente y la variable dependiente.

2. Halle el coeficiente de correlación y pruebe si estadísticamente es o no significativo con un $\alpha = 5\%$.

3. En base al modelo $\hat{Y}_i = \beta X_i$ calcule β y de el significado y las unidades de este parámetro.

4. Discuta porqué se utiliza el modelo
   $\hat{Y} = \beta X_i$

5. Que ventaja práctica tiene el hallar el β?

5. DISEÑO EXPERIMENTAL:

Antes de entrar a ver los diferentes tipos de diseños experimentales se darán conceptos sobre algunos terminos que son de uso muy frecuente en la experimentación agropecuaria.

Un experimento puede considerarse como ún medio que emplea el hombre para comunicarse con la naturaleza, promoviendo cambios bajo situaciones controladas.

En general el hombre aplica los tratamientos a las unidades experimentales y mide la respuesta de ellas con las "variables de respuesta". Este puede ser representado de la siguiente forma:

ESTIMULOS → [UNIDADES EXPERIMENTALES] → VARIABLES DE RESPUESTA

Los estímulos aplicados a las unidades experimentales son los tratamientos.

Las unidades experimentales son la base física sobre la cual se aplican los tratamientos o estímulos. Una unidad experimental puede ser una parcela en el campo, un animal, o un grupo de animales.

Los tratamientos provienen de factores cualitativos o cuantitativos; cualitativos puede ser por ejemplo: formas de aplicar un fertilizante, al voleo, en banda, variedades o especies de pastos sembrados y cuantitativos son aquellos que se pueden medir o cuantificar por ejemplo 10, 20, 50 o 100 Kg. de Nitrógeno por hectarea.

Un tratamiento también puede ser la combinación de factores cualitativos y cuantitativos asi por ejemplo un tratamiento puede ser: Aplicar 100 Kg. de Nitrógeno al voleo, o 50 Kg. de N en banda.

El diseño experimental es "La forma como se aplican los tratamientos a las unidades experimentales"

Principios fundamentales del Diseño Experimental:

- Aleatorización o asignación al azar de los tratamientos.
- Repeticiones
- Control del error experimental.

La aleatorización evita la introducción de errores sistemáticos en los experimentos, es decir evita que unos tratamientos se vean más favorecidos que otros. Así,si una variedad de arroz se siembra en un terreno fertil y la otra en un terreno con problemas de fertilidad, no puede concluirse que una variedad produce más que otra, porque no se puede saber, si la producción se debe a la variedad en si misma o a la fertilidad del suelo.

La repetición es el número de veces que un tratamiento se repite en un experimento, es deseable que el número de repeticiones sea igual para todos los tratamientos para compararlos con el mismo nivel de precisión.

El error experimental se puede minimizar controlando las fuentes de variación externas al experimento, tales como heterogeneidad del suelo, aplicación uniforme de riegos, pesticidas, etc.

Los diseños experimentales de uso más frecuente en la experimentación agropecuaría son:

- Completamente al azar
- Bloques al azar
- Parcelas divididas y subdivididas

5.1. Pasos a seguir en un experimento bien planeado:

En un experimento hay que hacer dos tipos de consideraciones. Las de tipo biológico previas a la realización del experimento y las de ejecución y diseño de este.

En las primeras, un factor importante es la selección de los tratamientos, en la cual debe primar el criterio del investigador, que selecciona sus tratamientos en base a la experiencia adquirida sobre el problema que desea estudiar, y a los aportes de otros profesionales.

En la ejecución del experimento, hay una cadena de pasos que deben seguirse, los cuales se presentan en el siguiente esquema 1/

1/ Tomado de Baena G. Elementos básicos para la evaluación del método experimental y análisis estadístico en la experimentación con fertilizantes. U.N. 1976.

Conocimiento de las Unidades exp rimentales → Forma de aleatorizar → Diseño experimental →

→ Modelo estadístico → Tipo de Análisis

El conocimiento de las unidades experimentales, lleva una forma de asignar los tratamientos (aleatorización) esto determina el diseño experimental a utilizar, cada diseño tiene su modelo estadístico, y de acuerdo a este se realiza el análisis.

Es muy importante seleccionar con anticipación las variables de respuesta que van a ser cuantificadas. Asi en ensayos de rendimiento, se mide la producción; en ensayos para control de malezas, además de la producción de arroz, debe tomarse una medida sobre el grado de control de malezas por ejemplo: Peso seco de malezas.

6. DISEÑO COMPLETAMENTE AL AZAR:

6.1. Características Generales:

- Es el diseño más simple, y se usa cuando las unidades experimentales son homogéneas.
- Se puede comparar cualquier número de tratamientos.
- Los tratamientos se aplican a las unidades experimentales al azar.
- Cualquier número de repeticiones por tratamiento es posible, es decir los tratamientos pueden tener diferente número de repeticiones.
- Proporciona más grado de libertad al error que cualquier otro diseño, con el mismo número de tratamientos y repeticiones.

Dada la restricción de que las unidades experimentales deben ser homogéneas, se usa muy poco en el campo, pero es de mucha utilidad en experimentos de laboratorio o de invernadero, donde se pueden uniformizar las unidades experimentales, por ejemplo en invernadero el suelo se puede homogenizar, y en el laboratorio se puede preparar un medio de cultivo homogéneo.

6.2. Modelo Estadístico:

Este diseño experimental está asociado al siguiente modelo estadístico:

$$\boxed{Y_{ij} = \mu + \tau_i + \varepsilon_{ij}}$$

Error experimental en la repetición j'esima del tratamiento i'esimo

Efecto del tratamiento i'esimo (i=1,2,...,t)

Efecto de la media general

Variable de respuesta del tratamiento i'esimo, en la repetición j'esima.

Con los siguientes supuestos:

$$\sum_{i=1}^{t} \tau_i = 0 \quad y \quad \varepsilon_{ij} \text{ NN } (0,\sigma)$$

Siendo:

τ_i = Efecto del tratamiento i ($\tau_i = \mu_i - \mu$)

μ_i = Promedio del tratamiento i

μ = Promedio general

6.3. <u>Aleatorización</u>:

Se asignan los tratamientos a las unidades experimentales al azar. Así por ejemplo, si se va a comparar el rendimiento de las siguientes variedades de arroz:

1-JUMA 58

2-TIKAL 2

3-MACUSPANA A-75

4-BG 90-2

Las cuales se asignan en el campo en una forma completamente al azar, como se observa en el siguiente esquema:

V_1	V_3	V_1	V_3
V_4	V_2	V_4	V_2

Los rendimientos obtenidos en toneladas por hectárea fueron los siguientes:*

No.	Variedad			Total($Y_i.$)	Promedio ($\bar{Y}_i.$)
1	JUMA-58	3.40	3.30	6.70	3.35
2	TIKAL-2	4.70	4.70	9.40	4.70
3	MACUSPANA A-75	2.00	2.10	4.10	2.05
4	BG 90-2	6.30	5.20	11.50	5.55

Gran Total $Y..$ = 31.70

Promedio General $\bar{Y}..$ = 3.96

Las hipótesis a probar son:

Hipotesis Nula: H_0: τ_i = o ó sea que no hay efecto de tratamiento lo cual índica que las 4 variedades rinden igual.

---

* Datos adaptados del Vivero Internacional de Rendimiento de arroz para América Latina (experimento realizado en Guatemala)

<u>Hipótesis Alternante</u>: H_a: Algunas medias de tratamiento difieren.

Para entrar a probar la hipótesis nula, se recurre al análisis de varianza, el cual se realiza en la siguiente forma:

Paso 1: Se calcula el factor de corrección (FC) o suma de cuadrados debidas a la media.

Siendo r = número de repeticiones

t = número de tratamientos

$$FC = \frac{(Y..)^2}{r \times t}$$

$$FC = \frac{(31.7)^2}{2 \times 4} = 125.61$$

Paso 2: Se halla la suma de Cuadrados de tratamientos (variedades SCTR)

$$SCTR = \frac{\sum_{i=1}^{4} Y_{i.}^2}{r} - FC$$

$$SCTR = (6.7^2+9.4^2+4.1^2+11.5^2)/2 - 125.61$$

$$\boxed{SCTR = 15.55}$$

Paso 3: Se halla la suma de Cuadrados Total (SCT)

$$SCT = \sum_{i=1}^{4} \sum_{j=1}^{2} Y_{ij}^2 - FC$$

$$SCT = (3.4^2+ 4.7^2+ ....+ 2.1^2+ 5.2^2) - 125.61$$

$\boxed{SCT = 16.16}$

Paso 4: Se obtiene por diferencia la suma de cuadrados del error (SCE)

SCE = SCT - SCTR

SCE = 16.16 - 15.65

$\boxed{SCE = 0.61}$

Paso 5: En base a los grados de libertad y a las sumas de cuadrados, se obtiene la Tabla de Análisis de Varianza, así:

Fuente de Variación	g.l.	SC	CM	F_c	F_t 5%
Variedades	3	15.55	5.18	34.55	6.59
Error	4	0.61	0.15		
$Total_{(c)}$	7	16.16			

En la cual los Cuadrados Medios son los cocientes resultantes al dividir las sumas de cuadrados entre los respectivos grados de libertad así:

CMTR = 15.55/3 = 5.18 y CME = $\frac{0.61}{4}$0.15

y la F_c o "F caculada" es la relación entre los cuadrados medios de variedades y error.

$\therefore F_c = 5.18/0.15 = 34.55$

Se habia lanzado la hipótesis nula de no hay efecto de tratamiento o que $\mu_1 = \mu_2 = \mu_3 = \mu_4$

La regla de decisión para probar la hipótesis es:

Acepto H : Si $F_c < F_t$

F_t se busca en la tabla de F, con un nivel de significancia dado, en nuestro caso 5%, con los grados de libertad de tratamientos y del error así:

GL Error	GL Tratamientos 3
4	$\boxed{6.59}$

ALEATORIZACION

EN EL DISEÑO

COMPLETAMENTE AL AZAR

V_1	V_3	V_1	V_3
V_4	V_2	V_4	V_2

UNIDADES EXPERIMENTALES HOMOGENEAS

7. DISEÑO EN BLOQUES COMPLETOS AL AZAR:

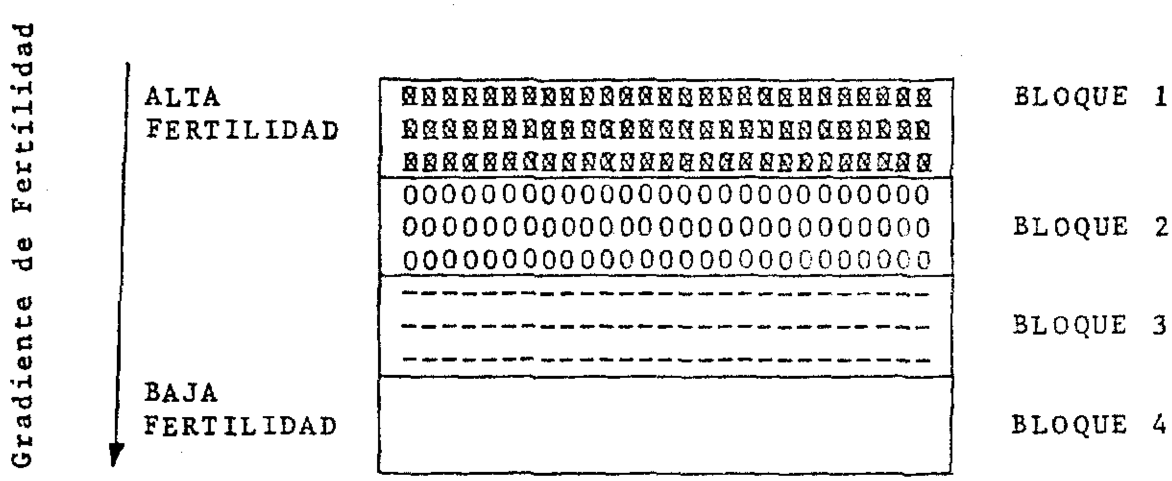
7.1. Características:

- Se utiliza cuando las unidades experimentales no son homogéneas y es posible estratificarlas en grupos más homogéneos.
- Cada bloque contiene a todos los tratamientos
- Se puede usar cualquier número de bloques.
- Desde el punto de vista teórico, se puede usar cualquier número de tratamientos por bloque, pero en la práctica es difícil conseguir en el campo bloques homogéneos para más de 10 o 12 tratamientos.
- Para que el diseño sea más eficiente, se necesita que haya homogeneidad dentro de bloques y heterogeneidad entre bloques.
- Es el diseño experimental más utilizado en la experimentación agrícola. Cerca del 70 u 80% de los experimentos son realizados bajo este diseño.

7.2. Concepto de bloque, ablocamiento, ejemplos:

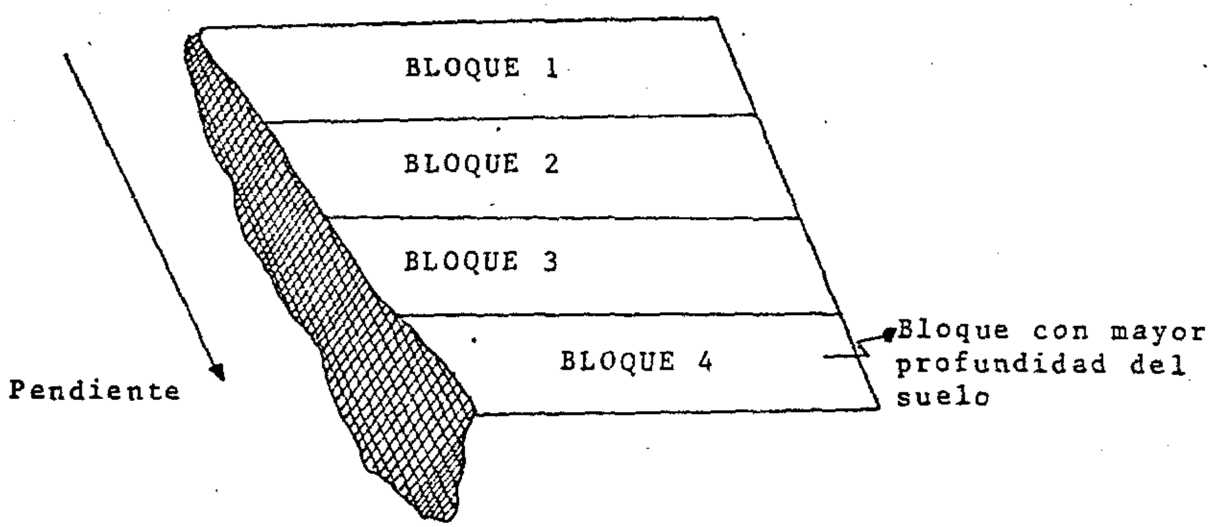
Se entiende por bloque a un conjunto de unidades experimentales homogéneas y por "ablocamiento" al conjunto de prácticas tendientes a ubicar mediante algun método ese conjunto de unidades experimentales homogéneas, con el objeto de reducir el error experimental. En general en experimentos agronómicos se asocia bloque, con una faja de terreno rectangular, en la cual se ubican los tratamientos, y es muy común encontrar que en la mayoría de los experimentos realizados bajo este diseño los bloques tienen esa forma, sin embargo no existe ninguna norma que determine como "mejor" dicha forma de ablocamiento.

Cuando se puede detectar por experiencias anteriores que existe en el terreno un gradiente de fertilidad definido, si es conveniente escoger bloques rectangulares así:

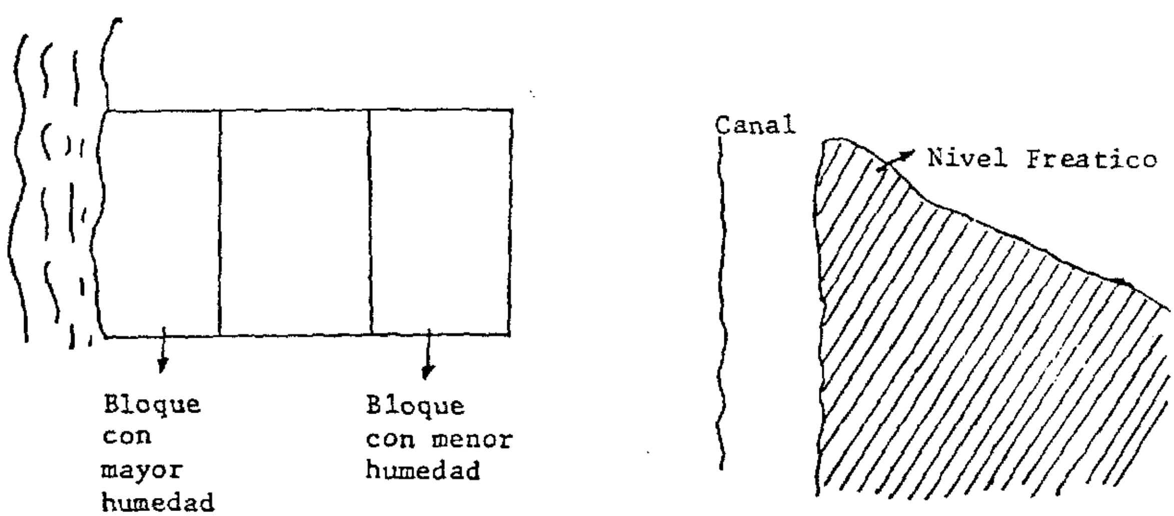


Porque de esta manera habrá más homogeneidad dentro del bloque, y más diferencia entre bloques. En este caso el largo de las parcelas debe ser paralelo al gradiente de fertilidad.

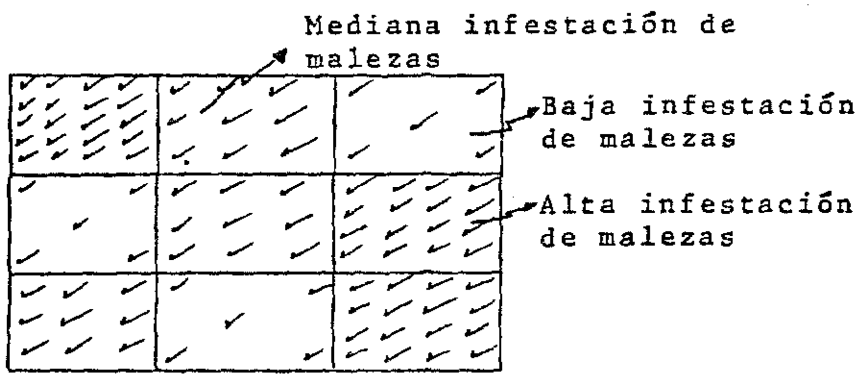
Otro caso en el que se puedan usar bloques rectangulares, ocurre cuando existe alguna pendiente en el terreno, ya que en la parte "baja" del lote habrá mayor deposición de materiales provenientes de la parte alta y por esta razón el suelo puede ser más "profundo" así:



Cuando hay canales de riego, sin revestimiento, las partes del lote más cercanas a este, tendrán más humedad así:



Cuando se quiere por ejemplo, probar el efecto de algunos tratamientos sobre el control de las malezas, se pueden detectar por reconocimiento del terreno zonas de alta, mediante y baja infestación de malezas, al "bloque" con forma geometrica definida deja de ser útil y se convierte en un conjunto de parcelas que tienen como característica común el grado de infestación de malezas así:



Si en el terreno no se detecta ningún gradiente de fertilidad, es conveniente usar bloques más compactos, en lo posible en forma "cuadrada", dado de que por experimentaciones realizadas en diferentes sitios, se ha encontrado que entre parcelas cercanas hay más afinidad en cuanto a fertilidad del terreno. Ejemplo:

BLOQUE 1	BLOQUE 2
BLOQUE 3	BLOQUE 4

7.3. Modelo Estadístico:

El modelo estadístico de este diseño experimental es:

$$\boxed{Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}}$$

Error experimental del tratamiento i'esimo en el bloque j'esimo

Efecto del bloque j'esimo

Efecto del tratamiento i'esimo

Efecto de la media general

Variable de respuesta del tratamiento i'esimo, en el bloque j'esimo.

7.4. Consideraciones Generales:

i) Sea:

$$\bar{Y}_{..} = (\sum_{i=1}^{t} \sum_{j=1}^{b} Y_{ij})/(bt) \qquad \text{Media General}$$

$$\bar{Y}_{i.} = (\sum_{j=1}^{b} Y_{ij})/b \qquad \text{Media del tratamiento i}$$

$$\bar{Y}_{.j} = (\sum_{i=1}^{t} Y_{ij})/t \qquad \text{Media del bloque J}$$

Entonces:

$\bar{Y}_{..}$ es un estimador de μ

$\bar{Y}_{i.}$ es un estimador de $\mu + \tau_j$

$\bar{Y}_{.j}$ es un estimador de $\mu + \beta_j$

Nota: b = número de bloques

t = número de tratamientos

ii) Las sumas de cuadrados de las desviaciones con respecto a $\overline{Y}..$ (SCTotal) se descompone así:

$$\Sigma\Sigma(Y_{ij}-\overline{Y}..)^2=\Sigma\Sigma(\overline{Y}_i.-\overline{Y}..)^2+\Sigma\Sigma(Y._j-\overline{Y})^2+\Sigma\Sigma(Y_{ij}-\overline{Y}_i.\overline{Y}._j+\overline{Y}..)^2$$

SCTotal = SCTratamiento + SCBloques + SCError

Las sumas de cuadrados anteriores estan asociadas con los siguientes grados de libertad; respectivamente:

Grados de libertad total:	b^t-1 $(b*t)-1$
Grados de libertad tratamientos:	$t-1$
Grados de libertad bloques:	$b-1$
Grados de libertad Error:	$(b-1)(t-1)$

Como en el diseño completamente al azar, el cociente entre una suma de cuadrados (S.C.) y sus correspondientes grados de libertad (GL.) es denominado "cuadrado medio" del efecto para el que se calcula. Así mismo el cuadrado medio del error (CME) es un estimador de σ^2 (varianza del experimento).

Cabe anotar que el cuadrado medio del error calculado en un diseño en BCA (CME_{BCA}) sera menor o igual al cuadrado medio del error calculado para un diseño completamente al azar (CME_{CA}) es decir:

$$CME_{BCA} \leq CME_{CA}$$

Esto es justamente el objetivo del bloqueo, aumentando la precisión de las comparaciones extrayendo de la variabilidad del diseño completamente al azar, aquella variabilidad debida a la no homogeneidad expresada mediante la diferencia de bloques.

iii) En el Diseño de BCA es posible probar <u>independientemente</u> estas 2 hipótesis nulas:

H_0: $\tau_i = 0$ para todo i (no hay diferencia entre tratamientos)

H_0: $\beta_j = 0$ para todo j (no hay diferencia entre bloques).

Las reglas de decisión para las 2 anteriores hipótesis son similares a las reglas de decición para el diseño completamente al azar.

La construcción de la tabla de análisis de varianza, se puede resumir así:

Fuentes de Variación	GL	SC	CM	F_c
Bloques	b-1	$\dfrac{\sum_{j=1}^{b} Y_{.j}^2}{t} - \dfrac{Y_{..}^2}{b x t}$	$\dfrac{SCB}{b-1}$	$\dfrac{CMB}{CME}$
Tratamientos	t-1	$\dfrac{\sum_{i=1}^{t} Y_{i.}^2}{b} - \dfrac{Y_{..}^2}{b x t}$	$\dfrac{SCTR}{t-1}$	$\dfrac{CMTR}{CME}$
ERROR	(b-1)(t-1)	SCTOTAL-SCB-SCTR	$\dfrac{SCE}{(b-1)(t-1)}$	
$TOTAL_{(c)}$	bt - 1	$\sum_{i=1}^{t} \sum_{j=1}^{b} Y_{ij}^2 - \dfrac{Y_{..}^2}{bt}$		

7.5. <u>Ejemplo numérico de un diseño en bloques al azar</u>:

Título: "Ensayo de rendimiento para cuatro lineas de arroz y una variedad testigo".

En la tabla siguiente se presentan los rendimientos por parcela, expresados en toneladas por hectárea.

Variedad No.	Descripción	BLOQUE 1	2	3	Total($Y_i.$)	Promedio $\overline{Y}_i.$
1	BR-SI-46-5	6.2	3.1	7.3	16.6	5.53
2	IET-1785	9.2	5.7	8.1	23.0	7.67
3	IR2823-399	2.3	1.4	2.3	6.0	2.00
4	BG-375-1	8.4	8.4	8.5	25.3	8.43
5	CICA	2.7	5.4	7.7	15.8	5.27
Total $Y._j$		28.8	24.0	33.9		
Promedio $\overline{Y}._j$		5.76	4.80	6.78		

Gran total $Y.. = 86.7$

Promedio general $\overline{Y}.. = 5.78$

Como se dijo en el punto 6.4, las hipótesis a probar son: Primero

H_{01}: $\tau_i = 0$ que expresada en otra forma:

H_{01}: $\mu_1 = \mu_2 = \mu_3 = ...= \mu_n$ (no hay diferencia entre promedios de tratamiento)

y segundo:

H_{02}: $\beta_j = 0$ que también puede expresarse como:

H_{02}: $\mu_1 = \mu_2 = ... = \mu_n$ (no hay diferencia entre promedios de bloque)

Para probar estas hipótesis se recurre al análisis de varianza y a la prueba de F.

Para el ejemplo se procede así:

Paso 1: Se calcula el factor de corrección (FC), llamado también suma de cuadrados debidos a la media, en la siguiente forma:

$$FC = \frac{(Y..)^2}{bxt} = \frac{(86.7)^2}{3x5} \qquad \therefore \boxed{FC = 501.126}$$

Paso 2: SE halla la suma de cuadrados de tratamientos (SCTR), en este caso variedad

$$SCTR = \frac{\sum_{i=1}^{5} Y_{i.}^2}{b} - FC$$

$$SCTR = [(16.6)^2+(23.0)^2+(6.0)^2+(25.3)^2+(15.8)^2]/3-501.126$$

$$\therefore \boxed{SCTR = 75.637}$$

Paso 3: Se halla la suma de cuadrados de bloque (SCB)

$$SCB = \frac{\sum_{j=1}^{b} Y_{.j}^2}{t} - FC$$

$$SCB = [(28.8)^2+(24.0)^2+(33.9)^2]/5 - 501.126$$

$$\therefore \boxed{SCB = 9.804}$$

Paso 4: Se halla la suma de cuadrados total (SCT)

$$SCT = \sum_{i=1}^{t} \sum_{j=1}^{b} Y_{ij}^2 - FC$$

$$\therefore SCT = (6.2)^2+(3.1)^2+\ldots+(5.4)^2+(7.7)^2 - 501.126$$

$$\therefore SCT = 605.73 - 501.126$$

$$\boxed{SCT = 104.604}$$

Paso 5: La suma de cuadrados del error (SCE) se calcula por diferencia así:

$$SCE = SCT - (SCTR + 9.804)$$

$$\therefore \boxed{SCE = 19.16}$$

Paso 6: En base a los grados de libertad y a la suma de cuadrados se obtiene la tabla de análisis de varianza de:

Fuentes de Variación	GL	SC	CM	F_c	$F_{t(5\%)}$
Bloques	2	9.804	4.902	2.04	4.46
Tratamientos	4	75.637	18.909	7.89	3.84
Error	8	19.163	2.395		
$Total_{(c)}$	14	104.604			

Los cuadrados medios se calculan así:

CMB = SCB/gl.bloque: CMB = 9.804/2 $\therefore$ $\boxed{\text{CMB}=4.902}$

CMTR = SCTR/gl.Trat $\therefore$ CMTR=75.637/4 $\therefore$ $\boxed{\text{CMTR}=18.909}$

CME = SCE/GL error $\therefore$ CME=19.163/8 $\therefore$ $\boxed{\text{CME}=2.395}$

Las F calculadas para Bloque y Tratamiento son:

F_c Bloque = CMB/CME = 4.902/2.395 = 2.04

F_c Trat = CMTR/CME = 18.909/2.395 = 7.89

Paso 7: La primera hipótesis nula

$H_{01}: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

Se acepta si $F_c < F_t$

F_t se busca en las tablas de F, con los grados de libertad de tratamientos y del error, con un nivel de significancia (α) de 5%, de la siguiente forma:

GL de error	GL de tratamientos
1	4
2	
8-0.05	3.84

Como F_c es mayor que 3.84, se rechaza la hipótesis nula y se concluye que algunos μ_i (promedios de tratamiento) difieren.

La segunda hipótesis nula

$H_{02}: \mu_1 = \mu_2 = \mu_3$

Se acepta si F calculada para el efecto de bloques es menor que F_t, la cual se ubica en la tabla con los grados de libertad para bloques, y para el error. Para este ejemplo, los grados de libertad para bloques son 2 y para tratamientos 8, Al buscar en la tabla con un α (nivel de significancia) 5%.

2 GL Bloques

GL Error 8-0.05 —— $\boxed{4.46}$

Se acepta para este caso la hipótesis nula y se concluye que no existe diferencia entre los promedios de bloque.

8. EXPERIMENTOS FACTORIALES:

Se llaman asi aquellos experimentos en los cuales se estudian a la vez dos o más factores, esto permite obtener información sobre la respuesta de la planta a cada uno de los factores, y al conjunto de ellos. Estos experimentos son muy útiles en la investigación biológica, porque, en la práctica una planta o un animal se ven afectados por varios factores que influyen simultáneamente sobre ellos y que de acuerdo a la forma como se presenten: Afectan en forma diferente, el organismo animal o vegetal; es decir los cambios en un factor pueden influir sobre los efectos del otro factor. Por ejemplo: Si se abona con Nitrógeno el arroz y hay un buen control de malezas, la planta responde al elemento, pero si se abona y no hay control de malezas, no se observa respuesta en la planta; asi mismo la respuesta a Nitrógeno puede depender de la variedad, y es aconsejable para este caso, usar más de una variedad, para obtener información sobre la respuesta de las variedades.

8.1. Interacción:

Es la respuesta diferencial de un factor ante los cambios en el otro factor. Para entender con más claridad este concepto, se presentan en la figura 1, tres ejemplos de interacción. En el primero, No Interacción, se observa paralelismo entre las dos rectas, es decir, que la presencia de un factor no afecta el comportamiento del otro, en el segundo ejemplo de Mediana Interacción se observa que la linea superior tiene mayor "pendiente" que la inferior, en los dos casos hay aumentos pero estos difieren en su magnitud; en el tercer caso de Alta Interacción hay un cruce de lineas, lo cual indica que al aumentar el nivel de un factor, se aumenta la respuesta en un nivel del otro factor, pero disminuye en un segundo nivel, presentándose un caso en el cual un factor influye en forma extrema sobre otro.

8.2. Usos de los experimentos factoriales:

Este tipo de experimentos se usan con bastante frecuencia en las investigaciones agrícolas donde se prueban varios factores, y se esta interesado en algunas interacciones. Asi por ejemplo: son muy comunes los ensayos de fertilización con uno o más elementos en diferentes épocas.

Lo que en general ha pasado con la experimentación, es un cambio de experimentos monofactoriales, a experimentos factoriales, se ha ido de lo simple a lo complejo, y en esto ha jugado papel importante la estadística, al desarrollar técnicas que permiten el análisis de ese tipo de experimentos.

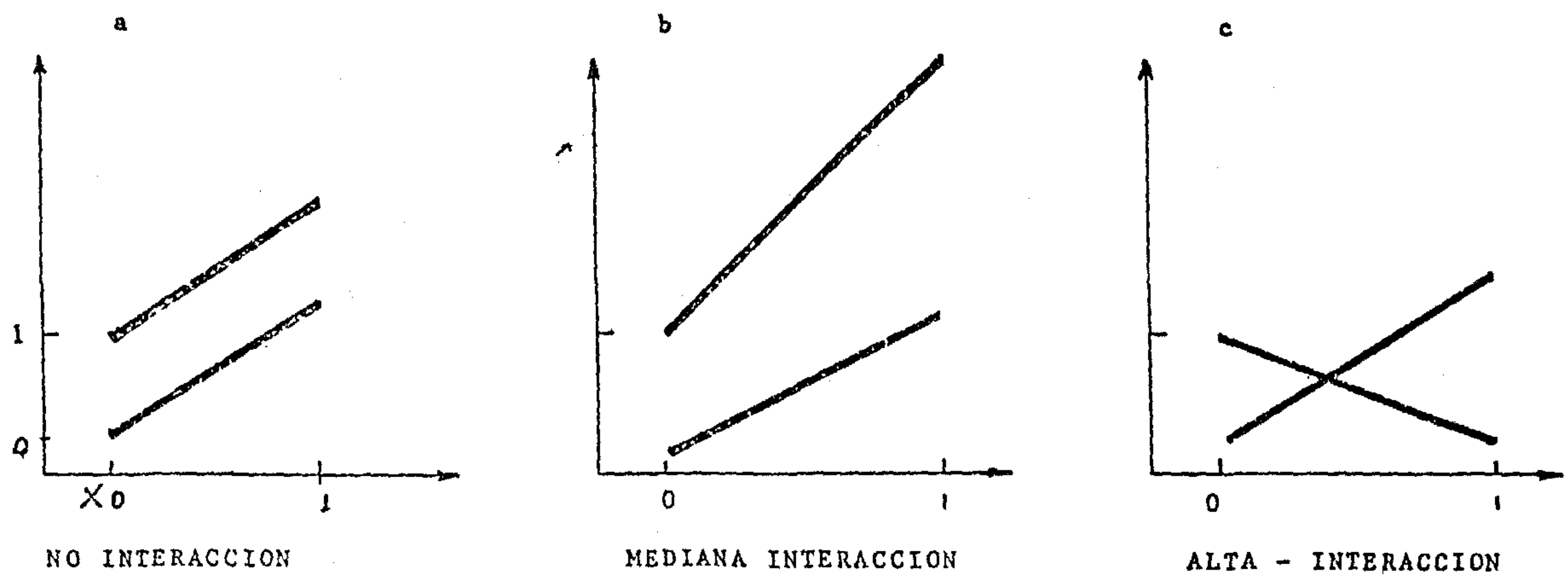


Fig. 1.: EJEMPLO DE INTERACCION ENTRE DOS FACTORES

8.3. Ejemplo de un experimento factorial, conducido bajo un diseño de bloques al azar:

Para indicar la manera como se analiza un experimento factorial, se presentan los datos de Rendimiento de arroz en Ton/ha, de dos variedades; la una de porte alto, y la otra de porte bajo, expuestas a tres niveles de fertilización con Nitrógeno (0,100, 200 Kg/ha), ver tabla 1.

Se puede apreciar que se tienen 6 tratamientos, los cuales surgen de todas las combinaciones posibles de Variedad y Nitrógeno. Este experimento es un factorial Variedad x Nitrógeno 2 x 3, es decir 2 modalidades en el factor variedad y 3 niveles en el factor Nitrógeno.

El diseño experimental utilizado fue el de Bloques al Azar, siendo el criterio de "Bloqueo" o agrupamiento de las unidades experimentales, la fertilidad del suelo; como los gradientes de fertilidad no tenian tendencias definidas, se agruparon las parcelas de forma que el Bloque quedara lo más"cuadrado posible" ver figura 2, en la cual también se muestra la aleatorización.

El modelo estadístico general, sería el de Bloques al Azar, así:

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

Tabla 1: RENDIMIENTO EN TON/HA DE DOS VARIEDADES DE ARROZ, BAJO TRES NIVELES DE NITROGENO

Trat. No.	Descripción		BLOQUE 1	2	3	4	Total Trat. $Y_{i.}$	Promedio Trat. $\overline{Y}_{i.}$
1	VAR ALTA	-No	7.5	6.0	7.0	8.5	29.0	7.25
2		-N100	8.5	6.8	7.3	8.4	31.0	7.75
3		-N200	7.6	5.9	7.3	8.2	29.0	7.25
							89.0	
4	VAR BAJA	-No	7.0	5.5	6.0	7.5	26.0	6.50
5		-N100	7.9	6.1	7.5	8.5	30.0	7.50
6		-N200	8.5	6.3	7.9	9.3	32.0	8.00
							88.0	
	Total Bloque($Y_{.j}$)		47.0	36.6	43.0	50.4	177.0	
	Promedio ($\overline{Y}_{.j}$)		7.83	6.10	7.16	8.40		

Fig. 2: Aleatorización

Diseño: Bloques al Azar

BLOQUE 1

A_0N_0	B_0N_{100}
B_0N_{200}	A_0N_{100}
B_0N_0	A_0N_{200}

BLOQUE 2

B_0N_{100}	B_0N_{200}
A_0N_{200}	A_0N_0
A_0N_{100}	B_0N_0

BLOQUE 3

A_0N_{200}	B_0N_0
A_0N_0	A_0N_{100}
B_0N_{100}	B_0N_{200}

BLOQUE 4

B_0N_0	A_0N_0
B_0N_{200}	A_0N_{200}
A_0N_{100}	B_0N_{100}

CONVENCIONES

A_0 Variedad de Arroz Alta

B_0 Variedad de Arroz Baja

N_0 0 KG de Nitrógeno aplicados

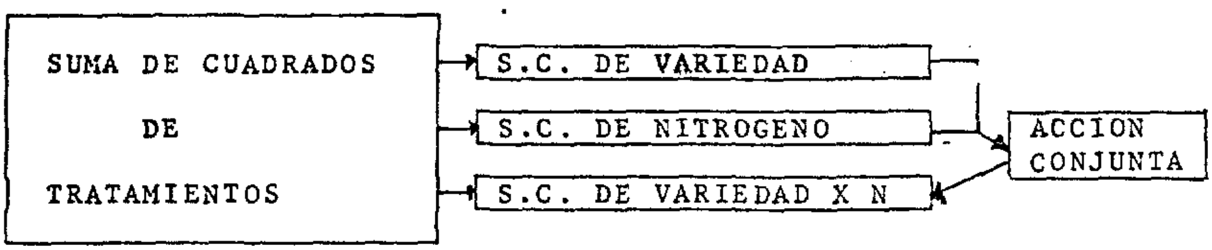
N_{100} 100 KG de Nitrógeno Aplicados

N_{200} 200 KG de Nitrógeno Aplioados

Pero debe tenerse en cuenta que el efecto de tratamiento, está compuesto de dos factores, variedad y Nitrógeno, y la interacción de ellos.

8.4. <u>Análisis de Varianza</u>:

El análisis de varianza, es el mismo que se usa, para un Diseño de Bloques al Azar, cuyos tratamientos son niveles de un solo factor, o modalidades de otro, pero descomponiendo la suma de cuadrados de tratamientos en sus respectivos componentes así:



Los calculos para sumas de cuadrados (SC) son los siguientes:

$\underline{FC} = \dfrac{Y_{..}^2}{bt}$ Siendo b = número de bloques

t = número de tratamientos

$\therefore \quad FC = \dfrac{(177)^2}{4 \times 6}$ $\boxed{FC = 1305.38}$

SCBLOQUES = SCB

$$SCB = \sum_{j=1}^{4} Y_{.j}^2/t - FC$$

$$SCB = (47.0^2+36.6^2+\ldots+50.4^2)/6 - 1305.38$$

$\therefore$ $\boxed{SCB = 17.58}$

SC TRATAMIENTOS = SCTR

$$SCTR = \sum_{i=1}^{6} Y_{i.}^2/b - FC$$

$\therefore$ $SCTR = (29.0^2+31.0^2+29.0^2+26.0^2+30.0^2+32.0^2)/4-FC$

$\therefore$ $\boxed{SCTR - 5.37}$

<u>SCTOTAL</u> = SCT

$$SCT = \sum_{i=1}^{6} \sum_{j=1}^{4} Y_{ij}^2 - FC$$

$\therefore$ $SCT = (7.5^2+6.0^2+\ldots+7.9^2+9.3^2) - 1305.38$

$\therefore$ $\boxed{SCT = 24.02}$

La <u>Suma de Cuadrados del Error</u> (SCE) se calcula por diferencia así:

SCE = SCT - SCB - SCTR

$\therefore$ SCE = 24.02 - 17.58 - 5.37

$\therefore$ $\boxed{SCE = 1.07}$

Para descomponer la Suma de <u>Cuadrados de Tratamientos</u> se procede así:

SCTR = SC Variedad+SC Nitrógeno+SC VAR x NITROGENO

Se halla $SCVARIEDAD = \sum_{vi=1}^{2} (TOTAL\ VARIEDAD)^2/bxN - FC$

Siendo N: Número de niveles de Nitrógeno

$\therefore$ $SCVARIEDAD = [(29.0+31.0+29.0)^2+(26+30+32)^2] / 4x3$

$\therefore$ $SCVARIEDAD = (89.0^2+88.0^2)/12 - 1305.38$

$\therefore$ $\boxed{SCVARIEDAD = 0.04}$

$SCNITROGENO = \sum_{n_i=1}^{3} (TOTAL\ NIVEL\ DE\ N)^2/b \times v - FC$

Para: v = número de variedades

$SCNITROGENO = [(29.0+26.0)^2+(31.0+30.0)^2+(29+32)^2] / 4x2-F($

$\therefore$ $SCNITROGENO = (55.0^2+61.0^2+61.0^2)/8 - 1305.38$

$\therefore$ $\boxed{SCNITROGENO = 3.00}$

La suma de cuadrados de la interacción Variedad x Nitrógeno se calcula por diferencia, así:

SCVARIEDAD x NITROGENO = SCTR - SCVARIEDADES - SCNITROGENO.

$\therefore$ SCVARIEDAD x NITROGENO = 5.37 - 0.04 - 3.00

$\therefore$ $\boxed{SCVARIEDAD \times NITROGENO = 2.33}$

Con estas sumas de cuadrados se puede reproducir la tabla de Análisis de Varianza:

Fuentes de Variación	G.L.	S.C.	CM	F_c	F_t 5%
Bloques	3	17.58	5.86	83.71	
Tratamientos	5	5.37	1.07	15.29	
- Variedad	1	0.04	0.04	0.57	
- Nitrógeno	2	3.00	1.50	21.43	
- Var x Nitrógeno	2	2.33	1.17	16.71	
Error	15	1.07	0.07		
$Total_{(c)}$	23	24.02			

Como se observa en la tabla de análisis de varianza, se puede dividir la suma de cuadrados de tratamientos, y se prueban aisladamente cada uno de los efectos, encontrándose que rechazó la hipótesis nula para:

Bloques Lo cual indica que el críterio de bloqueo fue efectivo.

En cuanto a tratamientos se dividió la SC y se encontró:

a) Que se acepta la hipótesis nula para el efecto variedad, y se concluye que no hay efecto varietal.

b) Que hay diferencia entre los niveles de Nitrógeno.

c) Que hay interacción entre variedsd y nitrógeno.

Este último lateral (c), nos está indicando que las variedades de arroz del experimento, responden en forma diferencial a la aplicación de Nitrógeno lo cual puede apreciarse en mejor forma en la figura 3.

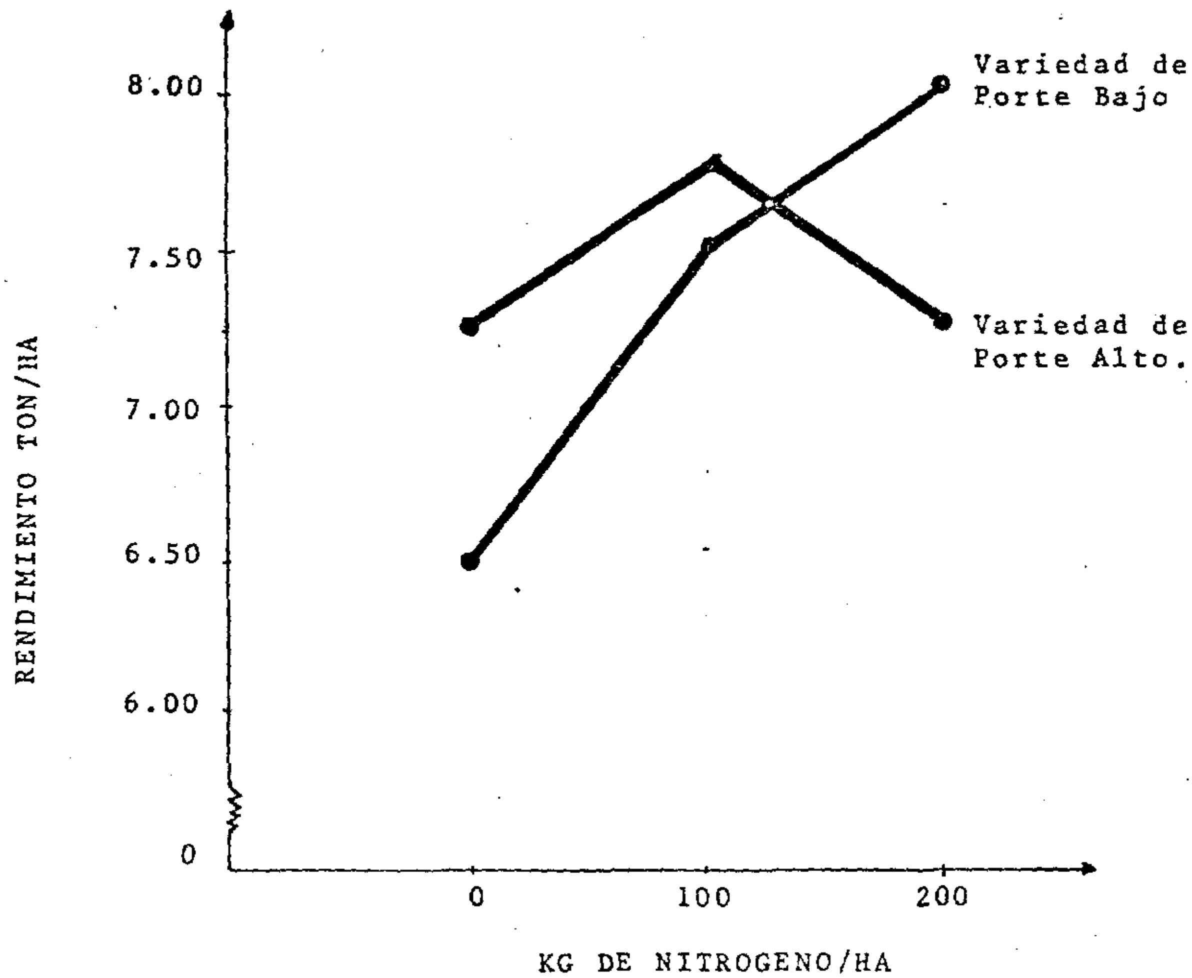


FIG. 3.: INTERACCION VARIEDAD Y NITROGENO

9. DISEÑO EN PARCELAS DIVIDIDAS:

Realmente no es un diseño experimental, sino un "arreglo de tratamientos".

9.1. Características:

Las características principales son:

- Se usa solamente para experimentos factoriales, generalmente cuando se estudian dos factores.
- Se utilizan dos tamaños de parcela, la de tamaño mayor se denomina parcela principal y la parcela de tamaño menor subparcela.
- La aleatorización se cumple en dos etapas, en la primera se asignan al azar las parcelas principales y en la segunda se asignan al azar las subparcelas en las parcelas principales.
- Se compara con mayor precisión el factor que se asigna en las subparcelas.
- Es ventajoso cuando por alguna razón práctica, un factor debe tener parcelas grandes, o cuando se desea comparar con mayor precisión un factor que otro.

En la mayoria de los casos, este arreglo, se realiza bajo un diseño en bloques al azar, razón por la cual, en la aleatorización y en el modelo estadístico se tendrá en cuenta el efecto de "Bloque".

PARCELAS DIVIDIDAS

MODELO ESTADISTICO

$$Y_{ijk} = \mu + R_i + P_j + (RxP)_{ij} + S_k + (PxS)_{jk} + (RS)_{ik} + (RxPxS)_{ijk}$$

Error B - Entre subparcelas

Efecto de la interacción parcela j por subparcela k.

Efecto de la subparcela k

Error A - Entre parcelas principales

Efecto de la parcela principal j

Efecto de la repetición i

Efecto de la media general

Observación de la unidad experimental i,jk.

En experimentación con arroz, ha sido muy útil porque hay factores como Nitrógeno, o riego que son muy difíciles de manejar en parcelas pequeñas. En el caso del Nitrógeno, porque habría movilización de este elemento de una parcela a otra, y habría que construir "Caballones" para cada parcela, y en el caso del riego, es muy difícil "inundar" una parcela, y no hacerlo con sus vecinas.
El investigador debe tener presente: "Cuál es el factor que quiere comparar con más precisión" y en base a ello, asignar ese factor a las subparcelas.

El modelo estadístico de este "diseño" se presenta en la figura 1. Note que aparecen dos tipos de Error, Error (a) y Error (b), el Error (a) sirve para comparar el efecto de repetición, y el efecto del factor asignado en la parcela principal, y el Error (b) sirve para comparar el efecto del factor asignado en la subparcela, y la interacción de los dos factores.
El esquema de las fuentes de variación y grados, de libertad se presenta a continuación:

Fuente de Variación	G.L.
Repetición	r-1
Factor A	a-1
Error (a)	(r-1)(a-1)
Factor B	b-1
Factor A x Factor B	(a-1)(b-1)
Error (b)	a(r-1)(b-1)
$\text{Total}_{(c)}$	rab - 1

9.2. Ejemplo numérico del diseño en parcelas divididas:

Se quiere comparar el efecto del Nitrógeno a 2 niveles 0, y 100 Kg/ha sobre cuatro variedades de arroz: CICA-7, CICA-8, CICA-9 e IR-8.

Para evitar la construcción de caballones alrededor de cada parcela, se eligió el "diseño" en parcelas divididas. La forma de aleatorización se presenta en sus dos etapas en la figura 2. Primero se asignan al azar los niveles de Nitrógeno, a las parcelas mayores correspondiendole a cada nivel cuatro subparcelas (Fig. 2a) y posteriormente se asignan las variedades a las parcelas principales (Fig. 2b) quedando cada variedad en una subparcela.

En este experimento se espera comparar con mayor precisión el efecto de variedad, y la interacción Variedad x Nitrógeno, por tal motivo se asigna el

FIG. 2: Aleatorización en el Diseño en parcelas divididas

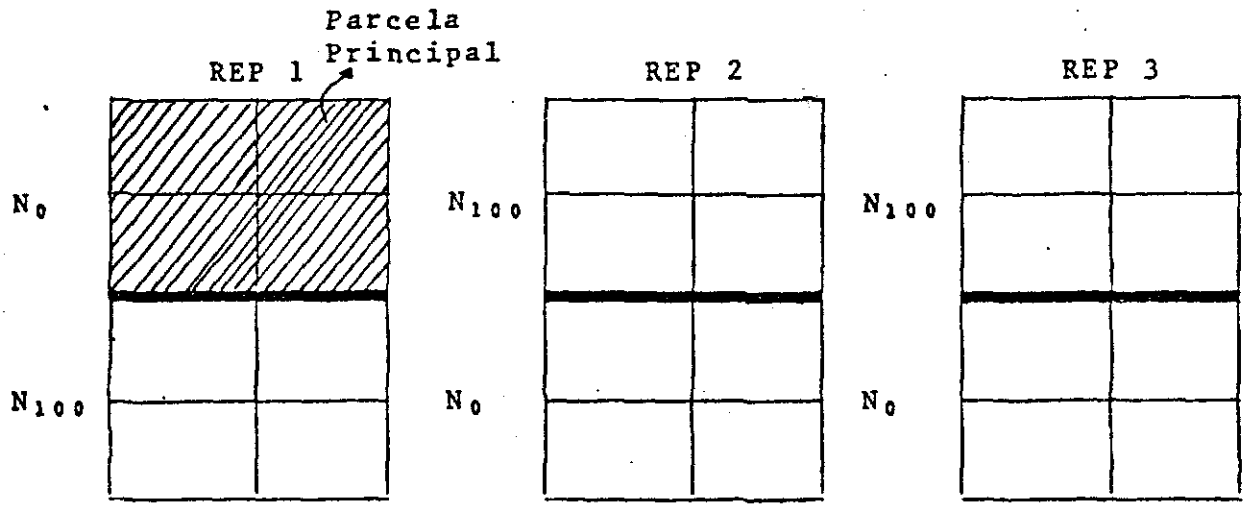


FIG. 2.A.: PRIMERA ETAPA DE ALEATORIZACION

Se asignan las parcelas principales

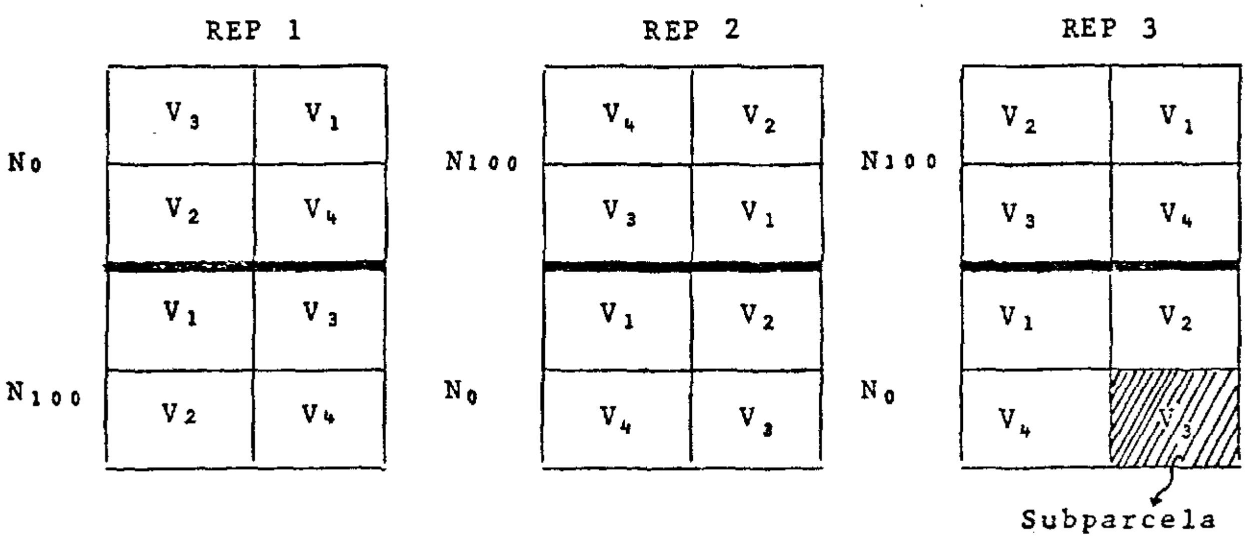


FIG. 2.B.: SEGUNDA ETAPA DE ALEATORIZACION

Se asignan las subparcelas a las parcelas principales.

nivel de Nitrógeno en la parcela principal, y las variedades en la subparcela.

Es importante recordar que, el tamaño de la subparcela, determina el de la parcela principal, y no al contrario.

En tabla 1, se presentan por repetición los resultados obtenidos, en Ton/ha.

9.3. Análisis de Varianza:

Para i: Repetición (i=1, 2, 3)

j: Nivel de N (j=1, 2)

K: Variedad (k=1, 2, 3, 4)

$$FC = [\Sigma Y_{ijk}]^2 / r \times n \times v$$

Siendo: r = número de repeticiones (3)

n = niveles de Nitrógeno (2)

v = número de variedades (4)

$\therefore$ $FC = (123.9)^2/24$

$\therefore$ $\boxed{FC = 639.63}$

SUMA DE CUADRADOS DE BLOQUES = SCB

$$SCB = \frac{Y_{i..}}{n x v}^2 - FC$$

$$SCB = [(39.5)^2+(38.4)^2+(46.0)^2]/8 - 639.63$$

Tabla 1: Rendimiento de arroz en Ton/ha, para cuatro variedades y dos niveles de nitrógeno.

		BLOQUE			
		1	2	3	
No.	CICA-7	3.9	3.6	5.9	13.4
	CICA-8	5.8	5.7	7.3	18.8
	CICA-9	5.6	5.8	6.0	17.4
	IR-8	2.5	2.7	3.5	8.7
		17.8	17.8	22.7	58.3
No. 100	CICA-7	4.4	4.0	6.0	14.4
	CICA-8	6.3	6.4	7.5	20.2
	CICA-9	7.0	6.4	6.1	19.5
	IR-8	4.0	3.8	3.7	11.5
		21.7	20.6	23.3	65.6
		39.5	38.4	46.0	123.9

TOTALES DE VARIEDAD

V_1 27.8

V_2 39.0

V_3 36.9

V_4 20.2

$\therefore$ $\boxed{SCB = 4.217}$

SUMA DE CUADRADOS DE NITROGENO = SCN

$$SCN = \frac{\Sigma Y._j.^2}{r \times v} - FC$$

$$SCN = [(58.30) + (65.60)^2] / 12 - 639.63$$

$\therefore$ $\boxed{SCN = 2.220}$

SUMA DE CUADRADOS DE PARCELAS PRINCIPALES DE N=SC (NPPAL)

$$SC(N\ PPAL) = (17.8^2+17.8^2+22.7^2+20.6^2+23.3^2)/v - FC$$

$\boxed{SC(N\ PPAL) = 7.148}$

SUMA DE CUADRADOS DEL ERROR A = SCE(a)

$$SCE(a) = 7.148 - SCB - SCN$$

$\therefore$ $SCE(a) = 7.148 - 4.217 - 2.220$

$\therefore$ $\boxed{SCE(a) = 0.711}$

SUMA DE CUADRADOS DE VARIEDADES = (SCV)

$$SCV = \frac{Y_{..j}^2}{r \times n} - FC$$

$\therefore$ $\boxed{SCV = 37.61}$

<u>SUMA DE CUADRADOS DE TRATAMIENTOS:</u> SCTR

$SCTR = (13.4^2+18.8^2+...+11.5^2)/3 - FC$

$\therefore$ $\boxed{SCTR = 40.15}$

<u>SUMA DE CUADRADOS DE VARIEDAD x NITROGENO:</u> SCV x N

$SCV \times N = SCTR - SCN - SCV$

$\therefore$ $SCVxN = 40.15 - 2.22 - 37.61 = 0.32$

$\therefore$ $\boxed{SCV \times N = 0.32}$

<u>SUMA DE CUADRADOS DEL ERROR B</u> = SCE(b)

Se calcula por diferencia de:

Suma de cuadrados Total: DCT - SC otros efectos

Para $SCT = \Sigma Y_{ijk}^2 - FC$

$\therefore$ $\boxed{SCT = 49.12}$

$$y\ SCE(b) = 49.12 - (SCB + SCN + SCE(a) + SCV + SCVxN)$$

$$= 49.12 - (4.217+2.220+0.711+37.61+0.32)$$

$\therefore$ $\boxed{SCE(b) = 4.048}$

Con estas sumas de cuadrados se construye la tabla 2.

Tabla 2: Análisis de Varianza para la variable Rendimiento de arroz en Ton/ha. Efecto de 2 niveles de Nitrógeno en cuatro variedades de arroz.

Fuentes de Variación	G.L.	SC.	CM	F_c
Repetición	2	4.217	2.109	5.99
Nitrógeno	1	2.220	2.220	6.31
Error(a)	2	0.705	0.352	
Variedad	3	37.615	12.538	37.20
Var x Nitrógeno	3	0.315	0.105	< 1
Error (b)	12	4.048	0.337	
$\text{Total}_{(c)}$	23	49.120		

CV = 11.2% E(b)

$\bar{X}$ = 5.16 Ton/ha.