

CAPITULO 4

DISEÑOS EN BLOQUES

4.1.BLOQUES AL AZAR.

4.1.1.Introducción

El diseño de experimentos busca lograr precisión en las conclusiones a mínimo costo. Este objetivo se ve afectado por dos tipos de variabilidad en los datos: variabilidad controlable y variabilidad no controlable por un diseño. La variabilidad no controlable se trata por medio de la aleatorización y la variabilidad controlable se trata de acuerdo al principio de control local. Si un efecto de tratamiento resulta impreciso debido a que el error experimental impide su captación el experimentador tiene a su alcance dos vías de acción. Mendenhall (1971) llama a éstas hacer diseños reductores del ruido (error) o incrementadores del volumen de la respuesta, partiendo de que el error experimental es como un ruido que dificulta la captación del mensaje que proporciona el tratamiento. Para reducir el error se cuenta con el bloqueo y el uso de covariables. El éxito de ambos mecanismos depende del conocimiento adecuado del material experimental para formar grupos apropiados de unidades.

Ejemplo 4.1. Supongamos que se dispone de un terreno experimental en pendiente donde se intenta probar cuatro variedades de un cultivo. El terreno se puede dividir en cuatro franjas al mismo nivel que son razonablemente de igual fertilidad, humedad, etc. La respuesta esperada de las parcelas de una franja dada difiere de la respuesta esperada en parcelas de las otras franjas, y se puede evitar que esa variabilidad identificada integre el error. Para ello se considera a cada faja como un bloque y se divide en cuatro parcelas que se adjudican al azar a cada variedad. La diferencia entre bloques afectará igualmente a cada variedad y la comparación entre variedades es independiente (ortogonal como se dice en la terminología estadística) al efecto de los bloques. Si, por otro lado, las franjas de terreno interaccionan con las variedades no pueden ser consideradas como bloques en el sentido en que acá lo usamos.

Figura 4.1. Bloqueo de parcelas.

Ejemplo 4.2. Falconer (1960) cita un caso de interacción entre dos factores de clasificación. Dos genotipos se prueban en dos ambientes A y B. Si el genotipo A es mejor en el ambiente 1 pero el B

es mejor en el ambiente 2 se dice que hay una interacción genotipo-ambiente. Si eso no sucede se considera que el genotipo y el ambiente son independientes y los ambientes actúan como bloques en la evaluación de los genotipos.

Ejemplo 4.3. Cuando se cuenta con cerdos mellizos para probar una ración alimenticia resulta conveniente sacar ventaja de las características del material experimental para reducir el error. Se adjudica al azar un animal dentro de cada par para ser tratado y se deja al otro como testigo. El diseño resultante en bloques de dos parcelas, se conoce como de parcelas apareadas.

A veces sucede que el material experimental parece homogéneo pero se prevee que habrá diferente manejo, por lo que puede resultar de utilidad formar bloques por grupos de parcelas a aplicar manejo similar.

Ejemplo 4.4. Si se quieren probar cinco raciones en vacas lecheras estabuladas no siempre se cuenta con animales de parición en época similar. Se toman las cinco primeras paridas y se les adjudica al azar un tratamiento a cada una. Tan pronto como se cuente con otras cinco se puede iniciar un nuevo grupo y el efecto de la fecha de parición pasa a constituir el efecto de bloque.

Ejemplo 4.5. Cochran y Cox (1978) comentan con respecto al hecho de que a veces no se previenen bloques al iniciarse el experimento. Si se prueban cinco productos para combatir a las malezas y se mide el efecto por conteo de malezas, toda lluvia que caiga atrasa el trabajo, además el conteo luego de una lluvia se verá afectado. En casos como este, al investigador le conviene contar todas las parcelas de un grupo que incluya igual número de parcelas de cada tratamiento por día, y si se presenta alguna distorsión, la puede aislar como efecto de bloque.

En cuanto a la forma del bloque y de las parcelas existen algunas pautas generales. Una de ellas es la de formar bloques compactos, al ser las parcelas vecinas de rendimiento similar. En algunos casos esa pauta debe dejarse de lado, como en el ejemplo 6.1 donde los bloques deben seguir las curvas de nivel. Las parcelas, al contrario que los bloques, interesa que sean iguales de modo que generalmente se tenderá al uso de parcelas alargadas y que, teniendo partes distintas, se lleven toda la heterogeneidad. En caso del ejemplo 6.1 deberán hacerse siguiendo la pendiente y perpendiculares a los bloques (Vase Li [1969]).

Ejemplo 4.6. Al hacer bloques con animales para evaluar el crecimiento, los bloques deben hacerse con animales lo mas parecidos posible en un bloque. No debe buscarse la igualdad entre los bloques de modo que "los promedios de todos los bloques sean iguales". Este objetivo se logra muchas veces poniendo en un mismo bloque animales de distinto peso con lo que se incrementa el error experimental.

4.1.2. Modelo para bloques al azar

El modelo para diseño en bloques al azar es: $Y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}$ $\epsilon_{ij} \sim \text{NID}(0; \sigma^2)$ donde se expresa que cada observación, en particular la observación que está en el bloque i y bajo el tratamiento j , es la suma de una media general μ ; un efecto del bloque en el que se encuentra: β_i ; un efecto del tratamiento que se le aplica: τ_j y un error aleatorio: ϵ_{ij} . Tomaremos como efecto del bloque la diferencia entre el promedio del bloque y la media general $\beta_i = \mu_i - \mu$; de modo que la suma de los efectos de bloque es nula $\sum \beta_i = \sum (\mu_i - \mu) = 0$. Del mismo modo el efecto del tratamiento será la diferencia entre el promedio de las parcelas sometidas a éste tratamiento y la media general: $\tau_j = \mu_j - \mu$ y $\sum \tau_j = 0$. El valor del error que afecta a cada parcela proviene al azar de una población de errores con distribución normal, media cero y varianza común sea cual sea el tratamiento o bloque en que actúe.

Las estimaciones de los parámetros del modelo que se obtienen por mínimos cuadrados son:

$$\hat{\mu} = \bar{Y}_{..} = \text{media general del conjunto de datos}, \hat{\tau}_j = t_j = \bar{Y}_{.j} - \bar{Y}_{..}, \hat{\beta}_i = b_i = \bar{Y}_{i.} - \bar{Y}_{..}$$

$$\hat{\epsilon}_{ij} = e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - (\bar{Y}_{..} + b_i + t_j) = Y_{ij} - \bar{Y}_{.j} - \bar{Y}_{i.} + \bar{Y}_{..}$$

Ejemplo 4.8. Los siguientes datos provienen del experimento que se citó anteriormente:

Tratamientos	1	2	3	4	T_i	$\bar{Y}_{i.}$	b_i
Bloques I	37	20	41	25	123	30,75	0,75
II	31	16	44	22	113	28,25	-1,75
III	38	15	48	15	116	29,00	-1,00
IV	34	29	47	18	128	32,00	2,00
Total T_j	140	80	180	80	480		
	35	20	45	20		30	
t_j	5	-10	15	-10			0

El ajuste de los datos será $\hat{\mu} = \bar{Y}_{..} = 30$, $\hat{\tau}_1 = t_1 = \bar{Y}_{.1} - \bar{Y}_{..} = 35 - 30 = 5$, $t_2 = 20 - 30 = -10$, $t_3 = 45 - 30 = 15$ y $t_4 = 20 - 30 = -10$, de modo que $\sum T_j = 0$. Las estimaciones de los efectos de los bloques serán $\hat{\beta}_j = b_j = \bar{Y}_{j.} - \bar{Y}_{..} = 30,75 - 30 = 0,75$, $b_2 = 28,75 - 30 = -1,75$, $b_3 = 29 - 30 = -1$ y $b_4 = 32 - 30 = +2$ y también así la suma de las estimaciones debe ser nula: $\sum b_i = 0$

$$\sum_{i=1}^n b_i = 0,75 + (-1,75) + (-1) + 2 = 0$$

Finalmente cada $\hat{Y}_{..} = \bar{Y}_{..} + t_j + b_i$ y $\hat{\epsilon}_{ij} = e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - (\bar{Y}_{..} + b_j + t_i) = Y_{ij} - \bar{Y}_{.j} - \bar{Y}_{i.} + \bar{Y}_{..}$

4.1.3. Análisis de varianza para bloques al azar

De acuerdo al modelo $Y_{ij} = \bar{Y}_{..} + b_i + t_j + e_{ij}$ restando $\bar{Y}_{..}$ de ambos lados $Y_{ij} - \bar{Y}_{..} = b_i + t_j + e_{ij}$ y elevando al cuadrado y sumando para todo i y todo j

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = k \sum b^2 + n \sum t^2 + \sum \sum e^2$$

y como $\sum \sum (Y_{ij} - \bar{Y}_{..})^2 = SC$ = suma de cuadrados total que mide la variación del conjunto de nk datos en torno a la media general.

$\sum b^2 = k \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2$ = suma de cuadrados de los bloques o explicada por los bloques.

$n \sum t^2 = n \sum (\bar{Y}_{.j} - \bar{Y}_{..})^2$ = suma de cuadrados de los tratamientos o explicada por los tratamientos

$\sum \sum e^2_{ij} = \sum \sum (Y_{ij} - \hat{Y}_{ij})^2$ = suma de cuadrados del error o del residuo no explicada por el modelo.

Tenemos $SC = SCB + SCT + SCE$ lo que se puede expresar en un cuadro de análisis de la varianza, donde se muestran además las fórmulas usuales de cálculo abreviado.

FUENTE VARIACIÓN	SUMAS CUADRADOS	GRADOS LIBERTAD	CUADRADO MEDIO	F
Bloques	$\frac{\sum T_i^2}{k} - C$	$n-1$	SCB/GLB	CMT/CME
Tratamientos	$\frac{\sum T_j^2}{n} - C$	$k-1$	SCT/GLT	
Error o residuo	por diferencia	$(n-1)(k-1)$	SCE/GLE	
TOTAL		$nk-1$		

$$y \quad C = \frac{T_{..}^2}{nk}$$

Ejemplo 4.8 (Continuación). En el caso de los datos del ejemplo tenemos: $C = \frac{480^2}{4 \times 4} = 14.400$

$SC = 37^2 + \dots + 18^2 - C = 2.040$ y $SCB = \sum T_{i.}^2 / k - C = (123^2 + 113^2 + 116^2 + 128^2) / 4 - C = 34,5$

$SCT = \sum T_{.j}^2 - C = (14^2 + 80^2 + 180^2 + 80^2) / 4 - C = 1.800$ y $SCE = SC - SCB - SCT = 205,5$ quedando el análisis de varianza del siguiente modo:

FUENTE VARIACIÓN	SUMAS CUADRADOS	GRADOS LIBERTAD	CUADRADO MEDIO	F
Bloques	34,5	3	11,5	26,27
Tratamientos	1800,0	3	600	
Error o residuo	205,5	9	22,83	
TOTAL	2040,0	15		

$F_{0,95} = 3,86$ y $F_{0,99} = 6,99$

Es interesante notar que si los datos no son ortogonales, la suma de cuadrados de tratamientos varía según se incluyan o no bloques en el modelo. Esta característica resulta difícil de entender para algunos, y es siempre difícil de valorar en el análisis. Por lo tanto, es muy aconsejable buscar la ortogonalidad siempre que sea posible.

4.1.4.Caso particular de dos tratamientos: Parcelas apareadas.

Cuando el bloque está integrado por solamente dos parcelas se habla de parcelas apareadas, y constituye un caso digno de destacarse entre los diseños en bloques, ya que el efecto de los tratamientos puede ser probado mediante la prueba F o la t sin ninguna diferencia, debido a que ambas coinciden exactamente. Este tipo de resultados revelan la generalidad de la prueba F y el hecho de que la t es un caso particular de ella.

Ejemplo 4.9. Analizaremos mediante ambas técnicas la diferencia entre las medias de dos muestras dadas abajo. En el ejemplo 1.16 (sección 1.3.4.2) se analizó la diferencia entre las medias considerando las observaciones apareadas y usando la variable diferencia.

Muestra 1	Muestra 2	d	d ²
161,30	149,64	11,66	135,96
148,26	163,30	-15,04	226,20
142,99	152,68	-9,69	93,90
184,47	161,64	22,83	521,21
146,69	157,69	-11,00	121,00
164,11	146,08	18,03	325,08
162,31	170,04	-7,73	59,75
171,22	173,27	-2,05	4,20
170,08	146,70	23,38	546,62
161,27	157,89		

$$t_{(8)} = \frac{\bar{d} - \mu_{\bar{d}}}{\hat{\sigma}_{\bar{d}}} = 0,65 \text{ (ns)}$$

FUENTE VARIACIÓN	SUMAS CUADRADOS	GRADOS LIBERTAD	CUADRADO MEDIO	F
Bloques	51.34222	1	51.34222	0,43
Tratamientos	1264.33000	8	158.04125	1,31
Error o residuo	965.46778	8	120.68347	
TOTAL	2281.14000	17		

con lo que se comprueba que el resultado es el mismo ya que $0,65^2 = 0,43$.

4.1.5.USO DE SAS

El ejemplo 4.8 analizado en el SAS seria:

```
DATA cuatro8;
  INPUT b t y @@;
  CARDS;
1 1 37 1 2 20 1 3 41 1 4 25
2 1 31 2 2 16 2 3 44 2 4 22
3 1 38 3 2 15 3 3 48 3 4 15
4 1 34 4 2 29 4 3 47 4 4 18
```

```
PROC ANOVA;
  CLASS b t;
  MODEL y=b t;
  MEANS t/LSD LINES;
  RUN;
```

la salida sería:

Analysis of Variance Procedure						
Class Level Information						
Class		Levels		Values		
B		4		1	2	3 4
T		4		1	2	3 4
Number of observations in data set = 16						
Analysis of Variance Procedure						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	6	1834.5000000	305.7500000	13.39	0.0005	
Error	9	205.5000000	22.8333333			
Corrected Total	15	2040.0000000				
R-Square		C.V.	Root MSE	Y Mean		
0.899265		15.928079	4.7784237	30.0000000		
Analysis of Variance Procedure						
Dependent Variable: Y						
Source	DF	Anova SS	Mean Square	F Value	Pr > F	
B	3	34.500000	11.500000	0.50	0.6892	
T	3	1800.000000	600.000000	26.28	0.0001	

el análisis de varianza esta de acuerdo a lo visto en la pagina 169.

La salida para la parte de separacion de medias es:

Analysis of Variance Procedure			
T tests (LSD) for variable: Y			
NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.			
Alpha= 0.05 df= 9 MSE= 22.83333			
Critical Value of T= 2.26			
Least Significant Difference= 7.6435			
Means with the same letter are not significantly different.			
T Grouping	Mean	N	T
A	45.000	4	3
B	35.000	4	1
C	20.000	4	2
C	20.000	4	4

Un segundo ejemplo corrido en el SAS es:

```
DATA uno;
  INPUT trat $ n p b rend;
  CARDS;
0-0      0    0 1 161.3
60-0     60   0 1 148.2
120-0    120  0 1 142.9
0-80     0    80 1 184.4
60-80    60   80 1 146.6
120-80   120  80 1 164.1
0-160    0   160 1 162.3
60-160   60  160 1 171.2
120-160  120 160 1 170.0
0-0      0    0 2 149.6
60-0     60   0 2 163.3
120-0    120  0 2 152.6
0-80     0    80 2 161.6
60-80    60   80 2 157.6
120-80   120  80 2 146.0
0-160    0   160 2 170.0
60-160   60  160 2 173.2
120-160  120 160 2 146.7
;
```

* Esta parte compara el PROC ANOVA con el PROC GLM;

```
PROC ANOVA;
  CLASS trat b;
  MODEL rend=b trat;
  RUN;
PROC GLM;
  CLASS trat b;
  MODEL rend=b trat;
  RUN;
```

Analysis of Variance Procedure - Class Level Information									
Class	Levels	Values							
TRT	9	0-0 0-160 0-80 120-0 120-160 120-80 60-0 60-160 60-80							
B	2	1 2							
Number of observations in data set = 18									
Analysis of Variance Procedure - Dependent Variable: REND									
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F				
Model	9	1315.67222	146.18580	1.21	0.3988				
Error	8	965.46778	120.68347						
Corrected Total	17	2281.14000							
	R-Square	C.V.	Root MSE		REND Mean				
	0.576761	6.886086	10.9856		159.53333				
Source	DF	Anova SS	Mean Square	F Value	Pr > F				
B	1	51.34222	51.34222	0.43	0.5325				
TRT	8	1264.33000	158.04125	1.31	0.3560				

General Linear Models Procedure										
Class Level Information										
Class	Levels	Values								
TRT	9	0-0 0-160 0-80 120-0 120-160 120-80 60-0 60-160 60-80								
B	2	1 2								
Number of observations in data set = 18										
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F					
Model	9	1315.67222	146.18580	1.21	0.3988					
Error	8	965.46778	120.68347							
Corrected Total	17	2281.14000								
		R-Square	C.V.	Root MSE	REND Mean					
		0.576761	6.886086	10.9856	159.53333					
Source	DF	Type I SS	Mean Square	F Value	Pr > F					
B	1	51.34222	51.34222	0.43	0.5325					
TRT	8	1264.33000	158.04125	1.31	0.3560					
Source	DF	Type III SS	Mean Square	F Value	Pr > F					
B	1	51.34222	51.34222	0.43	0.5325					
TRT	8	1264.33000	158.04125	1.31	0.3560					

4.1.5. Eficiencia del diseño.

El diseño en bloques al azar reduce el error al identificar y aislar variación experimental. No obstante cuando el experimentador diseña su ensayo se le presenta a menudo la duda de formar o no bloques si el factor presenta variación.

Ejemplo 4.10. Conversaciones con dos investigadores." Quiero probar 3 tratamientos. Tomamos 24 vacas, las ordenamos por producción de leche, porcentaje de grasa, número de lactancias y peso del animal. Luego las dividimos en grupos de 3 (*). Se elige una vaca de cada uno de los 8 grupos y se forman de ese modo 3 grupos de 8. Cada vaca es un bloque." Hay varios problemas semánticos en eso. Si al nivel del asterisco (*) les llamamos bloques se nos simplifican las definiciones. Dentro de cada bloque se adjudica una vaca a cada tratamiento. Se puede decir aun peor: "ordenamos las 24 vacas, hacemos grupos con una de las tres primeras, otra del siguiente trio, etc. A cada bloque (de 8 vacas) lo sometemos a un tratamiento." Pues bien, ese grupo de 8 vacas no es un bloque sino el conjunto de vacas sometidas a un tratamiento. El bloque tiene que tener (en general) tantas unidades experimentales como tratamientos, esas unidades son parecidas (lo mas parecidas posibles), dentro de cada bloque se adjudican al azar los animales a cada tratamiento. Otra observación: cuidado con decir "se hace un grupo y se le somete al tratamiento A". Con decir no hay problema, pero que no se trate a los animales sometidos al tratamiento A todos juntos, sino i) todos (los 24 animales) juntos o ii) juntos por bloques (de 3 vacas). En caso contrario puede haber un efecto sistemático que afecte a las 8 vacas del tratamiento A y que no afecte a las de los otros tratamientos.

Vale la pena insistir en que las unidades experimentales de un bloque deben ser (todo lo posible) iguales y tratadas igual.

Ejemplo 4.7. Los siguientes datos son de un ensayo en lechería diseñado en la EEMAC. Se contaba con el número de vacas mostrado y se querían probar 3 tratamientos. Se decidió pues hacer bloques de 3 vacas. De acuerdo al criterio del experimentador el factor a considerar en primer lugar (el factor mas importante) era la lactación. Por lo tanto se dividió a los animales en los de primera lactancia, los de segunda y los de tercera y más. El siguiente criterio en orden de prioridad era el mes de lactancia. El tercer criterio para bloquear fue la producción promedio en la semana anterior al ensayo. Los valores de la tabla estan en orden decreciente de produccion. Como habia 6 vacas de primera lactancia se decidió hacer dos bloques pese a que una de ellas era de segundo mes de producción.

ENSAYO DE SUPLEMENTACION DE INVIERNO

* PROMEDIO *

NO * PV LACT MES PRODa120M 22a128 *BLOQUE

527 *	430	1	2	17.2	18.2 *	I
614 *	475	1	1	16.0	16.8 *	I
622 *	360	1	1	17.1	16.2 *	I
617 *	439	1	1	16.8	16.1 *	II
654 *	370	1	1	17.2	15.7 *	II
605 *	505	1	1	12.6	14.6 *	II
451 *	450	2	1	19.5	20.3 *	III
506 *	442	2	1	19.5	20.1 *	III
553 *	440	2	2	20.0	19.9 *	III
427 *	498	2	1	19.6	19.7 *	IV
418 *	500	2	1	20.7	19.3 *	IV
537 *	432	2	1	17.5	18.3 *	IV
505 *	434	2	2	19.4	17.7 *	V
507 *	448	2	2	18.4	17.7 *	V
439 *	470	2	1	16.1	16.7 *	V
346 *	506	3	1	19.9	20.9 *	VI
941 *	466	7	2	20.6	20.3 *	VI
904 *	516	5	1	21.6	20.1 *	VI
329 *	556	3	2	19.3	19.8 *	VII
307 *	470	4	1	19.3	19.6 *	VII
450 *	520	3	2	17.5	18.4 *	VII
908 *	488	6	2	14.4	16.9 *	VIII
950 *	542	7	1	17.3	16.8 *	VIII
226 *	520	4	2	17.1	16.4 *	VIII
727 *	520	8	5	14.5	15.1 *	IX
833 *	540	7	7	14.8	14.5 *	IX
210 *	477	4	5	13.5	13.3 *	IX
543 *	463	1	5	14.7	12.5 *	X
651 *	430	1	8	13.7	13.7 *	X
409	508	2	7	13.2	13.1	X
542 *	429	1	5	17.0	16.2 *	XI
429 *	467	2	8	14.1	16.2 *	XI
445 *	460	2	6	16.5	15.2 *	XI

Del mismo modo se tuvo que obviar el criterio en los bloques III, IV y V. Si se dejaba las tres vacas de segundo mes de parición en el bloque V se incluía una de producción muy superior a las otras dos. El investigador optó por darle preferencia a la producción por sobre la etapa de la lactancia. En los bloques VI a VIII se manejaron las lactancias medias y etapas iniciales de lactancia. En el bloque IX se usaron las vacas de edad y lactancias avanzadas. En los X y XI las vacas nuevas de lactancia avanzada, también acá se dio preferencia a la producción por sobre la decisión de mantener vacas de segundo mes de lactancia juntas. Una de las razones invocadas por el investigador para no confiar tanto en el criterio mes de lactancia, fué el hecho de que un indicador mas exacto de la etapa de la lactancia en que se encontraba el animal hubiera sido días de lactancia pero se no se contaba con ese dato a mano

125 *	510	6	1	16.2	14.5 *	
944 *	484	6	1	16	16.4 *	NO
338 *	510	2	8	12.2	12.1 *	SECAR
534 *	480	2	1	15.3	16.4 *	
621 *	410	1	1	9.6	13.7 *	
510 *	450	2	5	12.6	12.1 *	

515 * 445 2 7 12 11.2 *

Las vacas que se muestran en último lugar no se usaron. Se considero que aportaban mas variabilidad que precisión al ensayo al no tener con quien colocarlas en un mismo bloque.

El sorteo se hizo del siguiente modo: se adjudico un número al azar a cada vaca. Luego se las ordenó dentro de cada bloque por ese número al azar. Las que estaban en el primer lugar se destinaron al tratamiento A, las del segundo al B y las del tercero al C. Las vacas ordenadas por tratamiento y bloque estan en la siguiente tabla.

ENSAYO DE SUPLEMENTACION DE INVIERNO. Situacion final.

*

PROMEDIO *

NO * PV LACT MES PRODa120M 22a128 *BLOQUE

527 *	430	1	2	17.2	18.2 *	I	A
605 *	505	1	1	12.6	14.6 *	II	A
506 *	442	2	1	19.5	20.1 *	III	A
537 *	432	2	1	17.5	18.3 *	IV	A
950 *	542	7	1	17.3	16.8 *	IX	A
439 *	470	2	1	16.1	16.7 *	V	A
346 *	506	3	1	19.9	20.9 *	VII	A
307 *	470	4	1	19.3	19.6 *	VIII	A
727 *	520	8	5	14.5	15.1 *	X	A
651 *	430	1	8	13.7	13.7 *	XI	A
542 *	429	1	5	17	16.2 *	XII	A

622 *	360	1	1	17.1	16.2 *	I	B
654 *	370	1	1	17.2	15.7 *	II	B
451 *	450	2	1	19.5	20.3 *	III	B
418 *	500	2	1	20.7	19.3 *	IV	B
226 *	520	4	2	17.1	16.4 *	IX	B
507 *	448	2	2	18.4	17.7 *	V	B
904 *	516	5	1	21.6	20.1 *	VII	B
450 *	520	3	2	17.5	18.4 *	VIII	B
210 *	477	4	5	13.5	13.3 *	X	B
543 *	463	1	5	14.7	12.5 *	XI	B
445 *	460	2	6	16.5	15.2 *	XII	B

614 *	475	1	1	16	16.8 *	I	C
617 *	439	1	1	16.8	16.1 *	II	C
553 *	440	2	2	20	19.9 *	III	C
427 *	498	2	1	19.6	19.7 *	IV	C
908 *	488	6	2	14.4	16.9 *	IX	C
505 *	434	2	2	19.4	17.7 *	V	C
941 *	466	7	2	20.6	20.3 *	VII	C
329 *	556	3	2	19.3	19.8 *	VIII	C
833 *	540	7	7	14.8	14.5 *	X	C
409 *	508	2	7	13.2	13.1 *	XI	C
429 *	467	2	8	14.1	16.2 *	XII	C

En la tabla se agregó la producción en otro período y el peso vivo para tener mas elementos de juicio en la decisión de si los bloques estaban bien hechos.

4.2.DISEÑO EN CUADRADO LATINO

4.2.1.Concepto y usos del Cuadrado Latino.

Se presenta a veces la situación de que existen dos razones diferentes para formar bloques, teniéndose bloques en dos sentidos o factores.

Ejemplo 4.11. Supóngase en el caso del Ejemplo 4.1 que además de la pendiente E-W, existiera otra N-S.

En situaciones como ésta se usa el diseño en Cuadrado Latino que contempla la formación de bloques en dos sentidos, se les suele llamar bloques fila y bloques columna o filas y columnas directamente. Los tratamientos deben ubicarse de tal modo que se presenten una sola vez en cada fila y cada columna. Esta condición lleva a que los Cuadrados Latinos tengan siempre tantas filas como columnas como tratamientos. La construcción del diseño asegura de este modo que los efectos de los tratamientos pueden estimarse independientemente de los efectos de las filas y las columnas. Daremos otras aplicaciones del diseño a través de ejemplos.

Ejemplo 4.12. Cochran y Cox (1978) destacan la utilidad del diseño en investigación preliminar de fuentes de variación, y presentan ejemplo de como en condiciones comunes y corrientes se pueden evaluar efectos desconocidos: Se considera que la música afectaba a la producción en una fábrica, y se analizó el efecto simultáneo de cuatro programas de radio y un testigo sin radio y el día de la semana, en una experiencia que duró cinco semanas realizada según el siguiente diseño:

SEMANA	Lunes	Martes	Miércoles	Jueves	Viernes
1	1	5	4	3	2
2	2	1	5	4	3
3	3	2	1	5	4
4	4	3	2	1	5
5	5	4	3	2	1

Ejemplo 4.13. Otra utilización muy importante del diseño en Cuadrado Latino es en experiencias con animales, a menudo con vacas lecheras. Supóngase que se desea analizar el efecto de tres sistemas de alimentación en vacas lecheras. La adjudicación al azar de los tratamientos a las vacas no dará buen resultado si la variación entre vacas es importante. Generalmente en estos casos no se detectan diferencias significativas entre tratamientos. Se puede considerar a cada vaca como un bloque, y alimentarla sucesivamente con los tres métodos. En ese caso se presenta el problema de que luego de un pico la vaca disminuye su producción a lo largo de la lactancia. Si a un tratamiento le tocan en varias vacas al final de la lactancia se verá perjudicado en la comparación. Con estos elementos, un diseño adecuado para probar tres tratamientos con tres vacas puede ser el siguiente:

ETAPA DE LA LACTACION			
VACA	1	2	3
1	A	B	C
2	C	A	B
3	B	C	A

donde las letras identifican a los tres tratamientos.

Ejemplo 4.14. Si en el caso del ejemplo 4.1 la pendiente fuera uniforme en un sentido solo, pero el campo experimental fuera de forma alargada y con parcelas perpendiculares al gradiente de fertilidad, se pueden tomar bloques de cuatro parcelas y dentro de ellos considerar como columnas

a la posición dentro del bloque. Cada tratamiento estará una vez en cada fila y en cada columna como se ve en el dibujo:

A	B	C	D	D	A	B	C	C	D	A	B	B	C	A	D
Bloques < 1 >				< 2 >				< 3 >				< 4 >			

Figura 4.2. Uso de Cuadrado Latino para controlar el gradiente en un solo sentido.

Ejemplo 4.10. (continuación) Notemos también que con la ordenación propuesta por Barretto y Montossi se puede pensar en un Cuadrado Latino del tipo presentado por Cochran y Cox (1957): se ordenan los animales en bloques (llamémosles bloques-fila) de 3, pero dentro de cada bloque-fila se ordenan de 1 a 3, se forman bloques-columna con las primeras de cada bloque-fila, otro con las segundas y un último con las terceras. Se adjudican los tratamientos de modo que le toque una vaca de cada bloque-fila y de cada bloque- columna en un Cuadrado Latino.

4.2.2. Modelo y análisis para Cuadrado Latino.

El modelo para el diseño en Cuadrado Latino es totalmente similar al de Bloques al Azar contemplando dos tipos de bloques: $Y_{ijh} = \mu + \beta_i + \kappa_h + \tau_j + \varepsilon_{ijh} \quad \varepsilon \sim \text{NID}(0; \sigma^2)$

es decir que la observación ijh es la suma de una media general, un efecto de la fila i , un efecto de la columna h , un efecto del tratamiento j y un error proveniente de una población de errores con media cero, varianza constante y distribución normal. Los estimadores dados por el método de los mínimos cuadrados como siempre son:

$$\begin{aligned}\hat{\mu} &= \bar{Y} \\ \hat{\beta}_i &= \bar{Y}_{i.} - \bar{Y}_{..} \\ \hat{\kappa}_k &= \bar{Y}_{..k} - \bar{Y}_{..} \\ \hat{\tau}_{.j.} &= \bar{Y}_{.j.} - \bar{Y}_{..} \\ e_{ijh} &= Y_{ijk} - \hat{Y}_{ijk}\end{aligned}$$

Y el análisis de varianza de acuerdo al modelo es: $SC = SCF + SCC + SCT + SCE$. Estos resultados se presentan en forma de cuadro donde se usan las fórmulas abreviadas de cálculo. Nótese que, a pesar de indicar a cada observación con tres subíndices, se utiliza solamente una doble sumatoria. En el cuadro puede verse además que filas, tratamientos y columnas tiene igual número de grados de libertad, debido a la característica del diseño de tener tantas filas y columnas como tratamientos.

FUENTE VARIACIÓN	SUMAS CUADRADOS	GRADOS LIBERTAD	CUADRADO MEDIO	F
Filas	$\bar{Y}_{..k} \cdot \bar{Y}_{..}$	n-1	SCB/GLB	CMT/CME
Columnas	$\sum T_i^2$			
Tratamientos	$\frac{\quad}{n} - C$	k-1	SCT/GLT	
Error o residuo	por diferencia	(n-1)(k-1)	SCE/GLE	
TOTAL		nk-1		

4.2.3. Elección del Cuadrado Latino.

Los cuadrados latinos no se aleatorizan como un bloques al azar debido a la estructura requerida. La técnica usual es elegir un cuadrado latino de una tabla como la que proporcionan Cochran y Cox (1978) y luego sortear cual es el tratamiento 1, cual es el 2, y así sucesivamente. Adicionalmente, se aconseja aleatorizar el orden de las columnas y de las filas, debido a que variando el orden de filas o columnas el cuadrado no pierde su condición de latino. Los cuadrados vienen en las tablas en formas de cuadrados estandarizados, es decir cuadrados latinos en los que la primera columna y la primera fila están ordenadas A,B,C, etc.

Ciertos autores alertan sobre la presencia de situaciones sistemáticas como la siguiente:

A	B	C	D	E	
D	E	A	B	C	
B	C	D	E	A	<-- Norte
E	A	B	C	D	
C	D	E	A	B	

Nótese que si la variedad E fuera de alto porte puede estar siempre sombreando a la D, excepto en la fila 4. El problema es darse cuenta, y la solución es estar alerta contra este tipo de situaciones.

Ejemplo 4.14. En el siguiente cuadro se muestra un ejemplo de Cuadrado Latino.

N 43	NP 61	(1) 35	P 21
NP 51	P 32	N 28	(1) 37
(1) 23	N 7	P 1	NP 60
P 19	(1) 24	NP 48	N 21

El análisis de varianza obtenido es el siguiente:

FUENTE VARIACIÓN	SUMAS	CUADRADOS	GRADOS LIBERTAD	CUADRADO MEDIO	F
Columnas	120		3	40,00	1,04
Filas	1512		3	504,00	13,09
Tratamientos	1853		3	617,67	16,04
Error	231		6	38,50	
TOTAL	3716		15		

4.2.5. Uso del SAS.

Basado en la característica de SAS de trabajar con el modelo que se le proporciona, el análisis de un Cuadrado Latino es tan fácil como el de un bloques al azar. La única característica consiste en proporcionarle un modelo de tres vías, tratamientos bloques-fila y bloque-columnas:

```
DATA cuatro15;  
INPUT f c t $ rend;  
CARDS;  
1 1 N 43  
1 2 NP 61  
1 3 (1) 35  
1 4 P 21  
2 1 NP 51  
2 2 P 32  
2 3 N 28  
2 4 (1) 37  
3 1 (1) 23  
3 2 N 7  
3 3 P 1  
3 4 NP 29  
4 1 P 19  
4 2 (1) 24  
4 3 NP 48  
4 4 N 21
```

```
PROC ANOVA;  
CLASS f c t;  
MODEL rend=f c t;  
MEANS t;  
RUN;
```

```
PROC GLM;  
CLASS f c t;  
MODEL rend=f c t;  
MEANS t/LSD LINES;  
CONTRAST 'N' t -1 -1 +1 +1;  
CONTRAST 'T' t -1 +1 -1 +1;  
CONTRAST 'INTERACCION' t +1 -1 -1 +1;  
RUN;
```

El análisis se corre con el ANOVA y el GLM a los efectos de mostrar las diferencias y similitudes, la más importante de las cuales es la disponibilidad de contrastes en el GLM.

Analysis of Variance Procedure					
Class Level Information					
Class	Levels	Values			
F	4	1	2	3	4
C	4	1	2	3	4
T	4	(1)	N	NP	P
Number of observations in data set = 16					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	3485.0000000	387.2222222	10.06	0.0055
Error	6	231.0000000	38.5000000		
Corrected Total	15	3716.0000000			
R-Square		C.V.	Root MSE	REND Mean	
0.937836		20.682789	6.2048368	30.0000000	
Source	DF	Anova SS	Mean Square	F Value	Pr > F
F	3	1512.000000	504.000000	13.09	0.0048
C	3	120.000000	40.000000	1.04	0.4404
T	3	1853.000000	617.666667	16.04	0.0029
Level of	-----REND-----				
	T	N	Mean	SD	
	(1)	4	29.7500000	7.2743843	
	N	4	24.7500000	14.9749791	
	NP	4	47.2500000	13.3759735	
	P	4	18.2500000	12.8419884	

La salida del GLM se muestra a continuación:

General Linear Models Procedure					
Class Level Information					
Class	Levels	Values			
F	4	1	2	3	4
C	4	1	2	3	4
T	4	(1)	N	NP	P
Number of observations in data set = 16					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	3485.0000000	387.2222222	10.06	0.0055
Error	6	231.0000000	38.5000000		
Corrected Total	15	3716.0000000			
R-Square		C.V.	Root MSE	REND Mean	
0.937836		20.682789	6.2048368	30.0000000	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
F	3	1512.000000	504.000000	13.09	0.0048
C	3	120.000000	40.000000	1.04	0.4404
T	3	1853.000000	617.666667	16.04	0.0029
Source	DF	Type III SS	Mean Square	F Value	Pr > F
F	3	1512.000000	504.000000	13.09	0.0048
C	3	120.000000	40.000000	1.04	0.4404
T	3	1853.000000	617.666667	16.04	0.0029
T tests (LSD) for variable: REND					
NOTE: This test controls the type I comparisonwise error rate					
not the experimentwise error rate. Alpha= 0.05 df= 6 MSE= 38.5					
Critical Value of T= 2.45 Least Significant Difference= 10.736					
Means with the same letter are not significantly different.					
T Grouping		Mean	N	T	
A		47.250	4	NP	
B		29.750	4	(1)	
B					
C		24.750	4	N	
C					
C		18.250	4	P	
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
N	1	121.0000000	121.0000000	3.14	0.1266
T	1	1156.0000000	1156.0000000	30.03	0.0015
INTERACCION	1	576.0000000	576.0000000	14.96	0.0083

Recordemos que, según lo visto en la sección 3.3, la comparación de todos contra todos no es lo indicado en un experimento factorial.

4.2.5. Cadenas de Cuadrados Latinos.

Dada la rigidez que posee el diseño en Cuadrado Latino se utiliza en general solamente para experimentos que involucren 4 a 8 tratamientos. Se trata de evitar que los tratamientos sean pocos o demasiados. Si los tratamientos son muy pocos los grados de libertad del error serán insuficientes, y si los tratamientos son muchos, resulta difícil de trabajar con el diseño. Si se tienen tres tratamientos (supongamos tres períodos en cada lactancia en vacas lecheras) se puede superar la situación de falta de grados de libertad repitiendo el experimento entero, con lo que se forman las llamadas cadenas de Cuadrados Latinos:

	Cuadrado 1			Cuadrado 2		
Vaca	1	2	3	1	2	3
1	A	B	C	A	B	C
2	B	C	A	C	A	B
3	C	A	B	B	C	A

El análisis de varianzas correspondientes será:

Fuentes de Variación	G. Libertad	G. Libertad
Entre cuadrados	1	2
Vacas dentro de cuadrados	4	6
Tratamientos	2	2
Filas	2	2
Error	8	14
TOTAL	17	26

Los grados de libertad entre vacas que son 5, se han descompuesto en grados de libertad entre cuadrados (1) más grados de libertad de las columnas (vacas) dentro de cuadrados que quedan en 4. También se podría separar el efecto de “filas-dentro-de-cuadrados” con dos grados de libertad, de la manera que lo hicimos arriba la interacción entre filas y cuadrados está en el error. Se pueden hacer tantos cuadrados como queramos (o podamos). La última columna del cuadro muestra la situación si hiciéramos una cadena de 3 cuadrados latinos. Si siguiéramos la regla de tener al menos 10 o 12 grados de libertad necesitaríamos una cadena de 3 cuadrados latinos como mínimo para lograrlo.

El modelo matemático para lograr ese análisis de varianzas será:

$$Y = \mu + \text{Cadena} + \text{Vaca}(\text{Cadena}) + \text{Tratamiento} + \text{Período} + \varepsilon$$

Donde, como siempre, Y es la variable de interés, μ es la constante (o media) y ε es el error.

En una situación como esta el investigador se pregunta por que la vaca 2 tiene que estar al lado de la 1. Supongamos que las vacas fueron tomadas en grupos de 3 animales semejantes. Si eso no sucede conviene aleatorizar el orden de las vacas y tenemos un diseño en retrocruzamiento (*cross over* o *change over*):

Vaca	1	2	3	1	2	3
1	A	B	C	A	B	C
2	B	C	A	C	A	B
3	C	A	B	B	C	A

4.2.6. Change Over y diseños relacionados.

Si tenemos dos tratamientos para ser aplicados a doce vacas, podemos intentar una cadena de 6 cuadrados latinos:

Cuadrado	1	2	3	4	5	6	7	8	9	10	11	12
Vaca	1	2	3	4	5	6	7	8	9	10	11	12
Período I	B	A	A	B	A	B	B	A	B	A	A	B
Período II	A	B	B	A	B	A	A	B	A	B	B	A

El diseño Change Over, o Cross Over como también se le dice, para un caso como el anterior consiste en adjudicar al azar los tratamientos a los animales, solamente exigiendo que a la mitad de ellos se le aplique el tratamiento A en primer lugar y a la otra mitad el B en primer lugar. No se exige en cambio, que si la vaca 1 le tocó la secuencia BA a la 2 le toque la secuencia AB, como lo hace la cadena de Cuadrados Latinos que vimos arriba. Es decir que, mediante sorteo, se puede lograr un diseño como éste:

Vaca	1	2	3	4	5	6	7	8	9	10	11	12
Período I	B	B	A	B	A	A	A	A	B	B	B	A
Período II	A	A	B	A	B	B	B	B	A	A	A	B

Reiteramos la idea anterior de que la utilización del diseño en change over debe hacerse en los casos en que efectivamente no haya diferencia esperada entre los cuadrados, es decir siempre que se considere que los cuadrados no son fuente de variación.

El análisis de un diseño change over como el mencionado se presenta abajo comparado al de una serie de seis Cuadrados Latinos:

SERIES DE CUADRADOS		CHANGE OVER
Fuentes de Variación	G. Libertad	Fuentes de Variación
Entre cuadrados	5	
Columnas (dentro de cuadrados)	6 11	Entre columnas
Tratamientos	1 1	Tratamientos
Filas (dentro de cuadrados)	6 1	Filas (Total)
Error	5 10	Error
TOTAL	23	TOTAL

El análisis es pues muy similar en ambos casos, pero si se cumple la situación de que la única diferencia entre cuadrados está dada por el azar, los grados de libertad entre cuadrados se incorporan al error ventajosamente. El análisis de éste tipo de diseño puede hacerse por la t de Student, como lo revela el hecho de que la F tiene un solo grado de libertad en el numerador.

El maestro C. C. Li (1969) explica inigualablemente los detalles del análisis, Federer (1973) describe los diseños *change over* en la página 122 y de los secuenciales en la página 127.

4.2.7. Cuadrado Greco-Latino y otros cuadrados.

Otro diseño similar a los que venimos estudiando es el diseño en Cuadrado Grecolatino.

Ejemplo 4.16. Supongamos que se dispone de 16 cerdos para probar cuatro raciones alimenticias. Los cerdos provienen de cuatro camadas distintas, y se dispone de cuatro establos diferentes para ubicarlos durante el ensayo. Como se piensa que ambos factores inciden en las ganancias de peso se proyecta utilizar el diseño en Cuadrado Latino, pero se observa que el peso de los animales es variable dentro de cada camada, aún cuando parecidos para los cerdos de igual posición en las distintas camadas. Es decir, parece aconsejable tomar el orden de peso del animal como otro factor de variación, formando un bloque con los más pesados en cada camada, otro con los que siguen, etc. El diseño resultante, que contempla bloques en tres sentidos, es un cuadrado grecolatino:

Camada	Establo 1	2	3	4
1	A α	B γ	C δ	D β
2	B β	A δ	D γ	C α
3	C γ	D α	A β	B δ
4	D δ	C β	B α	A γ

El análisis de varianza será:

Fuentes de Variación	G. Libertad
Entre camadas	3
Entre establos	3
Tratamientos latinos	3
Tratamientos griegos	3
Error	3
TOTAL	15

El modelo matemático será:

$$Y = \mu + \text{Camada} + \text{Establo} + \text{Tratamiento latino} + \text{Tratamiento griego} + \varepsilon$$

Diseños Row-and-Column. Existen otros diseños que utilizan mas de dos alfabetos y que reciben nombres de cuadrados hiper-greco-latinos, cuadrados mágicos, etc. Federer (1973) describe los diseños *Rectángulo Latino* en la página 123, *Cubo Latino* en la página 127 y *Cuadrado Mágico* en la página 125. Los ingleses hablan de diseños *Row-and-column* como una categoría que abarca todas estas, veremos algo de eso luego de introducir las ideas de los bloques incompletos.

4.3. BLOQUES INCOMPLETOS

4.3.1. Introducción a los bloques incompletos

Los diseños en bloques al azar y cuadrados latinos gozan de las ventajas de ortogonalidad y simplicidad de análisis. Es muy común, en la comparación de variedades y otros problemas en los que se comparan muchos tratamientos. Al aumentar el número de tratamientos aumenta el tamaño necesario del bloque y por lo tanto su heterogeneidad. Este tipo de situaciones llevó a la necesidad de introducir bloques incompletos. Los bloques incompletos se recomiendan a partir de las 15 variedades. Como estos diseños se usan generalmente en la comparación de variedades vamos a simbolizar el número de tratamientos por V.

CLASES DE DISEÑOS EN BLOQUES INCOMPLETOS. Cochran & Cox (1952) y Pimentel Gomes (1990) clasifican a los diseños en bloques incompletos en:

Tipo I. Se pueden agrupar en repeticiones. Si los bloques se pueden agrupar en repeticiones el diseño se dice "resolvable", en caso contrario "no-resolvable".

Tipo II. Se pueden agrupar en grupos de repeticiones.

Tipo III: No se pueden agrupar en repeticiones.

Supongamos que las variedades a comparar son 6:

Bloque					
I	A	B	C		
II	D	E	F		
III	A	E	F		
IV	B	C	D		

En el segundo caso los bloques I y II forman una repetición del experimento, los bloques III y IV forman una segunda.

Ejemplo 4.18. Los datos siguientes son de 4 tratamientos en 4 bloques de tamaño 3. Debido al tamaño de los bloques no caben todos los tratamientos. La idea básica es dejar afuera de cada bloque un tratamiento diferente, el tratamiento 1 quedó afuera del bloque 1, el 2 en el bloque 2, etc.

	Trat 1	Trat 2	Trat 3	Trat 4
Bloque 1		xx	xx	xx
2	xx		xx	xx
3	xx	xx		xx
4	xx	xx	xx	

Luego se aleatorizan el orden de los bloques y la situación queda así:

	Trat 1	Trat 2	Trat 3	Trat 4	Total
Bloque 1	74		74	76	224
2	75	76	76		227
3		68	69	73	208
4	72	73		76	221

Este caso es balanceado, es decir todas las comparaciones entre tratamientos tienen igual precisión.

ALGUNOS CONCEPTOS: BALANCE. Si se mide el número de veces que cada par de tratamientos se presenta en un mismo bloque (llamémosle λ) y el ese valor es igual para todos los pares de tratamientos el diseño se dice balanceado (el diseño se llama BIB – bloque incompleto balanceado). Si eso no es así el diseño es desbalanceado. Hay diseños parcialmente balanceados (PBIB) cuando algunos tratamientos se presentan con otros λ_1 veces y con otros mas λ_2 veces. Supongamos que tenemos 7 variedades a comparar en bloques de tamaño 3:

Bloques

I	A B C
II	D E F
III	G A B
IV	C D E
V	F G A
VI	B C D
VII	E F G

Hacer los bloques no es fácil. El diseño en bloques completos al azar prevee que todos los tratamientos están en todos los bloques. Muchas veces formamos bloques de tamaño menor que el número de tratamientos que tenemos. Entonces nos encontramos con que tenemos que usar bloques incompletos, es decir que no todos los tratamientos están en todos los bloques.

Tipo 1. Si tenemos 6 variedades en bloques de tamaño 3

Tipo 3. Si tenemos 7 variedades en bloques de tamaño 3

Tipo 1	Tipo 2	Tipo 3
A B C		A B C
D E F		D E F
A E F		G A B
B C D		C D E
		F G A
		B C D
		E F G

Notemos que el factor clave pasa a ser el tamaño de los bloques, el investigador forma bloques del tamaño que el material experimental le permite.

Ejemplo 4.19.

↓Trat/B→	1	2	3	4	5	6	7	8
1		52		43		114		34
2	10	78				80	130	
3		103		187	298			263
4	215	164			257		227	
5			27	122	120			153
6			26	78		56		89
7	95		55			65	139	
8	154		150		265		254	

4.3.2 Análisis

Antes. Antiguamente se hablaba así:

Análisis como bloques completos. Si el diseño es resolvable (como en el caso de los látices) se puede analizar como bloques completos. Esto no es recomendable, ya que estamos perdiendo información. Además el experimentador generalmente forma (como debe ser) los bloques como parcelas de igual respuesta y por tanto las repeticiones no tienen efecto. La única razón de este procedimiento es convencer al investigador de usar bloques incompletos: si el diseño es resolvable entonces el investigador no arriesga nada con usarlo.

Análisis intrabloques. El modelo para los bloques al azar es: $Y_{ij} = \mu + b_i + t_j + e_{ij}$ sean los bloques completos o incompletos. Si tenemos repeticiones podemos escribir:

$$Y_{ijk} = \mu + b_{i(k)} + R_k + t_j + e_{ijk}.$$

El caso más frecuente es que los efectos de los tratamientos sean fijos. El análisis intra-bloques asume que los efectos de los bloques son también fijosⁱ. La estimación de los efectos de los tratamientos procura entonces obtener el BLUE que se obtiene con el PROC GLM de SAS. Las medias ajustadas de los libros de estadística experimental son las "LS means" de SAS. Las varianzas de los contrastes son complicadas. Para la prueba t o MDS el número diferente de repeticiones, hace que el SAS calcule un p-value para cada contraste lo cual hace muy complicada la interpretación. Con estas salvedades y el tener cuidado con la descomposición de la Suma de Cuadrados de los Tratamientos, permiten asegurar que el análisis con GLM puede ser hecho bien. La eficiencia de los BI puede ser interesante de estudiar.

Análisis con recuperación de la información inter-bloques. Antiguamente se hablaba mucho de este procedimiento, mientras que hoy se piensa mejor en términos de proponer un modelo con efectos de los bloques aleatorios. La técnica tradicional consistía en obtener estimadores inter-bloques y combinarlos con los estimadores intra-bloques en una estimación más eficiente.

Ahora.

La técnica actual es usar el Proc Mixed. Recordemos que el GLM usa mínimos cuadrados y el Mixed usa máxima verosimilitud.

ⁱ Muchas personas, entre las que me incluyo, encuentran que eso es muy raro de ocurrir.

4.3.3.Lattices.

Para ilustrar la construcción de un Lattice partamos de la siguiente estructura cuadrada:

A	B	C
D	E	F
G	H	I

las filas del cuadrado generan los bloques de la repetición 1.

1	A	B	C
2	D	E	F
3	G	H	I

Las columnas generan bloques de la repetición 2 y las diagonales de la repetición 3.

4	A	D	G
5	B	E	H
6	C	F	I

7	A	D	I
8	B	E	H
9	C	F	G

Falta de balance de Lattice Simple o Triple. Cada par de variedades no aparecen juntas en un mismo bloque el mismo número de veces. En el Lattice Triple 3 x 3 que estamos analizando se observa la siguiente situación:

Par	Concurrencia	Par	Concurrencia
A,B	1	B,C	1
A,C	1	B,D	0
A,D	1	B,E	1
A,E	1	B,F	1
A,F	0	B,G	1
A,G	1	B,H	1
A,H	0	B,I	0
A,I	1		

Estos diseños entran en la categoría de no-ortogonales.

Ejemplo (pg213). 16 variedades en lattice 4x4 con 3 repeticiones.

Repetición1	Bloque1	1 2,0 2 2,9 3 2,2 4 3,9
	2	5 2,3 6 2,5 7 1,4 8 1,7
	3	9 1,6 10 3,0 11 1,5 12 2,1
	4	13 2,3 14 3,4 15 2,0 16 2,8
Repetición2	5	1 2,2 5 2,3 9 2,7 13 1,4
	6	2 3,1 6 2,8 10 2,6 14 2,8
	7	3 3,1 7 2,9 11 2,5 15 2,4
	8	4 4,0 8 2,8 12 2,7 16 1,5

Repetición3

9	1	3,0	6	2,9	11	2,6	16
		3,1					
10	2	1,8	5	1,9	12	2,9	15
		2,5					
11	3	1,7	8	2,0	9	1,4	14
		2,3					
12	4	4,4	7	3,7	10	3,3	13
		2,2					

```
proc mixed; class r b trt; model y=trt; random r b(r); ls means trt / pdiff; run;
```

4.3.4. Diseños alpha.

En las décadas del 30 y 40 era importante la facilidad de construcción y análisis de los diseños. La falta de flexibilidad para disponer de diseños en bloques incompletos resolvables para cualquier número de variedades era una limitante sensible en el trabajo. Patterson & Williams en 1976, introducen los α -Lattices o diseños Alpha, que son una generalización de los Lattices originales de Yates y que proporcionan una gran flexibilidad

$$V=24; k=4; s=6 \quad _V=26; k_1=5; k_2=4; s=6$$

Bloque	Rep			
1				
2				
3				
4				
5				
6				

Bloque	Rep			
1				
2				
3				
4				
5				
6				

La flexibilidad se obtuvo a un precio: la calidad de los α -Lattices es despareja.

Cómo elegir el tamaño de un diseño Alpha. Llamamos $\alpha(0,1)$ al diseño en el cual algunos de los tratamientos están juntos una vez y otros cero veces. Del mismo modo, llamamos $\alpha(0,1,2)$ al diseño en el cual algunos tratamientos están juntos 0 veces, otros 1 vez y otros 2 veces. Resulta claro que un diseño $\alpha(0, 1)$ es mejor que uno $(0, 1, 2)$ y debe ser preferido siempre que es posible.

Condición necesaria para la existencia de un $\alpha(0, 1)$: $k \leq s$. González (1991) da algunas reglas:

1. Si el número de tratamientos es mayor de 15, usar bloques incompletos, ya sean Lattices de Yates o α -Lattices.
2. Es conveniente que el número de repeticiones sea mayor a 3: $r \geq 3$
3. Tamaño del bloque igual o ligeramente menor que la raíz del numero de trat: $k \leq \sqrt{v}$
4. Una vez decididos v, r, k , se puede elegir el diseño usando Alphagen

Cómo construir un α -Lattice con tamaños de bloque diferentes, usando Alphagen.

Ejemplo: $v = 26$ $5 \times 5 < 26 < 5 \times 6$

Bloque	Rep			
1				
2				
3				
4				
5				
6				

4.3.7. Eficiencia de los diseños en bloques.

El factor de eficiencia (EF) es el resultado de dividir: (a) la variancia promedio con que un diseño en bloques completos estimaría las diferencias entre variedades, por (b) la variancia promedio con que el diseño en bloques incompletos estimaría esas diferencias. Si el cuadrado medio del error fuera el mismo en ambos casos el factor de eficiencia estará entre cero y uno: $0 < EF < 1$

Eficiencia relativa a) la variancia promedio de las diferencias entre variedades, en el análisis que ignora los bloques incompletos, y toma las replicaciones como bloques completos, dividida por b) la variancia promedio de dichas diferencias en el análisis de bloques incompletos (incluyendo la recuperación de información inter-bloque).

Cálculo de la EFICIENCIA RELATIVA del Lattice Simple 5x5 Ejemplo de Cochran & Cox

Delineamento	SED Promedio
Bloques incompletos	4,15
Bloques completos	5,48

Eficiencia relativa $= (5,48/4,15)^2 = 1,74$. Hubo un 74% de ganancia en eficiencia.

Ganancia en eficiencia. Normalmente la ganancia en eficiencia por el uso de los diseños en bloques incompletos es sensible. Patterson & Hunter (1983) reportan que, en 244 ensayos de cereales en el Reino Unido entre 1975 y 1982, la mediana de la eficiencia relativa fue de 1.23. En el programa de maíz de CIMMYT el uso de los α -Lattices desde 1985, en ensayos con más de 200 entradas ha mostrado una eficiencia relativa de entre el 100% y el 200%. El programa de trigo de CIMMYT, con el uso de los α -Lattices desde 1988, en ensayos de 100 a 400 entradas con riego ha logrado eficiencias relativas de entre 100% y 132%.

Cómo calcular el factor de eficiencia (HMEF) de cualquier diseño en bloques incompletos resolvable usando REML

1. Modelo con todos los factores fijos (para que no haya recuperación de inf. inter-bloque).
'fixed' Rep+Rep.Block+Varied.
2. Datos de rendimiento ficticios, sin ninguno repetido.
3. $r' = 2\alpha^2 / (SED)^2$
HMEF = r'/r

Cálculo de la Eficiencia Relativa (HMEF) de un diseño alpha para comparar $v = 26$ variedades. El diseño tiene $r = 3$ replicaciones, cada una de las cuales contiene dos bloques de 5 parcelas y cuatro bloques de 4 parcelas.

$v = 26$; $r = 3$; $k_1 = 5$; $k_2 = 4$; $s = 6$

$$HMEF = r'/r$$

$$r' = 2\alpha^2 / (SED)^2$$

$$r = 3$$

$$r' = 2 \times 10^{-7} / (0.000298)^2 = 2.252$$

$$HMEF = 2.252/3 = 0.75$$

Auditoría retrospectiva del uso de bloques incompletos Cómo predecir el valor de diseños en bloques incompletos mediante la técnica conocida como post-blocking (Ainsley [1985], Ainsley, Paterson & Patterson [1987]).

Cualquier experimento en BLOQUES COMPLETOS puede, es decir que es válido, ser analizado superponiéndole BLOQUES INCOMPLETOS sobre la disposición resultante de la aleatorización de los bloques completos, y haciendo como que los bloques incompletos fueron planificados desde el inicio.

Ejemplo 4.17. Rendimientos por parcela (t/h) en un ensayo en bloques al azar con 16 variedades

M	L	P	O	K	B	E	I
1.11	1.18	1.06	1.08	0.98	0.95	1.04	1.00
F	J	D	A	C	H	G	N
1.07	1.10	1.02	0.98	1.06	1.07	1.12	1.08
A	C	G	O	H	P	L	M
0.96	1.05	1.13	1.06	1.02	1.00	1.09	1.02
I	E	J	K	D	F	N	B
0.96	1.04	1.06	1.07	1.07	1.01	1.07	1.10
O	H	M	F	E	C	P	A
0.95	0.93	1.01	0.98	1.04	1.00	1.04	0.97
I	N	B	D	K	G	J	L
0.92	1.00	1.00	1.05	1.06	1.14	1.14	1.18

Ensayo de 16 variedades en tres bloques al azar. Estimación (conservativa) de la ganancia en eficiencia que hubiéramos obtenido si hubiéramos usado bloques incompletos: Eficiencia de bloques incompleto relativa a bloques completos: $= (0,0366/0,0314)^2 = 1,36$. Esto quiere decir que, para obtener, usando bloques completos, la misma precisión que hubiéramos obtenido usando bloques incompletos, hubiéramos precisado un 36% más de repeticiones $3 \times 1,36 = 4,08$. Es decir que hubiéramos precisado 4 bloques en lugar de 3.

Ensayo de cebada - La Estanzuela 1990. 24 variedades en bloques al azar con 4 repeticiones. Parcelas de 1.20 x 5.00m.

Repetición				
Bloque 1 5 parcelas	Bloque 2 5 parcelas	Bloque 3 5 parcelas	Bloque 4 5 parcelas	Bloque 5 4 parcelas

Estimación (conservadora) de la ganancia en eficiencia que hubiéramos obtenido de haber usado bloques incompletos en lugar de bloques completos: Eficiencia relativa $= (0,344/0,317)^2 = 1,18$. Para obtener, usando bloques completos, la misma precisión que hubiéramos obtenido usando bloques incompletos, en lugar de 4 repeticiones (bloques completos), hubiéramos precisado $4 \times 1,18 = 4,72$, es decir, 5 repeticiones.

4.3.9. Otros Diseños en Bloques Incompletos.

Yates (1940) introduce los *Lattice Squares*, que son diseños en bloques incompletos que usan dos sistemas de bloqueo para eliminar la heterogeneidad en dos direcciones. En estos Lattices, también el número v debe ser un cuadrado perfecto, y requieren un número alto de replicaciones. Los "*Lattice squares*" de Yates son diseños en *Filas-y-Columnas*, en bloques incompletos.

Patterson & Robinson (1989) presentan un catálogo de diseños en filas y columnas con $25 \leq v \leq 100$ y $r = 2$. Luego Patterson desarrolló un programa (RCX) que genera diseños eficientes en filas y columnas, para $16 \leq v \leq 400$ y $r = 2$.

TABLA 3. Clasificación de los principales diseños

One grouping		Two groupings
Bloques completos	Bloques aleatorizados	Cuadrados latinos
Resolvable	Square Lattice	Lattice square
Bloques incompletos	Rectang. Lattice	Triple Simple
	α -Lattice	Row-and-colum designs

Lattices aumentados.

Pimentel Gomes & Viegas (1978)ⁱⁱ recomendaban cosas como esta: partiendo de un diseño en látice

1	2	3	A
4	5	6	A
7	8	9	A
1	4	7	A
2	5	8	A
3	6	9	A

La idea que ellos manejan es que este diseño disminuye la varianza entre las medias ajustadas de los tratamientos.

Oliveira & Barbin (1990)ⁱⁱⁱ se entusiasman y agregan dos tratamientos comunes en cada bloque. Evidentemente estamos en un caso de lattices grandes, donde se comparan muchos tratamientos.

Finalmente podemos considerar que se pueden poner diferentes tratamientos en un bloque con la única condición de que haya tratamientos comunes en todos los bloques. Esa situación se parece a un grupo de experimentos en los cuales hay tratamientos comunes testigos. La veremos al final del libro.

ⁱⁱ Pimentel Gomes, F. & Viegas, G. P. Experiments in square lattice with a common treatment in all blocks. Revista Agricultura, Piracicaba, 53:35-43, 1978.

ⁱⁱⁱ Oliveira, A. C. & Barbin, D. Experimentos en reticulados cuadrados. Pesquisa Agropecuaria Brasileira, Brasília, 23(7): 717-723

Historia. Yates en 1936 introduce los Lattices cuadrados donde el tamaño del bloque es igual a la raíz cuadrado del número de tratamientos. Si el número de repeticiones es 2 se habla de Lattices simples, si es 3 de Lattices triples. Harshbarger en 1949 introduce los Lattices rectangulares, caracterizados porque el número de tratamientos es de la forma: $V = k(k+1)$. Los Lattices de Yates y Harshbarger son óptimos en el sentido de que son los mejores diseños disponibles para esas condiciones.

TABLA 4. Evolución de los diseños aleatorizados.

1910	Se introduce la <i>replicación</i> en la investigación agrícola
1919	Fisher ingresa a Rothamsted
1920's	Fisher introduce la <i>aleatorización</i> los diseños en <i>bloques</i> aleatorizados y el diseño en <i>Cuadrado Latino</i>
1936	Yates introduce los diseños en <i>Square Lattice</i>
1940	Yates introduce los <i>Lattice Square</i>
1949	Harshbarger introduce los diseños en <i>Lattice Rectangular</i>
1976	Patterson & Williams introducen los α - <i>designs</i>
1978	Patterson, Williams & Hunter construyen un catalogo de diseños <i>alpha</i> altamente eficientes.
1984	Patterson & Paterson desarrollan un algoritmo disponible para mainframes y minis, el cual automáticamente genera diseños α con alto nivel de factor de eficiencia
1987	El algoritmo se hace disponible para IBM PC's y compatibles
1988	Patterson & Robinson construyen un catalogo de diseños <i>row-and-column</i> muy eficientes.

Tesis y teorías del presente trabajo.

1. Lo importantes es el modelo (y no el nombre del diseño).

- Diseños que no tienen nombre (como es el caso del libro de Anderson, 1995, o la investigación de SCHATCH, 1996) tienen modelo.
- El SAS (y el GENSTAT) requieren el modelo, y la tendencia a trabajar así es creciente (el M-STAT no es así).
- Dos lattices juntos ya no son un lattice.
- Están ganando importancia si el modelo es mixto o no.

2. En la recuperación de la información inter-bloques lo importantes es el método de estimación de los σ_b^2 .

- El procedimiento de estimación de los BLUE's es unanimidad (GLS=MM), la estimación de las σ_b^2 no lo es.
- Los σ_b^2 pueden ser considerados componentes de varianza, eso forma parte de una visión de la realidad que no es única.

3. Modelo fijo mixto, esa es la cuestión.

- La duda de la visión clásica de recuperar o no la información inter-bloques no es tan importante. Si los bloques son aleatorios lo mejor (el BLUE) es la conjunta y no la intra-bloques. Si los σ_b^2 son determinados nulos, los bloques pueden ser ignorados.
- Ahora podemos trabajar así: si el modelo es fijo use mínimos cuadrados, si el modelo es mixto use MM, y el resultado concuerda con Yates.

4. El nuevo software hacen realizables diseños que antes no lo eran.

- El diseño de Oliveira & Barbin (1989) es un ejemplo de caso que puede ser analizado con el Proc Mixed.
- Los casos particulares pueden ser más fáciles de deducir a partir de las MM.
- El nuevo límite de análisis parece ser la conexión. Datos desconexos no pueden ser analizados conjuntamente. (Para los diseños el límite es la eficiencia.)

5. De los bloques incompletos a los grupos de experimentos es un paso. La nueva mentalidad es: estimadores de dos diferentes estratos de información que se ponderan adecuadamente.

Comparación entre los métodos clásico y moderno^{iv}.

Método clásico (Yates, 1940; Pimentel Gomes, 1991)	Método de los modelos mixtos Henderson (1984), SAS (1992)
1. Existen dos formas de análisis: intra-bloques y con recuperación de la información inter-bloques.	1. Los bloques pueden ser considerados fijos o aleatorios. Por lo tanto, el modelo es fijo o mixto. El procedimiento que debe ser utilizado es consecuencia de esto.
2. Si los bloques son fijos el análisis adecuado es intra-bloques	2. Si los bloques son fijos, el análisis adecuado es con el GLM.
3. Si los bloques son aleatorios tienen información sobre los tratamientos.	3. Si los bloques son aleatorios el modelo es mixto, entonces las ecuaciones MM proveen el BLUE (la mejor solución, la otra es inaceptable)
4. El parámetro importante es $a = \frac{\sigma_e^2}{\sigma_e^2 + k\sigma_b^2}$.	4. El punto clave es estimar σ_b^2
5. El método de estimación de varianzas más utilizado puede ser considerado método 3 de Henderson.	5. Otros métodos de estimación de varianzas como ML (máxima verosimilitud) y REML pueden ser mejores.
6. Las medias de los tratamientos deben ser corregidas.	6. Los BLUEs de los tratamientos en los modelos mixtos, corresponden con las medias corregidas tradicionales.
7. Los resultados son aproximados	7. El grado de aproximación de los resultados es bastante conocido. Existen resultados exactos (Ofversten, 1993) pero son complejos.

^{iv} Asumimos que los tratamientos también son fijos

4.4. ANALISIS DE COVARIANZA

4.4.1. Idea básica del análisis de covarianza.

Por análisis de covarianza se entiende generalmente a la técnica para corregir los resultados de una variable de interés por la influencia de otra, llamada la covariable. Supongamos que se compara el peso de dos razas de animales, puede suceder que los animales tengan diferentes edades e interesa comparar los pesos a igualdad de edades. Es decir que se debe estimar el peso que tendrán los animales a una determinada edad y efectuar la comparación de esos pesos. Puede haber, por lo tanto, una discusión acerca de a que edad se llevan todos los pesos. Una solución a veces tomada es llevar a todos los animales "a peso promedio". Esa solución a su vez esta condicionada a la media de ese grupo de animales. Otra postura es llevar a una edad arbitraria, por ejemplo 180 días (6 meses). Esta última es la actitud tomada generalmente en mejoramiento animal, pero la primera tiene cierto preferencia por tener mejor precisión en un ensayo determinado.

La idea básica del ajuste de valores se puede ver en la figura que sigue:

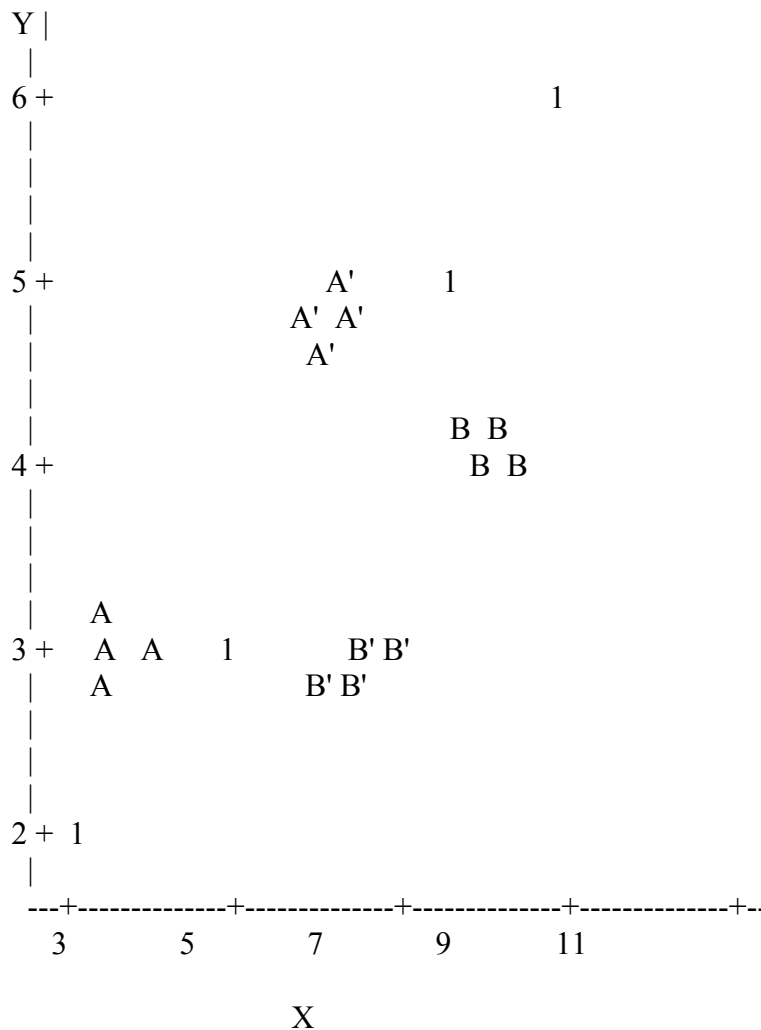


Figura 4.3. Corrección por covariables.

Los animales de la raza A son más livianos que los de la raza B, pero observando las edades (X) se puede determinar que son mas jóvenes. Los valores reales son los marcados con las letras y los valores corregidos por peso son los indicados con primas (apostrofes). En cierto modo los valores corregidos son los que se hubieran obtenido si hubiesen tenido la edad promedio (o una edad dada de antemano). En este caso, cuando se corrige por el efecto de la edad se observa que en realidad los animales de la raza A son más pesados (para su edad digamos) que los de la raza B.

Un ejemplo muy común en producción vegetal, puede ser cuando el rendimiento de un cultivo para dos variedades deba compararse, pero se lograron diferente número de plantas en cada una de las dos variedades y se intenta una comparación que no esté afectada por el número de plantas.

Del punto de vista de los modelos lineales el análisis de covarianza no es mas que una aplicación de la regresión. En realidad toda variable cuantitativa que pueda afectar a la variable de interés puede ser llamada covariable. Generalmente se usa la expresión análisis de covarianza para el uso de regresión especialmente destinado a corregir los valores por una covariable.

4.4.2. Enfoque clásico.

El enfoque clásico como lo muestran Snedecor (1988) y Steel y Torrie (1983) es analizar los resultados por medio de suma de cuadrados y un nuevo concepto similar que es la suma de productos. Se construye una cantidad similar a la suma de cuadrados que es la suma de productos y se trabaja con la línea del error. Por lo tanto el análisis tendrá diferencias según el diseño sea parcelas al azar, bloques al azar, cuadrado latino o experimentos factoriales. El análisis asume que se esta trabajando con la regresión "del error" es decir la respuesta a la regresión dentro de cada tratamiento.

El enfoque clásico del análisis de covarianza llevaba a largos y laboriosos cálculos. Por este motivo, no se hacia en general uso de regresiones curvilíneas o multivariadas.

4.4.3. Enfoque de modelos lineales.

Un enfoque tipo modelos lineales puede verse en Freund (1988) o Draper y Smith (1988). La idea puede resumirse en los siguientes términos. Para cada tratamiento se puede ajustar una recta de regresión peso contra edad (o rendimiento contra número de plantas):

$$Y = B_{0i} + B_{1i} X + E$$

Se puede ajustar una recta única peso contra edad (rendimiento por número de plantas en el ejemplo de producción vegetal):

$$Y = B_0 + B_1 X + E$$

Se puede ajustar una recta promedio de las rectas de cada tratamiento

$$Y = B_{0i} + B_{1i} X + E$$

Se puede ver que esa recta promedio es la que se obtiene de ajustar una recta de regresión con las sumas de cuadrados y de productos del error. Esta es la recta usada para el ajuste.

Aparte de la mayor información que normalmente se puede sacar del enfoque por modelos lineales se vuelve bastante fácil utilizar regresiones cuadráticas y dos o mas covariables, que el enfoque clásico no permite hacer con facilidad.

4.4.4. Ajuste de medias.

Según el enfoque clásico el tema del ajuste de medias se visualiza con mas facilidad, mientras que en el otro caso no es tan así. Un poco mal acostumbrados por el enfoque clásico muchos investigadores quieren tener las medias corregidas, perdiendo un poco de vista que el valor de esa corrección es relativo. Según el enfoque GLM se habla de LSmeans o medias mínimo-cuadráticas.

4.4.5. Eficiencia del análisis de covarianza.

Generalmente se considera que el análisis de covarianza se justifica si se obtiene una regresión significativa contra la covariable. También se puede observar la disminución del error que se obtiene con el uso de la covariable. En el caso de que el factor de bloque sea una variable cuantitativa se presenta la opción de usar bloques o covarianza. La filosofía de los bloques es considerar que todas las unidades experimentales son iguales, por lo tanto algunos autores argumentan que el análisis de covarianza describe mejor muchas situaciones. Otros dicen que se puede bloquear y, adicionalmente, usar covariable para remover la variabilidad remanente. También se presentan situaciones donde hay mas de un factor de bloqueo (como puede ser el caso de duda si bloquear animales por peso o por estado) y se opta por bloquear por un factor y usar el otro como covariable.

4.4.6. Análisis de covarianza en el SAS.

Un programa SAS para el análisis de covarianza sería:

```
DATA cuatro20;
TITLE 'Análisis de covarianza';
INPUT t x y @@;
CARDS;
1 5 3 1 9 5 1 11 6 1 3 2
2 4 5 2 7 6 2 0 2 2 8 7
2 10 8 2 2 3 2 12 9 2 5 8
3 8 7 3 7 6 3 3 4
3 2 3 3 8 6 3 2 4
PROC GLM;
  CLASS t;
  MODEL y=t x/SOLUTION;
  RUN;
```

Este programa ajusta efectos de tratamientos y regresión con X.

```
PROC GLM;
  CLASS t;
  MODEL y=t x t*x/SOLUTION;
  OUTPUT OUT=salida1 P=ygorro R=desvios;
  RUN;
```

Este otro ajusta además por una posible interacción entre tratamientos y la covariable, si esa interacción es significativa quiere decir que las regresiones no son iguales para los diferentes tratamientos.

Si las regresiones no son iguales se debe ajustar una regresión por tratamiento, del siguiente modo:

```
PROC SORT; b y t;
PROC PLOT; b y t;
  PLOT y*x=t;
  TITLE2 'Peso VS Edad por tratamiento';
  PLOT ygorro*x;
  RUN;
```

y la salida es:

ANALISIS DE COVARIANZA					
General Linear Models Procedure					
Class Level Information					
Class	Levels	Values			
T	3	1	2	3	
Number of observations in data set = 18					
Dependent Variable: Y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	67.8585859	22.6195286	34.23	0.0001
Error	14	9.2525253	0.6608947		
Corrected Total	17	77.1111111			
R-Square	C.V.	Root MSE	Y Mean		
	0.880010	15.56721	0.81295		
Source	DF	Type I SS	Mean Square	F Value	Pr > F
T	2	11.1111111	5.5555556	8.41	0.0040
X	1	56.7474747	56.7474747	85.86	0.0001
Source	DF	Type III SS	Mean Square	F Value	Pr > F
T	2	17.3485443	8.6742721	13.13	0.0006
X	1	56.7474747	56.7474747	85.86	0.0001
T for H0: Pr > T Std Error of					
Parameter		Estimate	Parameter=0		Estimate
INTERCEPT		2.323232323 B	5.28	0.0001	0.43999478
T	1	-2.070707071 B	-3.85	0.0018	0.53733061
	2	0.464646465 B	1.05	0.3118	0.44283044
	3	0.000000000 B	.	.	.
X		0.535353535 B	9.27	0.0001	0.05777414
NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.					

ANALISIS DE COVARIANZA					
General Linear Models Procedure					
Class Level Information					
Class	Levels	Values			
T	3	1	2	3	
Number of observations in data set = 18					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	68.0409357	13.6081871	18.00	0.0001
Error	12	9.0701754	0.7558480		
Corrected Total	17	77.1111111			
R-Square	C.V.	Root MSE	Y Mean		
0.882375	16.64799	0.86940	5.2222222		
Source	DF	Type I SS	Mean Square	F Value	Pr > F
T	2	11.1111111	5.5555556	7.35	0.0082
X	1	56.7474747	56.7474747	75.08	0.0001
X*T	2	0.1823498	0.0911749	0.12	0.8874
Source	DF	Type III SS	Mean Square	F Value	Pr > F
T	2	2.4915571	1.2457785	1.65	0.2331
X	1	43.1507138	43.1507138	57.09	0.0001
X*T	2	0.1823498	0.0911749	0.12	0.8874
T for H0: Pr > T Std Error of					
Parameter		Estimate	Parameter=0		Estimate
INTERCEPT		2.500000000 B	3.35	0.0057	0.74527426
T	1	-2.000000000 B	-1.55	0.1477	1.29240452
	2	0.131578947 B	0.14	0.8913	0.94265770
	3	0.000000000 B	.	.	.
X		0.500000000 B	3.81	0.0025	0.13106625
X*T	1	-0.000000000 B	-0.00	1.0000	0.18993304
	2	0.061403509 B	0.40	0.6977	0.15430038
	3	0.000000000 B	.	.	.
NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.					

ANALISIS DE COVARIANZA					
General Linear Models Procedure - Class Level Information					
Class		Levels		Values	
T		3		1 2 3	
Number of observations in data set = 18					
Dependent Variable: Y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	68.0409357	13.6081871	18.00	0.0001
Error	12	9.0701754	0.7558480		
Corrected Total	17	77.1111111			
R-Square	C.V.	Root MSE	Y Mean		
	0.882375	16.64799	0.86940 5.2222222		
Source	DF	Type I SS	Mean Square	F Value	Pr > F
T	2	11.1111111	5.5555556	7.35	0.0082
X	1	56.7474747	56.7474747	75.08	0.0001
X*T	2	0.1823498	0.0911749	0.12	0.8874
Source	DF	Type III SS	Mean Square	F Value	Pr > F
T	2	2.4915571	1.2457785	1.65	0.2331
X	1	43.1507138	43.1507138	57.09	0.0001
X*T	2	0.1823498	0.0911749	0.12	0.8874
T for H0: Pr > T Std Error of					
Parameter		Estimate	Parameter=0		Estimate
INTERCEPT		2.500000000 B	3.35	0.0057	0.74527426
T	1	-2.000000000 B	-1.55	0.1477	1.29240452
	2	0.131578947 B	0.14	0.8913	0.94265770
	3	0.000000000 B	.	.	.
X		0.500000000 B	3.81	0.0025	0.13106625
X*T	1	-0.000000000 B	-0.00	1.0000	0.18993304
	2	0.061403509 B	0.40	0.6977	0.15430038
	3	0.000000000 B	.	.	.
NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.					

La interacción tratamiento por covariable no es significativa, es decir que se puede sostener que los diferentes tratamientos responden igual al peso inicial, se puede ajustar una única regresión.

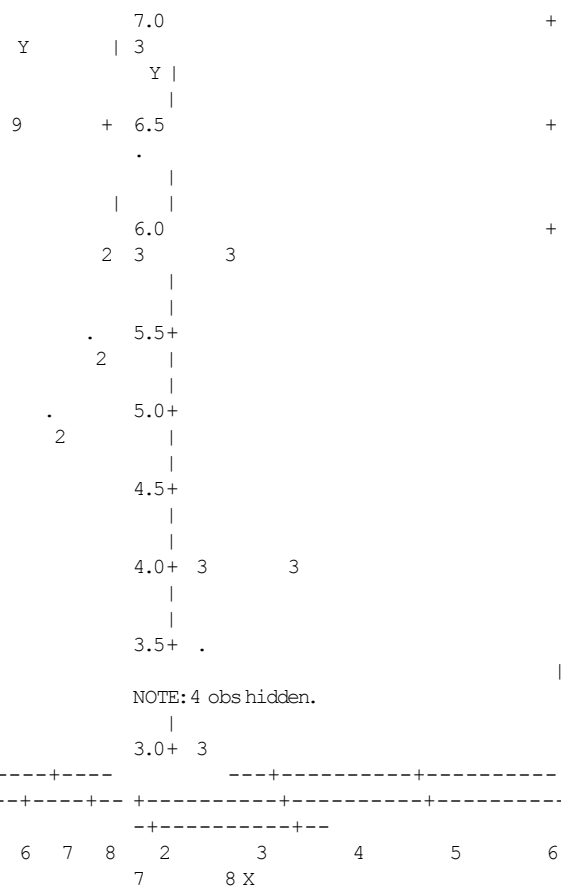
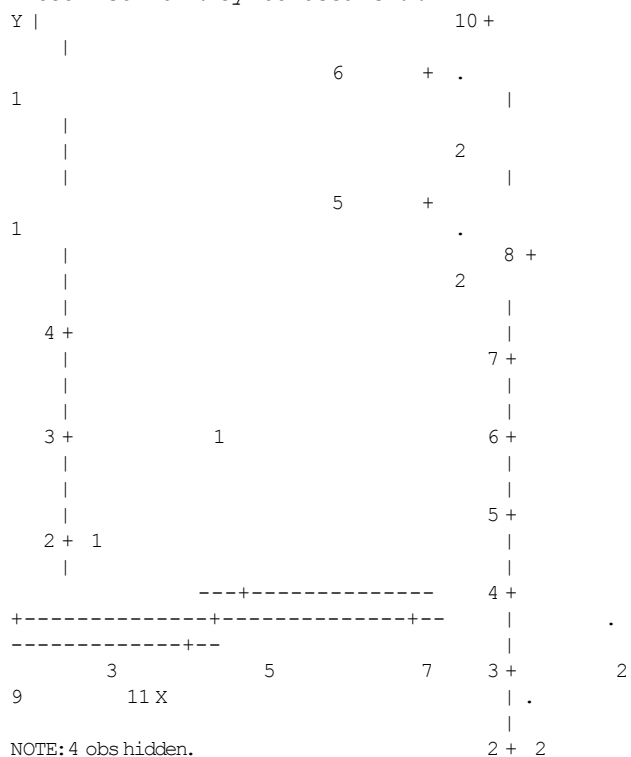
ANALISIS DE COVARIANZA

PESO VS EDAD POR TRATAMIENTO

-----T=1

Plot of Y*X. Symbol is value of T.

Plot of YGORRO*X. Symbol used is '!'.



OTROS CASOS DE ANALISIS DE COVARIANZA. COMO QUIEN VA Y COMO QUIEN VIENE

EL CASO DEL PESO AL NACER DE LOS NIÑOS. El modelo es $Y = \mu + S + EG + S*EG + \varepsilon$, donde Y es el peso al nacer de los bebés, s es el efecto del sexo del bebé, eg es la edad gestacional (tiempo desde la concepción al nacimiento), s*eg es la interacción y ε es el error. En un enfoque clásico de análisis de covarianza la variable cuantitativa (la edad gestacional) es una molestia para comparar los niveles de la variable cualitativa (el sexo en este caso). Por lo tanto se corrige por edad para obtener el peso al nacer de cada sexo. Pero en este caso el investigador en realidad lo que quiere hacer es modelar la situación, ambas variables son igualmente importantes, ni se menciona que es un análisis de covarianza. Sin embargo el modelo tiene variables cuantitativas y cualitativas.

EL CASO DE LAS PASTURAS. Saldanha & Saravia estudian la relación entre cantidad de agua disponible (X_1), cantidad de agua demandada (X_2) y crecimiento de las pasturas en 3 diferentes suelos y para las 4 estaciones. Entonces hacen una regresión para cada una de las 12 combinaciones de suelos y estaciones. Pero si queremos un modelo para todas las situaciones tenemos:

$$Y = \mu + S + E + S*E + X_1 + X_2 + \varepsilon$$

Suponiendo que procedemos al revés: el investigador ajusta el modelo completo y observa interacciones significativas, al margen que puede proceder de otra manera en realidad lo que le interesa es para cada combinación de suelo y estación (que reaccionen igual – es decir que no interactúen) saber cual es la relación entre la tasa de crecimiento de la pastura con X_1 y X_2 .

EL CASO DE LA DENSIDAD DE POBLACIÓN. En una investigación se intentó medir la densidad de población. Cuando se observaron las poblaciones obtenidas se vio que no coincidían con lo intentado y se propuso un análisis de covarianza. En ese caso parece mejor hacer una regresión (con el número de plantas implantadas como predictora) que un análisis de covarianza propiamente dicho (es decir considerar al número de plantas una variable independiente que afectó la situación).

COVARIANZA VS. BLOQUES. Muchos investigadores se plantean el dilema de si debe usar bloques o hacer un análisis de covarianza, como en el ejemplo 4.1 de este capítulo. Recordemos las propiedades.

CUANDO LA COVARIABLE ES AFECTADA POR LOS TRATAMIENTOS. En ese caso recomendamos una estrategia del tipo coeficientes de paso.

la ASO DE