DISEÑO DE EXPERIMENTOS

Diseños experimentales con factores de bloqueo

Licenciatura en Estadística
Profesores:
Dr. José Alberto Pagura
Lic. Julia Inés Fernández

Diseño en bloque completos aleatorizados

El diseño en bloques completos aleatorizados

- El Diseño Completamente Aleatorizado presupone que las unidades experimentales son homogéneas respecto a la influencia que pueden tener sobre las variables de respuesta.
- En muchos experimentos esto no ocurre, como puede apreciarse en los siguientes ejemplos:
- En un estudio sobre el rendimiento de un cultivo se sembrará en parcelas de difierentes ubicaciones lo que presupone diferentes calidades de suelo.
- Se desea saber si la posición del capot de un auto en una cabina de pintura tiene efecto sobre el número de imperfecciones que se producen al pintarlo. Los ensayos se deben realizar en diferentes turnos con diferentes operarios.

El diseño en bloques completos aleatorizados

- En casos como los ejemplos mencionados, si se realiza un diseño completamente aleatorizado la variancia del error experimental podría aumentar, lo que traería como consecuencia que no se pueda apreciar la existencia de efectos que afectan la respuesta .
- Es conveniente ensayar los tratamientos a estudiar sobre unidades experimentales homogéneas, por ejemplo, ensayar en el mismo turno y con los mismos operarios todas las posiciones de los capots que se consideran. Luego hacer otra replica en condiciones homogéneas etc.
- Esta clase de problemas pueden tratarse mediante la implementación de un diseño en bloques.

El diseño en bloques completos aleatorizados

- Un bloque estará compuesto por un grupo de unidades experimentales homogéneas – los diferentes capots a pintar en un turno- y sobre cada una de ellas se aplicará un nivel diferente del factor, asignado de forma aleatoria.
- Conducir un experimento de esta forma, permitirá reducir la variancia del error experimental mejorando la sensibilidad en la detección de efectos existentes.

El modelo en el diseño en bloques completos aleatorizados

- Se consideran los efectos de dos factores:

  Factor de interés para el investigador

  Bloques: los que desearíamos que su influencia fuera eliminada, si existe

- El modelo: $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$ *i=1…a; j=1…b*

  $\varepsilon_{ij} \sim NID(0;\sigma^2)$

Tabla ANOVA

FV	GL	SC	CM	F
Bloque	b - 1	$\sum_{j=1}^{b} a\ (\bar{y}_{.j} - \bar{y}_{..})^2$	$SC_{bloque}/(b-1)$	--------------------
Factor	a - 1	$\sum_{i=1}^{a} b\ (\bar{y}_{i.} - \bar{y}_{..})^2$	$SCf_{actor}/(a-1)$	CM_{factor}/CM_{error}
Error	N - a	$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$SC_{error}/(N-a)$	
Total	N - 1	$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2$		

Ejemplo

- Se comparan tres soluciones para lavar envases de leche.
- La efectividad de las mismas se mide en el retraso que se logra en el crecimiento de bacterias en los envases lavados
- Se aplicarán las tres soluciones a tres envases, se realizará un recuento bacteriano y se experimentará en cuatro días diferentes
- Sospechando la existencia de variabilidad entre los días, se aplicarán las tres soluciones cada uno de los días, decidiendo aleatoriamente en que envase se utiliza cada solución.

Resultados (recuento bacteriano)

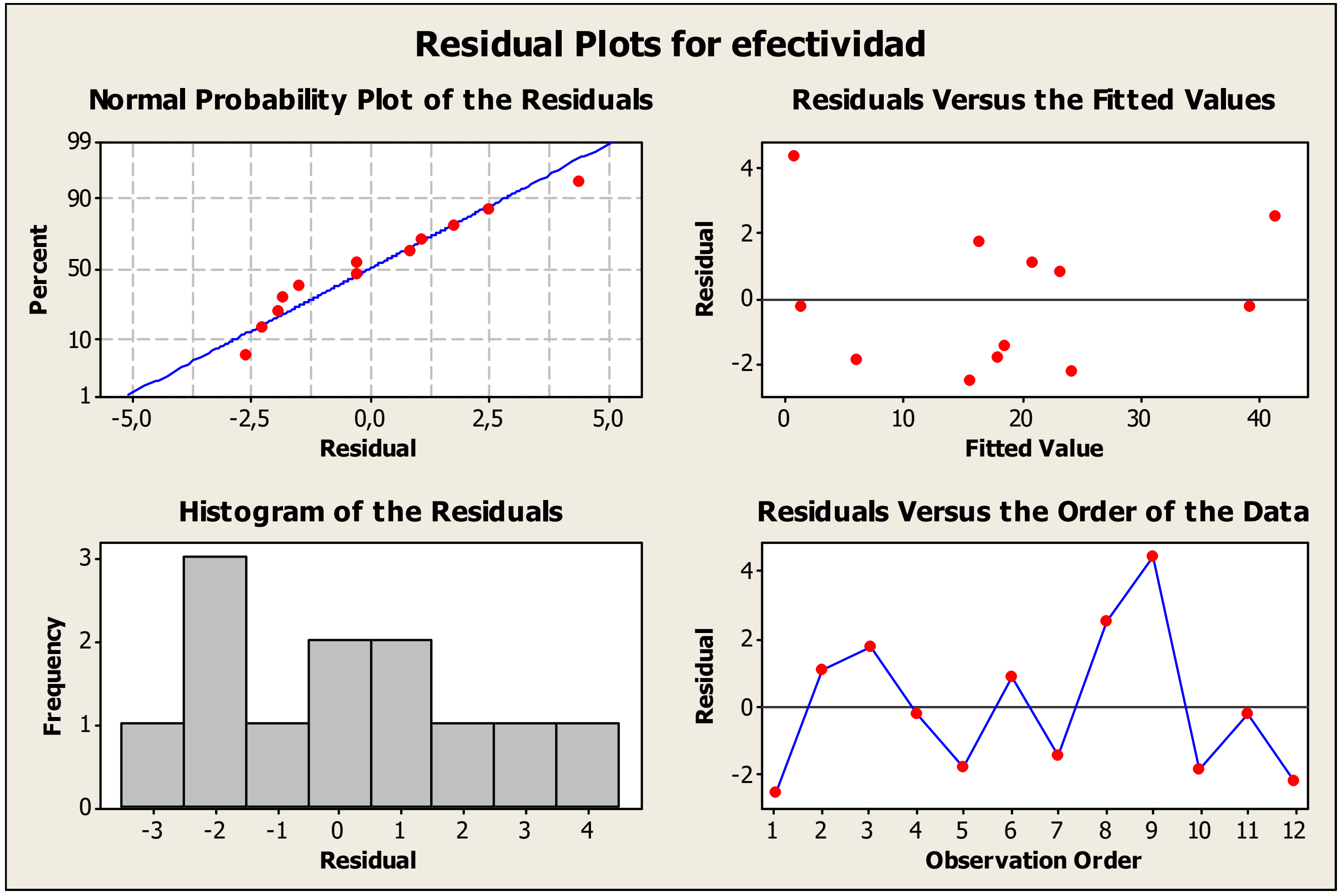
Solución	**Días**			
	1	**2**	**3**	**4**
1	13	22	18	39
2	16	24	17	44
3	5	4	1	22

El ANOVA

- La tabla ANOVA:

FV	GL	SC	CM	F	P
solución	2	703,50	351,75	40,72	0,000
día	3	1106,92	368,97	42,71	0,000
Error	6	51,83	8,64		
Total	11	1862,25			

Validación del modelo



Primeras conclusiones

- Hay diferencias estadísticamente significativas entre los retrasos promedios del crecimiento de bacterias para las diferentes soluciones
- La decisión de utilizar el día como bloque, permite reducir el error experimental y haciendo eficiente la comparación entre los tratamientos

¿Cuál "solución" es la "mejor"?

- Teniendo en cuenta que la hipótesis de igualdad de medias no fue aceptada, la realización de comparaciones múltiples permitirá decidir "cual es la mejor", *en este caso, que "solución es la que más retrasa el crecimiento de bacterias*.

Comparaciones múltiples

```
Tukey Simultaneous Tests
Response Variable efectividad
All Pairwise Comparisons among Levels of solución
solución = s1  subtracted from:

           Difference       SE of              Adjusted
solución     of Means  Difference  T-Value   P-Value
s2               2,25       2,078    1,083    0,5578
s3             -15,00       2,078   -7,217    0,0009


solución = s2  subtracted from:

           Difference       SE of              Adjusted
solución     of Means  Difference  T-Value   P-Value
s3             -17,25       2,078   -8,300    0,0004
```

Diferencias observadas

- El “recuento bacteriano” promedio para la solución 3 presenta diferencias estadísticamente significativas con respecto a los de las soluciones 1 y 2.

- Este promedio, resulta menor que los otros dos.

Estimación de la respuesta media para cada "solución"

- Las medias estimadas y sus errores estándares son:

solución	Media	EE(Media)
1	23,00	5,64
2	25,25	6,50
3	8,00	4,74

Diseño de cuadrado latino

Diseño de cuadrado latino

- Este diseño usa el principio de la formación de bloques para reducir el error experimental controlando dos factores perturbadores.
- EJEMPLO
- Un experimentador estudia los efectos que tienen cinco fórmulas diferentes de la carga propulsora utilizada en los sistemas de expulsión de la tripulación de un avión sobre la velocidad de combustión.
- Cada lote de materia primera alcanza para realizar cinco compuestos.

Ejemplo (continuación)

- Las fórmulas son preparadas por distintos operadores que por su capacitación, pueden tener influencia sobre la respuesta.
- Hay dos factores "perturbadores" que deberían controlarse en el diseño: lote y operador. El diseño de cuadrado latino es adecuado para este fin.
- En el ejemplo, el experimento consiste en tomar cinco lotes de materia prima, hacer que cinco operarios realicen una vez cada una de las cinco fórmulas

¿Cómo se construye un cuadrado latino?

- En términos generales, un cuadrado latino para p tratamientos, o cuadrado latino $p \times p$ es un cuadrado con p filas y p columnas. Cada una de las p^2 celdas contiene una de las p letras con las que designa cada tratamiento, y cada letra ocurre una y solo una vez en cada fila y columna.
- Puede obtenerse un cuadrado latino estándar escribiendo la primera fila en orden alfabético y cada fila sucesiva escribiendo las mismas letras de la fila anterior pero corridas un lugar a la izquierda

Cuadrado latino para el ejemplo

- Un cuadrado latino para el ejemplo puede ser:

Lotes de materia prima	Operador				
	1	2	3	4	5
1	A	B	C	D	E
2	B	C	D	E	A
3	C	D	E	A	B
4	D	E	A	B	C
5	E	A	B	C	D

- La fórmula A se ensayará por el operador 1 con el lote 1, por el operario 2 con el lote 5, etc.

Modelo estadístico

- El modelo estadístico: $y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \varepsilon_{ijk}$

  con $\varepsilon_{ijk} \sim NID(0, \sigma^2)$

  α_i efecto de "fila", $i=1 \ldots p$

  τ_j efecto del factor, $j=1 \ldots p$

  β_k efecto de "columna", $k=1 \ldots p$

Realizando la partición de la SCtotal en las tres fuentes de variación consideradas además del error se construye la tabla ANOVA, ,

Análisis de la variancia

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F
Factor	p-1	$SC_{factor} = p\sum_{j=1}^{p}\left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2$	$CM_{factor} = SC_{factor} / (p-1)$	$Fo = \dfrac{CM_{factor}}{CM_{error}}$
Filas	p-1	$SC_{fila} = p\sum_{i=1}^{p}\left(\bar{y}_{i..} - \bar{y}_{...}\right)^2$	$CM_{fila} = SC_{fila} / (p-1)$	
Columnas	p-1	$SC_{columna} = p\sum_{k=1}^{p}\left(\bar{y}_{..k} - \bar{y}_{...}\right)^2$	$CM_{columna} = SC_{columna} / (p-1)$	
Error	(p-2)(p-1)	SC_{error}	$CM_{error} = SC_{error} / \left[(p-2)(p-1)\right]$	
Total	P^2-1	$SC_{total} = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\left(y_{ijk} - \bar{y}_{..}\right.$		

Adecuación del modelo

- Como en otros casos, se podrá evaluar el cumplimiento de los supuestos del modelo considerando los gráficos de residuos vs. Ajustados, residuos en “papel normal” y residuos vs. orden de los ensayos.

- Los valores ajustados serán: $y_{ijk} = \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - 2\bar{y}_{...}$

- Los residuales resultan: $e_{ijk} = y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...}$

Resultados

- Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
fórmula	4	330,00	82,50	7,73	0,003
lote	4	68,00	17,00	1,59	0,239
operador	4	150,00	37,50	3,52	0,040
Error	12	128,00	10,67		
Total	24	676,00			

- Que conclusiones pueden extraerse?

Diseño de cuadrado grecolatino

Cuadrado grecolatino

- Permiten controlar tres fuentes de variación atribuibles a factores "perturbadores".
- Se construye superponiendo dos cuadrados latinos, uno con letras latinas y otro con letras griegas, de modo que cada letra latina aparece con cada letra griega una y solo una vez.
- Existen cuadrados grecolatinos para $p \geq 3$ excepto para $p = 6$

Cuadrado grecolatino 4x4

Fila	Columna			
	1	2	3	4
1	Aα	Bβ	Cγ	Dδ
2	Bδ	Aγ	Dβ	Cα
3	Cβ	Dα	Aδ	Bγ
4	Dγ	Cδ	Bα	Aβ

Modelo estadístico

- El modelo estadístico:

$$y_{ijkl} = \mu + \theta_i + \tau_j + \omega_k + \psi_i + \varepsilon_{ijkl}$$

Donde i, j, k, l toman valores de 1 a p

y_{ijkl} es la observación correspondiente a la fila i y la columna l para la letra latina j y la letra griega k.

ANOVA

Fuente de variación	g.l.	Suma de cuadrados
Factor con letras latinas	p-1	$SC_{columnas} = \frac{1}{p}\sum_{j=1}^{p} y_{...l}^2 - \frac{y_{...}^2}{N}$
Factor con letras griegas	p-1	$SC_G = \frac{1}{p}\sum_{j=1}^{p} y_{...k}^2 - \frac{y_{...}^2}{N}$
Filas	p-1	$SC_{filas} = \frac{1}{p}\sum_{j=1}^{p} y_{i...}^2 - \frac{y_{...}^2}{N}$
Columnas	p-1	$SC_{columnas} = \frac{1}{p}\sum_{j=1}^{p} y_{...l}^2 - \frac{y_{...}^2}{N}$
Error	(p-3)(p-1)	Diferencia entre SC_{total} y las cuatro sumas de cuadrados anteriores
Total	p^2-1	$SC_{total} = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} y_{ijkl}^2 - \frac{y_{...}^2}{N}$

Se calculan los cocientes CM_L/CM_{error} y CM_G/CM_{error} y se calcula el p-value con una $F_{p-1;(p-3)(p-1)}$.