Diseño en Bloques Completamente Aleatorizados

DISEÑO DE UN FACTOR-Diferentes Fuentes de Variabilidad

Si se desean controlar fuentes de variabilidad externas que pueden influir para detectar diferencias entre los TRATAMIENTOS existen los siguientes diseños:

- Diseño en **bloques** completos Aleatorizados (UNA fuente de variabilidad)
- Diseño en **Cuadrados Latinos** (DOS fuentes de variabilidad)
- Diseño en **Cuadrados Grecolatinos** (TRES fuentes de variabilidad)

Diseño en Bloques Completamente Aleatorizados

Introducción

1. En el **DECA** no es posible controlar algunos factores extraños de variación (o Factores de confusión).
2. Si los factores de confusión existen, el $\mathbf{MS_E}$ será un estimador **sesgado** para la varianza del error experimental, es decir:

$$\mathbf{E}[\mathbf{MS_E}] = \sigma^2 + \text{otra Cantidad}$$

   es decir, el $\mathbf{MS_E}$ es más grande de lo normal y por tanto la prueba $\mathbf{F}$ tenderá al NO rechazo de la hipótesis nula global de igualdad de medias.
3. En el caso anterior, es recomendable utilizar un diseño que controle estos factores extraños de variación (o factores de confusión).

Diseño en Bloques Completamente Aleatorizados

Introducción Parte 2

1. ¿En caso de existir una variación de qué tipo es y a qué se debe?
   La variación que existe es de tipo sistemática y puede ser debida a:
   - La naturaleza del problema o
   - Puede ser " inducida " o introducida por el investigador para ampliar las inferencias acerca de los tratamientos.
2. ¿Cuál es el principal objetivo de la experimentación?
   Controlar las fuentes de variación extrañas con el fin de que incidan directamente en la reducción asociada a la variación del término de error.
3. ¿Qué se pretende con la técnica de bloques?
   - Conseguir una mayor homogeneidad entre los sujetos o unidades experimentales intra-bloque y
   - Reducir el tamaño del error experimental. La formación de bloques homogéneos se realiza a partir de los valores de una variable altamente relacionada con la variable dependiente.

Diseño en Bloques Completamente Aleatorizados

4. ¿Cómo se garantiza la aleatorización al utilizar bloques?
   La aleatorización se garantiza ya que, dentro de los bloques, las UE son asignadas aleatoriamente a las distintas condiciones experimentales.
5. ¿Qué se consigue con la técnica de bloques?
   - Un material experimental mucho más homogéneo.
   - Se reduce la magnitud del error experimental.
   - Se incrementa el grado de precisión del experimento.

Diseño en Bloques Completamente Aleatorizados

Algunos ejemplos de Bloques de variación sistemática natural

- En experimentos sobre terreno a campo abierto.
  Usualmente cada bloque consiste en:
  - Un grupo de parcelas aprox. cuadradas; o
  - parcelas que dependen del gradiente de fertilidad (debido a la pendiente del terreno, por ejemplo) tal que las **UE** sobre el mismo nivel de gradiente son más semejantes que las de diferentes niveles;
  - o puede ser que una quebrada atraviese el terreno y así las parcelas equidistantes de la quebrada son más semejantes que aquellas que no lo están.
- En experimentos con animales.
  Los animales se colocan en grupos de resultados o bloques con base en características tales como:
  - Peso inicial,
  - condición del animal,
  - raza,
  - sexo,
  - edad,
  - etapa de lactancia,
  - producción de leche en el ganado,
  - camadas en cerdos y ratones, entre otros.

Ejemplos de Bloques de variación Inducida o introducida por el investigador

- Un investigador en un **experimento industrial** puede decidir obtener el material experimental de diferentes distribuidores quienes usan diferentes procesos de producción.

- En un ensayo de **alimentación de ganado** puede ser importante incluir animales de diferentes razas.

- En un experimento para probar diferentes **marcas de llantas** se quieren incluir carros de diferentes fabricantes y diferentes modelos de cada fabricante.

Diseño en Bloques Completamente Aleatorizados

Diseño de un factor con un BLOQUE

- Es uno de los diseños más usados en diferentes áreas, creado en 1925 por *R. A Fisher* quien buscaba métodos para el mejoramiento de experimentos en el campo agrícola. Su nombre se debe a experimentación en esta área, el cultivo se dividía en bloques y estos en parcelas que recibían los tratamientos.
- La situación más común se presenta cuando hay **UN** factor extraño de variación (o un factor de confusión). En este caso, la precisión de la estimación de σ^2 puede mejorarse usando un Diseño de Bloques Completamente al Azar (**DEBCA**).
- Por medio del agrupamiento de las UE en **subgrupos homogéneos**, denominados **BLOQUES** la variación asociada con el factor de confusión, se puede remover de la estimación de la media cuadrática del error.

Diseño en Bloques Completamente Aleatorizados

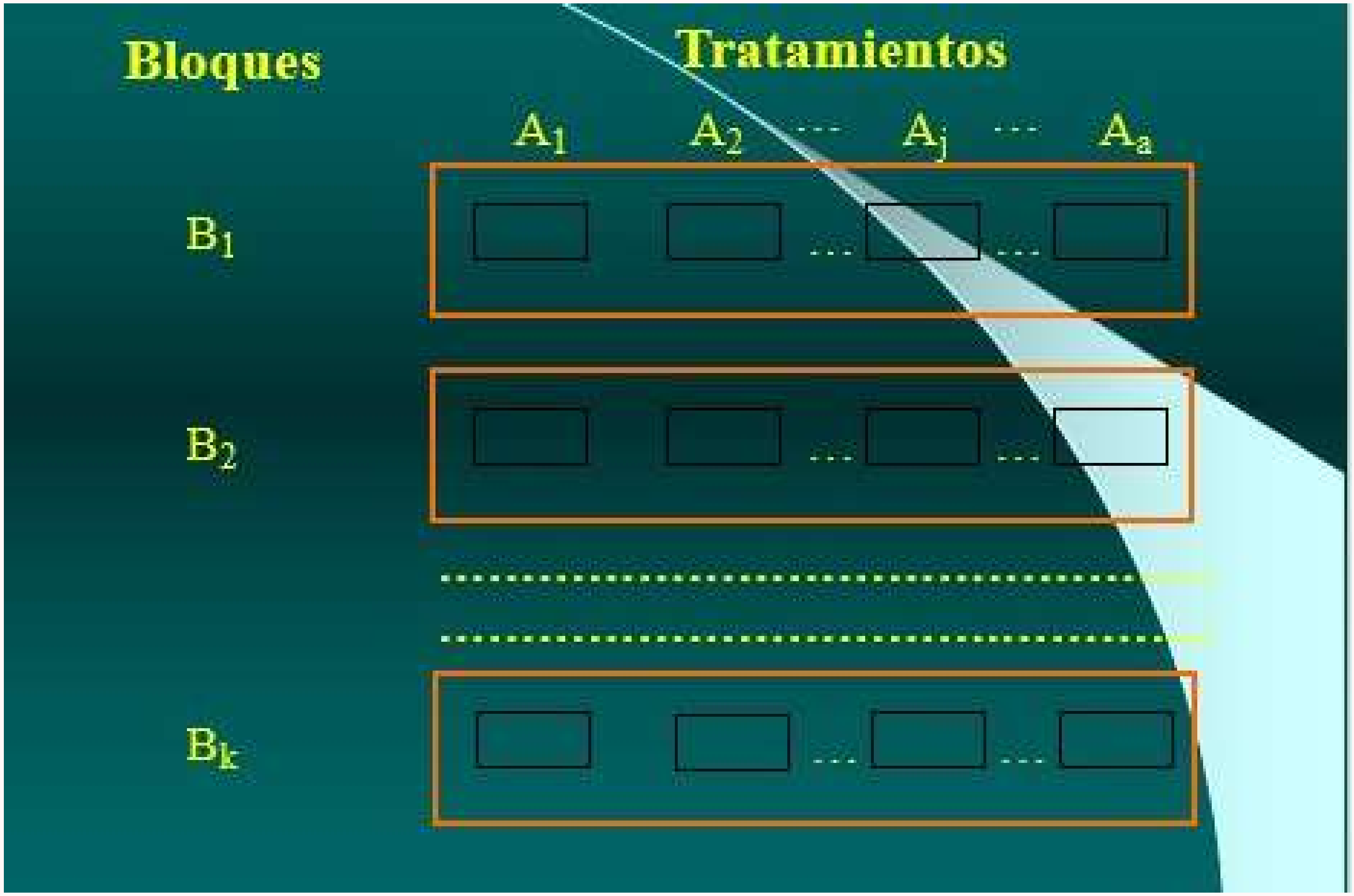
...

- El **DEBCA** es un diseño donde las **UE** a las que se le aplican los tratamientos se subdividen en grupos Homogéneos llamados BLOQUES de modo que el número de **UE** en un bloque debe ser igual al número de tratamientos en estudio.
- Cada tratamiento aparece en cada **bloque** y cada **bloque** recibe todos los tratamientos.

  En resumen: Los factores de confusión son **controlados** por medio del bloqueo.

Diseño en Bloques Completamente Aleatorizados

Gráficamente como se realiza un Diseño En Bloques



Diseño en Bloques Completamente Aleatorizados

OBJETIVO

Aislar y **eliminar** del término de error la variación atribuíble a los bloques y asegurar que las medias del tratamiento estén libres de los efectos del **BLOQUE**.

Notas

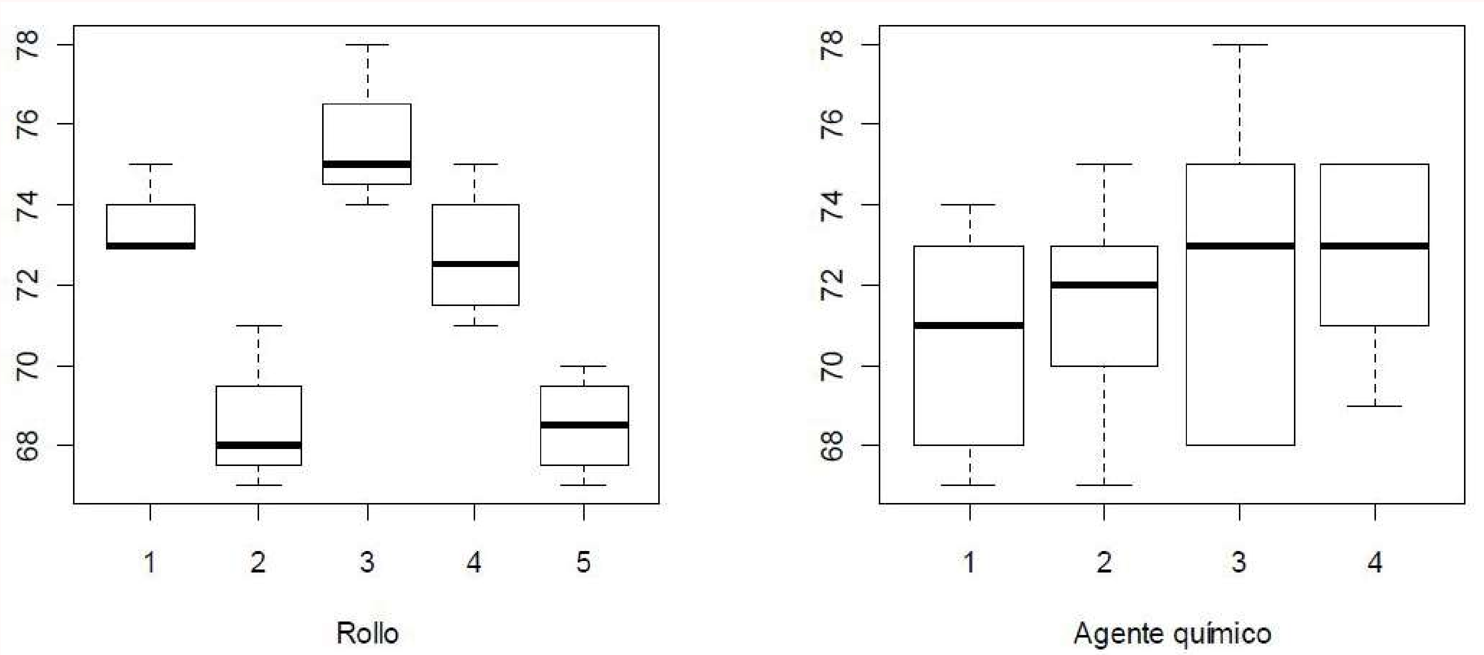
- La efectividad del método depende de la capacidad de conseguir **bloques homogéneos** de **UE** que a su vez depende del conocimiento de los investigadores acerca del material experimental.
- Si el diseño se usa apropiadamente, el **MSE** disminuye aumenta la **F** y mejora la probabilidad de rechazar $\mathbf{H_0}$.
- Los factores de **bloqueo** que aparecen en la práctica son: **TURNO, LOTE, DIA, TIPO DE MATERIAL, LÍNEA DE PRODUCCIÓN, OPERADOR, MÁQUINA, MÉTODO.**
- **NOTA**: Al existir bloques hace que no sea práctico aleatorizar en su totalidad.

Caso de Estudio 3: Resistencia de una tela

Un químico quiere probar el efecto de cuatro(4) agentes químicos sobre la **resistencia** de un tipo particular de tela. Selecciona cinco(**5**) rollos y aplica cuatro(4) agentes químicos de manera aleatoria a cada rollo. Las resistencias a la tensión resultantes son:

	Rollo				
Agente Químico	1	2	3	4	5
1	73	68	74	71	67
2	73	67	75	72	70
3	75	68	78	73	68
4	73	71	75	75	69

Continuación Caso de estudio 3



Diseño en Bloques Completamente Aleatorizados

Un ejemplo: Ejemplo 4 de la Introducción: Octanaje

Se desea determinar cual de **cinco**(**5**) tratamientos a la gasolina produce en promedio mayor octanaje, para lo cuál, se seleccionaron cuatro(**4**) barriles de gasolina y cada barril se dividió en cinco(5) partes, a cada una de las partes, en forma aleatoria se le aplicó cada uno de los distintos tratamientos. Los datos son los siguientes:

Tratamiento	Barril 1	Barril 2	Barril 3	Barril 4
A	91,7	91,2	90,9	90,6
B	91,7	91,9	90,9	90,9
C	92,4	91,2	91,6	91,0
D	91,8	92,2	92,0	91,4
E	93,1	92,9	92,4	92,4

Diseño en Bloques Completamente Aleatorizados

Disposición de los Datos

	Bloque 1	Bloque 2	…	Bloque b
T_1	Y_{11}	Y_{12}	…	Y_{1b}
T_2	Y_{21}	Y_{22}	…	Y_{2b}
$\vdots$	$\vdots$	$\vdots$	…	$\vdots$
T_a	Y_{a1}	Y_{a2}	…	Y_{ab}

con: $\mathbf{Y_{i.}}$: Total de observaciones bajo el $\mathbf{i}$-ésimo tratamiento, $\mathbf{i} = \mathbf{1}, \mathbf{2}, \cdots, \mathbf{a}.$

$\mathbf{Y_{.j}}$: Total de observaciones bajo el j-ésimo bloque, $\mathbf{j} = \mathbf{1}, \mathbf{2}, \cdots, \mathbf{b}.$

$\overline{\mathbf{Y}}_{\mathbf{i.}}$: Promedio de observaciones bajo el i-ésimo tratamiento, $\mathbf{i} = \mathbf{1}, \mathbf{2}, \cdots, \mathbf{a}.$

$\overline{\mathbf{Y}}_{\mathbf{.j}}$: Promedio de observaciones bajo el j-ésimo bloque, $\mathbf{j} = \mathbf{1}, \mathbf{2}, \cdots, \mathbf{b}.$

$\mathbf{Y_{..}}$: Total de todas las observaciones.

$\overline{\mathbf{Y}}_{\mathbf{..}}$: Promedio de todas las observaciones.

Diseño en Bloques Completamente Aleatorizados

Notación

$$\mathbf{Y_{i\,.}} = \sum_{j=1}^{b} \mathbf{Y_{ij}} \quad \mathbf{i} = 1, 2, \ldots, \mathbf{a}.$$

$$\mathbf{Y_{.\,j}} = \sum_{i=1}^{a} \mathbf{Y_{ij}} \quad \mathbf{j} = 1, 2, \ldots, \mathbf{b}.$$

$$\mathbf{Y_{.\,.}} = \sum_{i=1}^{a}\sum_{j=1}^{b} \mathbf{Y_{ij}} = \sum_{i=1}^{a} \mathbf{Y_{i\,.}} = \sum_{j=1}^{b} \mathbf{Y_{.\,j}}$$

Promedios

$$\bar{\mathbf{Y}}_{\mathbf{i}\,.} = \frac{\mathbf{Y_{i\,.}}}{\mathbf{b}}$$

$$\bar{\mathbf{Y}}_{.\,\mathbf{j}} = \frac{\mathbf{Y_{.\,j}}}{\mathbf{a}}$$

$$\bar{\mathbf{Y}}_{.\,.} = \frac{\mathbf{Y_{.\,.}}}{\mathbf{N}}$$

Modelo Estadístico: Modelo de Efectos

$$\boxed{\mathbf{Y_{ij}} = \mu + \tau_{\mathbf{i}} + \beta_{\mathbf{j}} + \varepsilon_{\mathbf{ij}}, \quad \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{a}; \quad \mathbf{j} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{b}}$$

$\tau_{\mathbf{i}}$: Efecto asociado al **tratamiento** $\mathbf{i}$

$\beta_{\mathbf{j}}$: Efecto asociado al **bloque** $\mathbf{j}$.

Supuestos: $\varepsilon_{\mathbf{ij}} \sim \mathbf{NID}(\mathbf{0}, \sigma^{\mathbf{2}})$

Restricciones: $\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{a}} \tau_{\mathbf{i}} = \mathbf{0}$ y $\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}} \beta_{\mathbf{j}} = \mathbf{0}$.

Diseño en Bloques Completamente Aleatorizados

En otras palabras: Los **supuestos** del modelo son:

- Normalidad
- Independencia
- Homogeneidad de varianzas
- No interacción de Bloques con tratamientos (Aditividad o modelo aditivo)

Equivalentemente: **Modelo de Medias**:

$$\boxed{\mathbf{Y_{ij} = \mu_{ij} + \varepsilon_{ij}, \;\; i = 1, 2, \ldots, a; \; j = 1, 2, \ldots, b,}}$$

donde

$$\mu_{ij} = \mu + \tau_i + \beta_j$$

Diseño en Bloques Completamente Aleatorizados

OBJETIVO

Hipótesis Central: Probar igualdad entre las medias de los tratamientos:

$$\mathbf{H_0}: \mu_{\mathbf{1}} = \mu_{\mathbf{2}} = \ldots = \mu_{\mathbf{a}} = \mu \quad \mathbf{H_1}: \mu_{\mathbf{i}} \neq \mu_{\mathbf{j}} \text{ para algún } \mathbf{i} \neq \mathbf{j}$$

donde

$$\mu_{\mathbf{i}} = \frac{\mathbf{1}}{\mathbf{b}} \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{b}} (\mu + \tau_{\mathbf{i}} + \beta_{\mathbf{j}}) = \mu + \tau_{\mathbf{i}}.$$

La hipótesis anterior es equivalente a:

$$\mathbf{H_0}: \tau_{\mathbf{1}} = \tau_{\mathbf{2}} = \ldots = \tau_{\mathbf{a}} = \mathbf{0} \quad \mathbf{H_1}: \tau_{\mathbf{i}} \neq \mathbf{0} \text{ para algún } \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{a}.$$

Diseño en Bloques Completamente Aleatorizados

Anotaciones

Observe:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij}-\bar{Y}_{..})^2 = b\sum_{i=1}^{a}(\bar{Y}_{i.}-\bar{Y}_{..})^2 + a\sum_{j=1}^{b}(\bar{Y}_{.j}-\bar{Y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij}-\bar{Y}_{.j}-\bar{Y}_{i.}+\bar{Y}_{..})^2$$

$$SST = \mathbf{SSA} + \mathbf{SSBloque} + \mathbf{SSE}$$

(gl) $\quad ab-1 = (a-1) + (b-1) + (a-1)(b-1)$

Diseño en Bloques Completamente Aleatorizados

Resultado

Bajo los supuestos del modelo se tiene:

a. $\mathbf{E[MSA] = E[\frac{SSA}{a-1}] = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1}}$

b. $\mathbf{E[MSB] = E[\frac{SSB}{b-1}] = \sigma^2 + \frac{a\sum_{j=1}^{b}\beta_j^2}{b-1}}$

c. $\mathbf{E[MSE] = \sigma^2}$

Además:

$\frac{\mathbf{SSA}}{\sigma^2} \sim \chi^2_{\mathbf{a-1}},$ Bajo $H_0 : \tau_1 = \ldots = \tau_a = 0$

$\frac{\mathbf{SSB}}{\sigma^2} \sim \chi^2_{\mathbf{b-1}},$ Bajo $H_0 : \beta_1 = \ldots = \beta_b = 0$

$\frac{\mathbf{SSE}}{\sigma^2} \sim \chi^2_{\mathbf{(a-1)(b-1)}},$ Bajo los supuestos del modelo, no se necesita que H_0 sea cierta

Diseño en Bloques Completamente Aleatorizados

Prueba de Hipótesis Central del DISEÑO

Estadístico de Prueba

$$F = \frac{MSA}{MSE} \sim F_{a-1,(a-1)(b-1)}$$

Región de Rechazo: Rechazo H_0 si

$$F_{\text{cal}} > f_{a-1,(a-1)(b-1),\alpha}$$

- Otra prueba **IMPORTANTE** es determinar si el bloqueo utilizado contribuye de manera significativa para el modelo:

$$\mathbf{H_0} : \beta_1 = \beta_2 = \ldots = \beta_b = \mathbf{0} \quad \mathbf{H_1} : \beta_j \neq \mathbf{0} \text{ para algún } \mathbf{j} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{b}.$$

Estadístico de Prueba:

$$\mathbf{F} = \frac{\mathbf{MSB}}{\mathbf{MSE}} \sim \mathbf{F}_{\mathbf{b-1},(\mathbf{a-1})(\mathbf{b-1})}.$$

Diseño en Bloques Completamente Aleatorizados

ANOVA

Fuente de Variación	SS	gl	MS	F_0	VP
Tratamientos	SSA	$a-1$	MSA	$\frac{MSA}{MSE}$	
Bloque	SSB	$b-1$	MSB	$\frac{MSB}{MSE}$	
Error	SSE	$(a-1)(b-1)$	MSE		
Total	SST	$ab-1$			

Algunas Expresiones Útiles

$$
\begin{aligned}
SST &= \sum\sum Y_{ij}^2 - \frac{Y_{..}^2}{N} \\
SSA &= \frac{1}{b}\sum_{i=1}^{a} Y_{i.}^2 - \frac{Y_{..}^2}{N} \\
SSB &= \frac{1}{a}\sum_{j=1}^{b} Y_{.j}^2 - \frac{Y_{..}^2}{N}
\end{aligned}
$$

Anotaciones

Una vez se rechaza $\mathbf{H_0}$, igual que con análisis de un factor, se buscan que tratamientos se pueden agrupar con los diferentes métodos usando los g.l del MSE $((\mathbf{a}-\mathbf{1})(\mathbf{b}-\mathbf{1}))$

En forma análoga se **verifican los supuestos** para el modelo usando los residuales.

Estimación de los parámetros

$$
\begin{aligned}
\hat{\mu} &= \bar{\mathbf{Y}}_{..} \\
\hat{\tau}_{\mathbf{i}} &= \bar{\mathbf{Y}}_{\mathbf{i}.} - \bar{\mathbf{Y}}_{..} \\
\hat{\beta}_{\mathbf{j}} &= \bar{\mathbf{Y}}_{.\mathbf{j}} - \bar{\mathbf{Y}}_{..} \\
\hat{\mathbf{Y}}_{\mathbf{ij}} &= \hat{\mu} + \hat{\tau}_{\mathbf{i}} + \hat{\beta}_{\mathbf{j}} = \bar{\mathbf{Y}}_{\mathbf{i}.} + \bar{\mathbf{Y}}_{.\mathbf{j}} - \bar{\mathbf{Y}}_{..} \\
\mathbf{e}_{\mathbf{ij}} &= \mathbf{Y}_{\mathbf{ij}} - \hat{\mathbf{Y}}_{\mathbf{ij}}
\end{aligned}
$$

Diseño en Bloques Completamente Aleatorizados

Retomando Caso de Estudio 3.

Un químico quiere probar el efecto de cuatro agentes químicos sobre la resistencia de un tipo particular de tela. Selecciona cinco rollos y aplica cuatro agentes químicos de manera aleatoria a cada rollo. Las resistencias a la tensión resultantes son:

	Rollo				
Agente Químico	1	2	3	4	5
1	73	68	74	71	67
2	73	67	75	72	70
3	75	68	78	73	68
4	73	71	75	75	69

```
y      <-   c(73,68,74,71,67,73,67,75,72,70,75,68,78,73,68,73,71,75,75,69)
agente <- as.factor(c(rep(1,5),rep(2,5),rep(3,5),rep(4,5)))
rollo  <- as.factor(c(rep(1:5,4)))
modelo <- aov(y~agente+rollo)
summary(modelo)
```

```
              Df  Sum Sq Mean Sq F value    Pr(>F)
agente         3   12.95    4.32   2.376     0.121
rollo          4  157.00   39.25  21.606 2.06e-05 ***
Residuals     12   21.80    1.82
---------------------------------------------
Total         19  191.75
```

Con $\alpha = \mathbf{0.05}$ se tiene que: $\mathbf{F_{\alpha;a-1,(a-1)(b-1)}} = \mathbf{F_{0.05;3,12}} = \mathbf{3.49}$.

Como $F_{Cal} = \frac{MS_{Trat.}}{MS_E} = \frac{4.32}{1.82} = 2.376 < 3.49 = F_{Tabla}$, entonces no se rechaza

$$H_0 \ : \ \mu_1 = \mu_2 = \ldots = \mu_4 = \mu$$

Conclusión similar se obtiene dado que el valor-$P = 0.121 > 0.05 = \alpha$, de donde con un nivel de significancia del $5\,\%$ se concluye que las medias de los tratamientos son iguales.

Diseño en Bloques Completamente Aleatorizados

...

Para la prueba de significancia de bloques se tiene que: $F_{\alpha;b-1,(a-1)(b-1)} = F_{0.05;4,12} = 3.26$, luego como

$$F_{Cal} = \frac{MS_B}{MS_E} = \frac{39.25}{1.82} = 21.606 > 3.26 = F_{Tabla},$$

entonces se rechaza

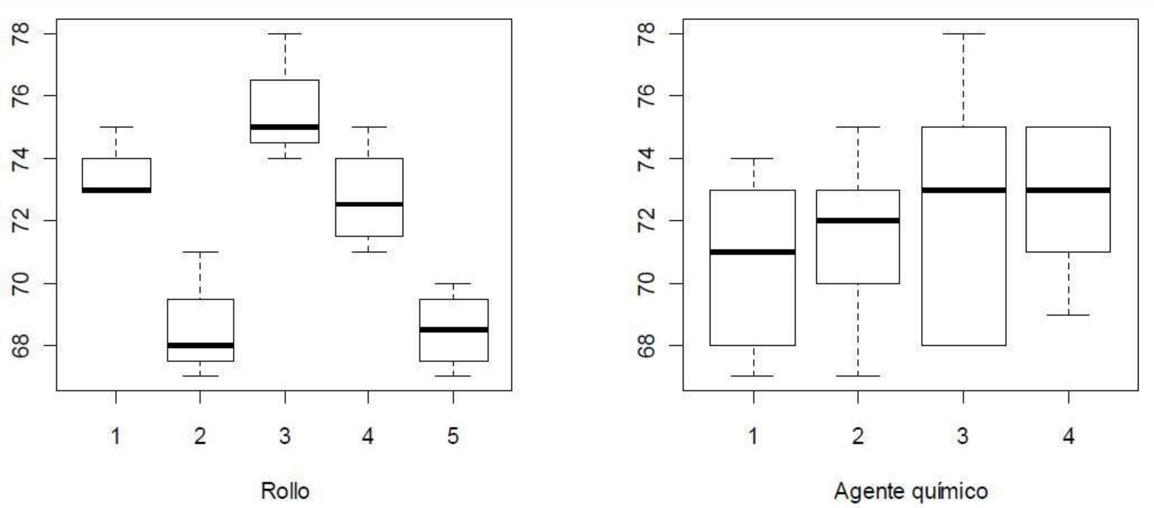
$$H_0 \ : \ \beta_1 = \beta_2 = \ldots = \beta_b = 0.$$

Similarmente, como el valor-$\mathbf{P} = \ \mathbf{2.06} \times \mathbf{10^{-5}} < \ \mathbf{0.05} = \alpha$, de donde con un nivel de significancia del $\mathbf{5}\,\%$ se concluye que el efecto de los bloques es **significativo**.

Diseño en Bloques Completamente Aleatorizados

...

```
boxplot(split(y,agente),xlab='Agente',ylab='Resistencia')
boxplot(split(y,rollo),xlab='Rollo',ylab='Resistencia')
```

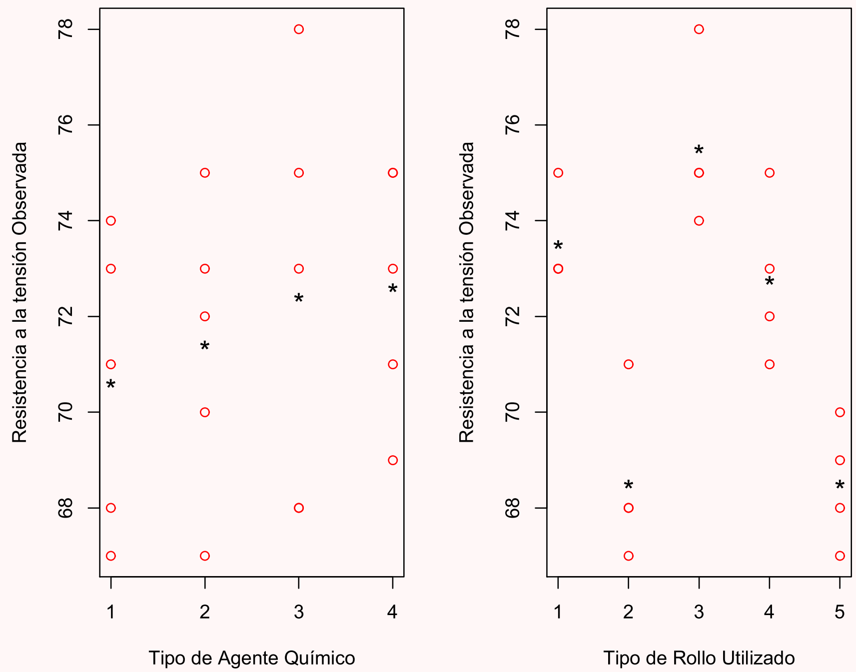


Diseño en Bloques Completamente Aleatorizados

...

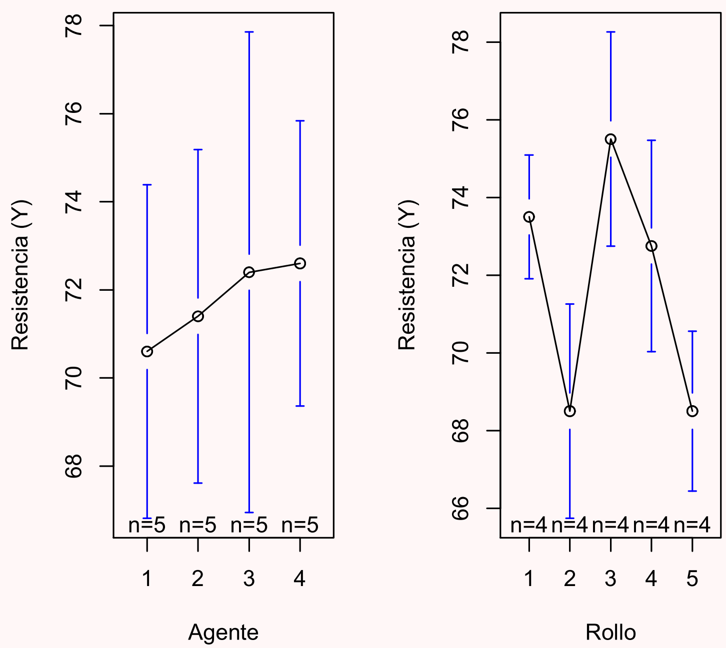
```
stripchart(y~agente,vert=T,method="overplot",pch=1,col="red",
ylab="")
stripchart(as.numeric(grp.means1)~as.numeric(names(grp.means1)),

pch="*",cex=1.5,vert=T,add=T)
```



...

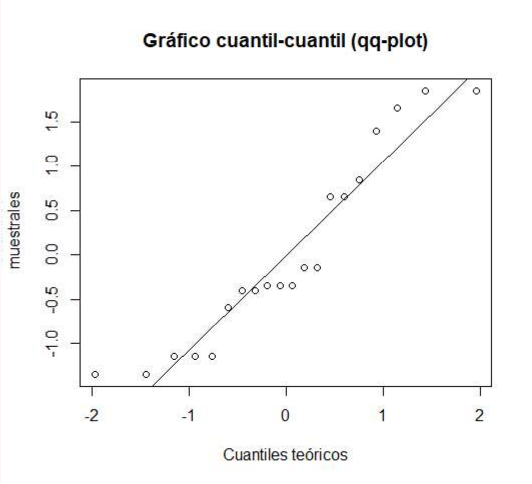
```
require(gplots)
plotmeans(y~agente,p=0.95,main="",xlab="Agente",ylab="Res (Y)")
```



NOTA: Los demás supuestos se verifican de manera análoga al procedimiento utilizado en el diseño completamente aleatorizado.

Verificación de los supuestos

Normalidad:




```
Shapiro-Wilk normality test
data:  residuales
W = 0.8996, p-value = 0.04054
```

Conclusión: A un nivel del $1\,\%$, se concluye que los errores del modelo tienen distribución normal.

Diseño en Bloques Completamente Aleatorizados

Supuesto de Homogeneidad de Varianza

Prueba de Bartlet:

```
residuales<-residuals(aov(y~rollo+agente))
bartlett.test(residuales,rollo)
bartlett.test(residuales,agente)
######################################################
Bartlett test of homogeneity of variances
data:  residuales and rollo
Bartlett's K-squared = 0.65699, df = 4, p-value = 0.9565
######################################################
Bartlett test of homogeneity of variances
data:  residuales and agente
Bartlett's K-squared = 2.6757, df = 3, p-value = 0.4444
```