Diseño de Experimentos I

Diseños completamente aleatorizados DCA

UTB

Universidad Tecnologica de Bolívar

Posgrado UTB 2020

Competencias

-  *Identificar dentro de la familia de los diseños experimentales con un sólo factor, los diseños utilizados para comparar tratamientos y partiendo de las hipótesis de estudio, construir sus modelos estadísticos y el analisis de varianza para probar las hipotesis respectivas y validar el modelo.*

Competencias

-  *Identificar dentro de la familia de los diseños experimentales con un sólo factor, los diseños utilizados para comparar tratamientos y partiendo de las hipótesis de estudio, construir sus modelos estadísticos y el analisis de varianza para probar las hipotesis respectivas y validar el modelo.*
-  *Diferenciar las pruebas de rangos múltiples en diseños de experimentos con un sólo factor, identificando la prueba más robusta en cada diseño experimental.*

Competencias

-  *Identificar dentro de la familia de los diseños experimentales con un sólo factor, los diseños utilizados para comparar tratamientos y partiendo de las hipótesis de estudio, construir sus modelos estadísticos y el analisis de varianza para probar las hipotesis respectivas y validar el modelo.*
-  *Diferenciar las pruebas de rangos múltiples en diseños de experimentos con un sólo factor, identificando la prueba más robusta en cada diseño experimental.*
-  *Verificar los supuestos del modelo estadístico asociado a un diseño de experimento validando los resultados obtenidos.*

Ejemplo de un diseño completamente al azar

- **Problema de aplicación**. *Un ingeniero de desarrollo de productos tiene interés en investigar la Resistencia a la tensión de una fibra sintética nueva que se usará para hacer tela de camisas para caballeros. El ingeniero sabe por experiencia previa que la resistencia a la tensión se afecta por el peso porcentual del algodón utilizado en la mezcla de materiales de la fibra. Además, sospecha que al aumentar el contenido de algodón se incrementará la resistencia, al menos en un principio. Sabe asimismo que el contenido de algodón deberá variar entre 10% y 40% para que el producto final tenga otras características de calidad que se desean (como la capacidad de ser sometido a un tratamiento de planchado permanente). El ingeniero decide probar ejemplares en cinco niveles del peso porcentual del algodón: 15%, 20%, 25%, 30% y 35%. También decide probar cinco ejemplares en cada nivel del contenido del algodón.*

Ejemplo de un diseño completamente al azar

- Factor de interés: *Peso porcentual de algodón en la fibra*.

Ejemplo de un diseño completamente al azar

- Factor de interés: *Peso porcentual de algodón en la fibra*.
- Niveles del Factor: 15%, 20%, 25%, 30% y 35% (cinco niveles $k = 5$).

Ejemplo de un diseño completamente al azar

- Factor de interés: *Peso porcentual de algodón en la fibra*.
- Niveles del Factor: 15%, 20%, 25%, 30% y 35% (cinco niveles $k = 5$).
- Variable de interés: $Y =$ Resistencia a la tensión de la fibra.

Ejemplo de un diseño completamente al azar

- Factor de interés: *Peso porcentual de algodón en la fibra*.
- Niveles del Factor: 15%, 20%, 25%, 30% y 35% (cinco niveles $k = 5$).
- Variable de interés: $Y =$ Resistencia a la tensión de la fibra.
- Replicas por nivel: $n = 5$.

Ejemplo de un diseño completamente al azar

- Factor de interés: *Peso porcentual de algodón en la fibra.*
- Niveles del Factor: 15%, 20%, 25%, 30% y 35% (cinco niveles $k = 5$).
- Variable de interés: $Y =$ Resistencia a la tensión de la fibra.
- Replicas por nivel: $n = 5$.
- Procedimiento inicial: aleatorizar el orden de las corridas:

Peso %	corridas				
15	1	2	3	4	5
20	6	7	8	9	10
25	11	12	13	14	15
30	16	17	18	19	20
35	21	22	23	24	25

Se selecciona un número entre 1 y 25. Suponga que este número es 12. Entonces, la observación número 12 (25%) se corre primero y así sucesivamente hasta obtener las 25 observaciones.

Ejemplo de un diseño completamente al azar

- Suponga que la secuencia de la prueba es:

Secuencia	corrida	Peso %	Secuencia	corrida	Peso %
1	22	20	14	7	10
2	18	30	15	1	15
3	10	20	16	24	35
4	23	35	17	21	35
5	17	30	18	11	25
6	5	15	19	2	15
7	14	25	20	13	25
8	6	20	21	22	35
9	15	25	22	16	30
10	20	30	23	25	35
11	9	20	24	19	30
12	4	15	25	3	15
13	12	25			

- Datos Y $(lb/pul\hat{\ }2)$:

Peso %	Y_{ij}				
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

Ejemplo de un diseño completamente al azar

- Modelo estadístico:

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

donde $Y_{ij} :=$ es la resistencia a la tensión de la j-ésima fibra sintética con i-ésimo porcentaje de algodón, $\tau_i :=$ es la medida del efecto a la tensión debida al i-ésimo porcentaje de algodon, $\varepsilon_{ij} :=$ es el error aleatorio y $\mu :=$ es la media global real de todos los tratamientos.

Ejemplo de un diseño completamente al azar

- Modelo estadístico:

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

donde $Y_{ij} :=$ es la resistencia a la tensión de la j-ésima fibra sintética con i-ésimo porcentaje de algodón, $\tau_i :=$ es la medida del efecto a la tensión debida al i-ésimo porcentaje de algodon, $\varepsilon_{ij} :=$ es el error aleatorio y $\mu :=$ es la media global real de todos los tratamientos.

- Hipótesis del problema:

$$\begin{array}{ll} H_0 : & \mu_{15\%} = \mu_{20\%} = \mu_{25\%} = \mu_{30\%} = \mu_{35\%} = \mu \\ H_1 : & \mu_i \neq \mu_j \text{ para algunos } i,\ j \end{array}$$

Ejemplo de un diseño completamente al azar

- Modelo estadístico:

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

  donde $Y_{ij} :=$ es la resistencia a la tensión de la j-ésima fibra sintética con i-ésimo porcentaje de algodón, $\tau_i :=$ es la medida del efecto a la tensión debida al i-ésimo porcentaje de algodon, $\varepsilon_{ij} :=$ es el error aleatorio y $\mu :=$ es la media global real de todos los tratamientos.

- Hipótesis del problema:

$$\begin{array}{ll} H_0: & \mu_{15\%} = \mu_{20\%} = \mu_{25\%} = \mu_{30\%} = \mu_{35\%} = \mu \\ H_1: & \mu_i \neq \mu_j \text{ para algunos } i,\ j \end{array}$$

- Significancia de la Prueba: $\alpha = 0.05$.

Ejemplo de un diseño completamente al azar

- Resumen de los datos:

Peso %	Y_{ij}					$Y_{i.}$	$\overline{Y}_{i.}$
15	7	7	15	11	9	49	9.8
20	12	17	12	18	18	77	15.4
25	14	18	18	19	19	88	17.6
30	19	25	22	19	23	108	21.6
35	7	10	11	15	11	54	10.8
						$Y_{..} = 376$	$\overline{Y}_{..} = 15.04$

Ejemplo de un diseño completamente al azar

- Suma de cuadrados:

Ejemplo de un diseño completamente al azar

-  Suma de cuadrados:
  - $SCTr = \frac{1}{5}\sum_{i=1}^{5} Y_{i\cdot}^2 - \frac{Y_{\cdot\cdot}^2}{N} = 475,76$

Ejemplo de un diseño completamente al azar

- Suma de cuadrados:
  - $SCTr = \frac{1}{5}\sum_{i=1}^{5} Y_{i\cdot}^{2} - \frac{Y_{\cdot\cdot}^{2}}{N} = 475,76$
  - $SCT = \sum_{i=1}^{5}\sum_{j=1}^{5} Y_{ij}^{2} - \frac{Y_{\cdot\cdot}^{2}}{N} = 636,96$

Ejemplo de un diseño completamente al azar

- Suma de cuadrados:
  - $SCTr = \frac{1}{5}\sum_{i=1}^{5} Y_{i\cdot}^2 - \frac{Y_{\cdot\cdot}^2}{N} = 475,76$
  - $SCT = \sum_{i=1}^{5}\sum_{j=1}^{5} Y_{ij}^2 - \frac{Y_{\cdot\cdot}^2}{N} = 636,96$
  - $SCE = SCT - SCTr = 161,2$

- Tabla ANOVA para Resistencia según Porcentaje de Algodon:

Fuente	SC	Gl	CM	Cociente-F	P-Valor
Tratamientos	475,76	4	118,94	14,76	0,0000
Error	161,2	20	8,06		
Total (Corr.)	636,96	24			

- Tabla ANOVA para Resistencia según Porcentaje de Algodon:

Fuente	SC	Gl	CM	Cociente-F	P-Valor
Tratamientos	475,76	4	118,94	14,76	0,0000
Error	161,2	20	8,06		
Total (Corr.)	636,96	24			

- La tabla ANOVA descompone la varianza de Resistencia en dos componentes:

- Tabla ANOVA para Resistencia según Porcentaje de Algodon:

Fuente	SC	Gl	CM	Cociente-F	P-Valor
Tratamientos	475,76	4	118,94	14,76	0,0000
Error	161,2	20	8,06		
Total (Corr.)	636,96	24			

- La tabla ANOVA descompone la varianza de Resistencia en dos componentes:
  - *un componente entre grupos (Tratamientos),* y

Ejemplo de un diseño completamente al azar

Tabla Anova

- Tabla ANOVA para Resistencia según Porcentaje de Algodon:

Fuente	SC	Gl	CM	Cociente-F	P-Valor
Tratamientos	475,76	4	118,94	14,76	0,0000
Error	161,2	20	8,06		
Total (Corr.)	636,96	24			

- La tabla ANOVA descompone la varianza de Resistencia en dos componentes:
  - *un componente entre grupos (Tratamientos)*, y
  - *un componente dentro de los grupos (Error)*.

- Tabla ANOVA para Resistencia según Porcentaje de Algodon:

Fuente	SC	Gl	CM	Cociente-F	P-Valor
Tratamientos	475,76	4	118,94	14,76	0,0000
Error	161,2	20	8,06		
Total (Corr.)	636,96	24			

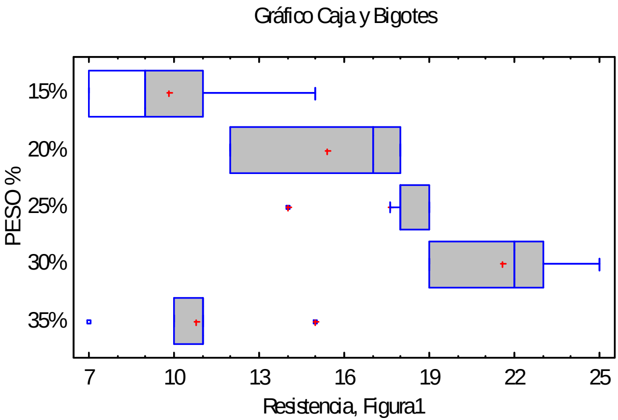
- La tabla ANOVA descompone la varianza de Resistencia en dos componentes:
  - *un componente entre grupos (Tratamientos)*, y
  - *un componente dentro de los grupos (Error)*.
- El Cociente F es igual a $14,76$ que resulta ser un valor alto y que tiene un p-valor inferior a 0.05

- Tabla ANOVA para Resistencia según Porcentaje de Algodon:

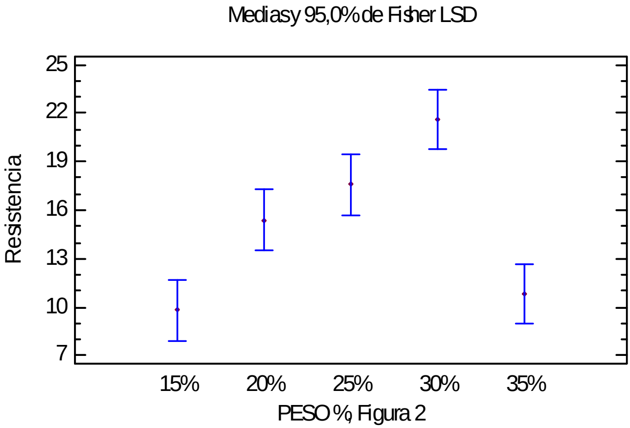
Fuente	SC	Gl	CM	Cociente-F	P-Valor
Tratamientos	475,76	4	118,94	14,76	0,0000
Error	161,2	20	8,06		
Total (Corr.)	636,96	24			

- La tabla ANOVA descompone la varianza de Resistencia en dos componentes:
  - *un componente entre grupos (Tratamientos)*, y
  - *un componente dentro de los grupos (Error)*.
- El Cociente F es igual a $14,76$ que resulta ser un valor alto y que tiene un p-valor inferior a 0.05
- Esto prueba que hay diferencia estadísticamente significativa en la resistencia media entre los diferentes niveles de algodón con una confianza de 95%.

- Diagramas de caja para los porcentajes de algodón (figura1)



- Grafico de medias (figura2)



Ejemplo de un diseño completamente al azar

Análisis gráficos

- Grafico de medias (figura2)

- Grafico de medias (figura2)
- Ambas gráficas indican que la resistencia a la tensión se incrementa hasta cerca de 30% de algodón, después de 30%, hay un marcado descenso a la tensión.

Ejemplo de un diseño completamente al azar
Análisis gráficos

- Grafico de medias (figura2)
- Ambas gráficas indican que la resistencia a la tensión se incrementa hasta cerca de 30% de algodón, después de 30%, hay un marcado descenso a la tensión.
- Con base en este análisis se tienen firmes sospechas de que:

- Grafico de medias (figura2)
- Ambas gráficas indican que la resistencia a la tensión se incrementa hasta cerca de 30% de algodón, después de 30%, hay un marcado descenso a la tensión.
- Con base en este análisis se tienen firmes sospechas de que:
  - El contenido de algodón afecta la resistencia a la tensión y

- Grafico de medias (figura2)
- Ambas gráficas indican que la resistencia a la tensión se incrementa hasta cerca de 30% de algodón, después de 30%, hay un marcado descenso a la tensión.
- Con base en este análisis se tienen firmes sospechas de que:
  - El contenido de algodón afecta la resistencia a la tensión y
  - alrededor de 30% de algodón produce la resistencia máxima.

Ejemplo de un diseño completamente al azar

- Haciendo uso del método de *comparaciones múltiples de Fisher* o de mínima diferencias significativas o LSD nos muestra:

Ejemplo de un diseño completamente al azar

- Haciendo uso del método de *comparaciones múltiples de Fisher* o de mínima diferencias significativas o LSD nos muestra:
- $$LSD = t_{\alpha/2, N-k} \cdot \sqrt{\frac{2}{n} CME} = t_{0.025, 30-5} \cdot \sqrt{\frac{2}{6} CME} = 3,74546$$

Ejemplo de un diseño completamente al azar

- Haciendo uso del método de *comparaciones múltiples de Fisher* o de mínima diferencias significativas o LSD nos muestra:

- $$LSD = t_{\alpha/2,N-k} \cdot \sqrt{\frac{2}{n}CME} = t_{0.025,30-5} \cdot \sqrt{\frac{2}{6}CME} = 3,74546$$

-

Nivel	Frec	Media	Grupo		
15%	5	9,8	X		
35%	5	10,8	X		
20%	5	15,4		X	
25%	5	17,6		X	
30%	5	21,6			X

Ejemplo de un diseño completamente al azar

- Estimaciones de los parámetros del modelo:

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

Ejemplo de un diseño completamente al azar

- Estimaciones de los parámetros del modelo:

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

- Los estimadores de mínimos cuadrados de los parámetros μ y τ_i son:

Ejemplo de un diseño completamente al azar

- Estimaciones de los parámetros del modelo:

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

- Los estimadores de mínimos cuadrados de los parámetros μ y τ_i son:
  - $\widehat{\mu} = \overline{Y}_{..}$

Ejemplo de un diseño completamente al azar

- Estimaciones de los parámetros del modelo:

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

- Los estimadores de mínimos cuadrados de los parámetros μ y τ_i son:
  - $\widehat{\mu} = \overline{Y}_{..}$
  - $\widehat{\mu}_i = \overline{Y}_{i.}$

Ejemplo de un diseño completamente al azar

- Estimaciones de los parámetros del modelo:

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

- Los estimadores de mínimos cuadrados de los parámetros μ y τ_i son:
  - $\widehat{\mu} = \overline{Y}_{..}$
  - $\widehat{\mu}_i = \overline{Y}_{i.}$
  - $\widehat{\tau}_i = \overline{Y}_{i.} - \overline{Y}_{..}$

Ejemplo de un diseño completamente al azar

- Estimaciones de los parámetros del modelo:

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

- Los estimadores de mínimos cuadrados de los parámetros μ y τ_i son:
  - $\widehat{\mu} = \overline{Y}_{..}$
  - $\widehat{\mu}_i = \overline{Y}_{i.}$
  - $\widehat{\tau}_i = \overline{Y}_{i.} - \overline{Y}_{..}$
- Intervalos de confianza $(1-\alpha) \times 100\%$, usando el método LSD se obtienen:

Ejemplo de un diseño completamente al azar

- Estimaciones de los parámetros del modelo:

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

- Los estimadores de mínimos cuadrados de los parámetros μ y τ_i son:
  - $\widehat{\mu} = \overline{Y}_{..}$
  - $\widehat{\mu}_i = \overline{Y}_{i.}$
  - $\widehat{\tau}_i = \overline{Y}_{i.} - \overline{Y}_{..}$

- Intervalos de confianza $(1-\alpha)\times 100\%$, usando el método LSD se obtienen:
  - $IC(\mu_i) = \overline{Y}_{i.} \pm t_{\alpha/2,N-k} \cdot \sqrt{\frac{CME}{n}}$

Ejemplo de un diseño completamente al azar

- Estimaciones de los parámetros del modelo:

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

- Los estimadores de mínimos cuadrados de los parámetros μ y τ_i son:
  - $\widehat{\mu} = \overline{Y}_{..}$
  - $\widehat{\mu}_i = \overline{Y}_{i.}$
  - $\widehat{\tau}_i = \overline{Y}_{i.} - \overline{Y}_{..}$

- Intervalos de confianza $(1-\alpha) \times 100\%$, usando el método LSD se obtienen:
  - $IC(\mu_i) = \overline{Y}_{i.} \pm t_{\alpha/2,N-k} \cdot \sqrt{\frac{CME}{n}}$
  - $IC(\mu_i - \mu_j) = \left(\overline{Y}_{i.} - \overline{Y}_{j.}\right) \pm t_{\alpha/2,N-k} \cdot \sqrt{2\frac{CME}{n}}$

Ejemplo de un diseño completamente al azar

- Del ejemplo:

$$
\begin{gathered}
\widehat{\mu} = \overline{Y}_{..} = 15,04 \\
\widehat{\tau}_1 = 9,80 - 15,04 = -5,24 \\
\widehat{\tau}_2 = 15,40 - 15,04 = 0,36 \\
\widehat{\tau}_3 = 17,60 - 15,04 = -2,56 \\
\widehat{\tau}_4 = 21,60 - 15,04 = 6,56 \\
\widehat{\tau}_5 = 10,80 - 15,04 = -4,24
\end{gathered}
$$

Ejemplo de un diseño completamente al azar

- Del ejemplo:

$$\widehat{\mu} = \overline{Y}_{..} = 15,04$$
$$\widehat{\tau}_1 = 9,80 - 15,04 = -5,24$$
$$\widehat{\tau}_2 = 15,40 - 15,04 = 0,36$$
$$\widehat{\tau}_3 = 17,60 - 15,04 = -2,56$$
$$\widehat{\tau}_4 = 21,60 - 15,04 = 6,56$$
$$\widehat{\tau}_5 = 10,80 - 15,04 = -4,24$$

- Intervalos

$$IC(\mu_4) = 21,60 \pm 2,65 \tag{1}$$

- Tabla de Medias para Resistencia según Porcentaje de Algodon con 95,0 intervalos de Bonferroni

Nivel	Frec	Media	Error estandard	Limite inf	Limite sup
15%	5	9, 8	1, 26965	6, 96895	12, 6311
35%	5	10, 8	1, 26965	12, 5689	18, 2311
20%	5	15, 4	1, 26965	14, 7689	20, 4311
25%	5	17, 6	1, 26965	18, 7689	24, 4311
30%	5	21, 6	1, 26965	7, 96895	13, 6311

Verificación de la adecuación del modelo

- Modelo
$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$
donde ε_{ij} son variables aleatorias independientes y ademas $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Verificación de la adecuación del modelo

- Modelo
$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$
donde ε_{ij} son variables aleatorias independientes y ademas $\varepsilon_{ij} \sim N(0, \sigma^2)$.
- Si éstos supuestos se satisfacen, el procedimiento de Análisis de varianza es una prueba exacta de la hipótesis de que no hay diferencias en la medias de los tratamientos.

Verificación de la adecuación del modelo

- Modelo
$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$
donde ε_{ij} son variables aleatorias independientes y ademas $\varepsilon_{ij} \sim N(0, \sigma^2)$.
- Si éstos supuestos se satisfacen, el procedimiento de Análisis de varianza es una prueba exacta de la hipótesis de que no hay diferencias en la medias de los tratamientos.
- Sinembargo, es comun que en la práctica esto supuestos no se satisfagan exactamente.

Verificación de la adecuación del modelo

- Modelo
$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$
donde ε_{ij} son variables aleatorias independientes y ademas $\varepsilon_{ij} \sim N(0, \sigma^2)$.
- Si éstos supuestos se satisfacen, el procedimiento de Análisis de varianza es una prueba exacta de la hipótesis de que no hay diferencias en la medias de los tratamientos.
- Sinembargo, es comun que en la práctica esto supuestos no se satisfagan exactamente.
- Por consiguiente, en general no es prudente confiar en el ANOVA antes de haber verificado éstos supuestos.

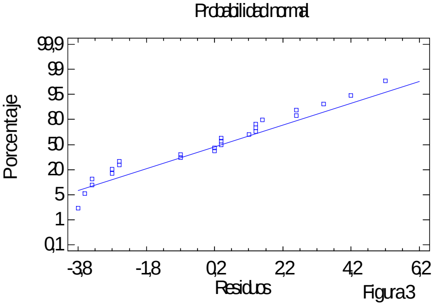


Figura 3

- En la gráfica de probabilidad normal los residuales tienen distribución normal si la gráfica tendrá la apariencia de una línea recta.

Verificación de la adecuación del modelo

Supuesto de normalidad

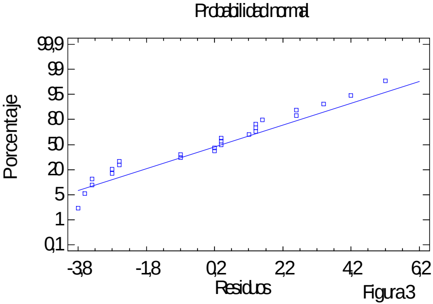
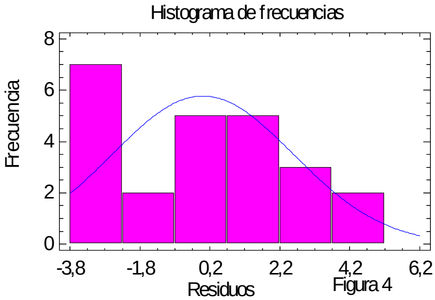


Figura 3

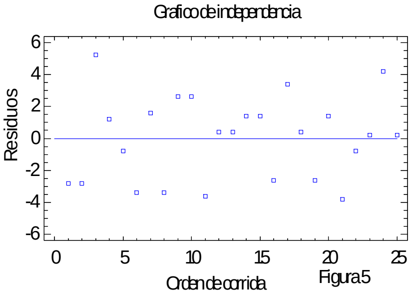
- En la gráfica de probabilidad normal los residuales tienen distribución normal si la gráfica tendrá la apariencia de una línea recta.
- Para visualizar la línea recta deberá prestarse más atención a los valores centrales de la gráfica que a valores extremos.

Histograma para Residuales (Figura 4)



- La distribución de los errores puede tener un ligero sesgo, con la cola derecha siendo mas larga que la cola izquierda. Esta gráfica no muestra una desviación marcada de la distibución normal.

Grafica de los residuales en secuencia del tiempo (Figura 5)



- No hay razón para sospechar cualquier violación de los supuestos de independencia o de una varianza constante.

- La graficación de los residuales en el orden temporal de la recolección de los datos es útil para detectar correlaciones entre los residuales.

- La graficación de los residuales en el orden temporal de la recolección de los datos es útil para detectar correlaciones entre los residuales.
  - Una tendencia a tener corridas de residuales positivos y negativos indica una correlación positiva.

- La graficación de los residuales en el orden temporal de la recolección de los datos es útil para detectar correlaciones entre los residuales.
    - Una tendencia a tener corridas de residuales positivos y negativos indica una correlación positiva.
    - Esto implicaría que el supuesto de independencia de los errores ha sido violado.

- La graficación de los residuales en el orden temporal de la recolección de los datos es útil para detectar correlaciones entre los residuales.
    - Una tendencia a tener corridas de residuales positivos y negativos indica una correlación positiva.
    - Esto implicaría que el supuesto de independencia de los errores ha sido violado.
    - Se trata de un problema potencialmente serio y cuya solución es difícil, por lo que de ser posible es importante evitar el problema cuando se colecten los datos.

- La graficación de los residuales en el orden temporal de la recolección de los datos es útil para detectar correlaciones entre los residuales.
    - Una tendencia a tener corridas de residuales positivos y negativos indica una correlación positiva.
    - Esto implicaría que el supuesto de independencia de los errores ha sido violado.
    - Se trata de un problema potencialmente serio y cuya solución es difícil, por lo que de ser posible es importante evitar el problema cuando se colecten los datos.
    - La aleatorización adecuada del experimento es un paso importante para conseguir la independencia.

Gráfica de los residuales contra los valores ajustados:

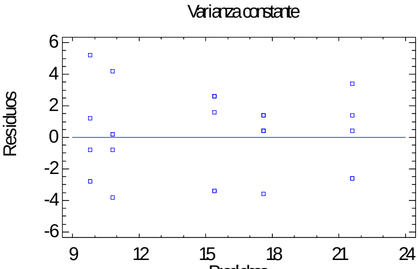


Figura6

- Esta gráfica no muestra algún patrón obvio. No es evidente ninguna estructura inusual.

- **Prueba de Bartlett.**

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$$
$$H_1 : \sigma_i^2 = \sigma_j^2, \text{ algunos } i, \ j$$

Verificación de la adecuación del modelo

Supuesto de varianza constante

- **Prueba de Bartlett.**

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$$
$$H_1 : \sigma_i^2 = \sigma_j^2, \text{ algunos } i,\ j$$

  - Nivel de significancia $\alpha = 0,05$

- **Prueba de Bartlett.**

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$$
$$H_1 : \sigma_i^2 = \sigma_j^2, \text{ algunos } i, \ j$$

  - Nivel de significancia $\alpha = 0,05$
  - Estadístico de prueba

Verificación de la adecuación del modelo

Supuesto de varianza constante

- **Prueba de Bartlett.**

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$$
$$H_1 : \sigma_i^2 = \sigma_j^2, \text{ algunos } i,\ j$$

  - Nivel de significancia $\alpha = 0,05$
  - Estadístico de prueba
  - $\chi_0^2 = 2,3026 \times q \diagup c$

Verificación de la adecuación del modelo

Supuesto de varianza constante

- **Prueba de Bartlett.**

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$$
$$H_1 : \sigma_i^2 = \sigma_j^2, \text{ algunos } i,\ j$$

  - Nivel de significancia $\alpha = 0,05$
  - Estadístico de prueba
  - $\chi_0^2 = 2,3026 \times q \diagup c$
  - $q = (N - k)\log_{10} S_p^2 - \sum_{i=1}^{k} (n_i - 1)\log_{10} S_i^2$

Verificación de la adecuación del modelo

Supuesto de varianza constante

- **Prueba de Bartlett.**

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$$
$$H_1 : \sigma_i^2 = \sigma_j^2, \text{ algunos } i,\ j$$

- Nivel de significancia $\alpha = 0,05$
- Estadístico de prueba
- $\chi_0^2 = 2,3026 \times q \big/ c$
- $q = (N-k)\log_{10} S_p^2 - \sum_{i=1}^{k}(n_i - 1)\log_{10} S_i^2$
- $S_p^2 = \sum_{i=1}^{k}(n_i - 1)S_i^2 \big/ (N-k)$

Verificación de la adecuación del modelo

Supuesto de varianza constante

- **Prueba de Bartlett.**

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$$
$$H_1 : \sigma_i^2 = \sigma_j^2, \text{ algunos } i, \ j$$

- Nivel de significancia $\alpha = 0,05$
- Estadístico de prueba
- $\chi_0^2 = 2,3026 \times q \diagup c$
- $q = (N-k)\log_{10} S_p^2 - \sum_{i=1}^{k} (n_i - 1)\log_{10} S_i^2$
- $S_p^2 = \sum_{i=1}^{k} (n_i - 1) S_i^2 \diagup (N-k)$
- $c = 1 + \left(\sum_{i=1}^{k} (n_i - 1)^{-1} - (N-k)^{-1} \right) \diagup 3(k-1)$

- donde S_i^2 es la varianza muestral de la población $i-$ésima.

- donde S_i^2 es la varianza muestral de la población $i-$ésima.
- De los datos:

- donde S_i^2 es la varianza muestral de la población $i-$ésima.
- De los datos:
  - $S_p^2 = 8.06$

- donde S_i^2 es la varianza muestral de la población $i-$ésima.
- De los datos:
  - $S_p^2 = 8.06$
  - $q = 0.45$

- donde S_i^2 es la varianza muestral de la población $i-$ésima.
- De los datos:
  - $S_p^2 = 8.06$
  - $q = 0.45$
  - $c = 1.1$

- donde S_i^2 es la varianza muestral de la población $i-$ésima.
- De los datos:
  - $S_p^2 = 8.06$
  - $q = 0.45$
  - $c = 1.1$
  - $\chi_0^2 = 2,3026 \times 0.45 \diagup 1.1 = 0.93$

- donde S_i^2 es la varianza muestral de la población $i-$ésima.
- De los datos:
  - $S_p^2 = 8.06$
  - $q = 0.45$
  - $c = 1.1$
  - $\chi_0^2 = 2,3026 \times 0.45 \diagup 1.1 = 0.93$
  - comparando con
$$\chi^2_{\alpha,k-1} = \chi^2_{0.05,4} = 9.3$$

- donde S_i^2 es la varianza muestral de la población $i-$ésima.
- De los datos:
  - $S_p^2 = 8.06$
  - $q = 0.45$
  - $c = 1.1$
  - $\chi_0^2 = 2,3026 \times 0.45 \diagup 1.1 = 0.93$
  - comparando con
$$\chi^2_{\alpha,k-1} = \chi^2_{0.05,4} = 9.3$$
  - no se rechaza la hipótesis nula y se concluye que las cinco varianzas son iguales.

- **Prueba estadística para la igualdad de varianzas usando el Statgraphics**

- **Prueba estadística para la igualdad de varianzas usando el Statgraphics**
- Contraste C de Cochran: $0,277916$, cuyo *valor* P$= 1,0$
  Contraste de Bartlett: $1,05266$, cuyo *valor* P$= 0,919766$
  Test de Levene: $0,317949$, cuyo *valor* P$= 0,862586$

- **Prueba estadística para la igualdad de varianzas usando el Statgraphics**
- Contraste C de Cochran: $0,277916$, cuyo *valor* P$= 1,0$
  Contraste de Bartlett: $1,05266$, cuyo *valor* P$= 0,919766$
  Test de Levene: $0,317949$, cuyo *valor* P$= 0,862586$
- El tercer estadístico mostrado en esta tabla, comprueba la hipótesis nula de que la desviación típica de Resistencia dentro de cada uno de los 5 niveles de Porcentaje de Algodon, es la misma.

Prueba de Shapiro-Wilks para la normalidad

- Hipotesis

$$
\begin{aligned}
H_0 &: \text{Los datos proceden de una distribución normal} \\
H_1 &: \text{Los datos no proceden de una distribución normal}
\end{aligned}
$$

Prueba de Shapiro-Wilks para la normalidad

- Hipotesis

$$
\begin{aligned}
H_0 &: \text{Los datos proceden de una distribución normal}\\
H_1 &: \text{Los datos no proceden de una distribución normal}
\end{aligned}
$$

- Ordenamos los datos de los residuales $X_{(1)}$ =-3,8, $X_{(2)}$ =-3,6, $X_{(3)}$ =-3,4, $X_{(4)}$ =-3,4, $X_{(5)}$ =-2,8, $X_{(6)}$ =-2,8, $X_{(7)}$ =-2,6, $X_{(8)}$ =-2,6, $X_{(9)}$ =-0,8, $X_{(10)}$ =-0,8, $X_{(11)}$ =0,2, $X_{(12)}$ =0,2, $X_{(13)}$ =0,4, $X_{(14)}$ =0,4, $X_{(15)}$ =0,4, $X_{(16)}$ =1,2, $X_{(17)}$ =1,4, $X_{(18)}$ =1,4, $X_{(19)}$ =1,4, $X_{(20)}$ =1,6, $X_{(21)}$ =2,6, $X_{(22)}$ =2,6, $X_{(23)}$ =3,4, $X_{(24)}$ =4,2, $X_{(25)}$ =5,2

Prueba de Shapiro-Wilks para la normalidad

i	$a_{i,25}$	$X_{(25-i+1)} - X_{(i)}$	$a_{(i),n} \times \left(X_{(25-i+1)} - X_{(i)}\right)$
1	0.4450	9	4.005
2	0.3069	7.8	2.39382
3	0.2543	6.8	1.72924
4	0.2148	6.0	1.2888
5	0.1822	5.4	0.98388
6	0.1539	4.4	0.67716
7	0.1283	4.0	0.5132
8	0.1046	4.0	0.4184
9	0.0823	2.2	0.18106
10	0.0610	2.0	0.122
11	0.0403	0.2	0.00806
12	0.0200	0.2	0.004
13	0.000	0	0
		Total	12.32462

Prueba de Shapiro-Wilks para la normalidad

- Se calcula el estadístico W definido por

$$W = \frac{1}{(n-1)\,S^2}\left[\sum_{i=1}^{k} a_{(i),n} \times \left(X_{(n-i+1)} - X_{(i)}\right)\right]^2$$

donde S^2 es la varianza muestral, los coeficientes a_i, se obtienen de la tabla Shapiro-Wilks.

Prueba de Shapiro-Wilks para la normalidad

- Se calcula el estadístico W definido por

$$W = \frac{1}{(n-1)\,S^2}\left[\sum_{i=1}^{k} a_{(i),n} \times \left(X_{(n-i+1)} - X_{(i)}\right)\right]^2$$

donde S^2 es la varianza muestral, los coeficientes a_i, se obtienen de la tabla Shapiro-Wilks.

- De manera que

$$\begin{aligned} W &= \frac{1}{(n-1)\,S^2}\left[\sum_{i=1}^{13} a_{(i),25} \times \left(X_{(25-i+1)} - X_{(i)}\right)\right]^2 \\ &= \frac{1}{24 \times 6.71667}\left[12.32462\right]^2 = 0.9422 \end{aligned}$$

Prueba de Shapiro-Wilks para la normalidad

- Otra expresión para el cálculo de W viene dado por

Prueba de Shapiro-Wilks para la normalidad

- Otra expresión para el cálculo de W viene dado por
- $W = \left[\sum_{i=1}^{13} a_{(i),25} \times \left(X_{(25-i+1)} - X_{(i)} \right) \right]^2 \diagup \sum_{i=1}^{13} X_{(i)}^2 = [12.32462]^2 \diagup 161,2 = 0.9422$

Prueba de Shapiro-Wilks para la normalidad

- Otra expresión para el cálculo de W viene dado por
- $W = \left[\sum_{i=1}^{13} a_{(i),25} \times \left(X_{(25-i+1)} - X_{(i)}\right)\right]^2 \Big/ \sum_{i=1}^{13} X_{(i)}^2 = [12.32462]^2 \big/ 161,2 = 0.9422$
  - donde $\sum_{i=1}^{13} X_{(i)}^2 = (-3,8)^2 + (-3,6)^2 + (-3,4)^2 + (-3,4)^2 + \cdots + (4,2)^2 + (5,2)^2 = 161,2$

Prueba de Shapiro-Wilks para la normalidad

- Otra expresión para el cálculo de W viene dado por
- $W = \left[\sum_{i=1}^{13} a_{(i),25} \times \left(X_{(25-i+1)} - X_{(i)}\right)\right]^2 \Big/ \sum_{i=1}^{13} X_{(i)}^2 = [12.32462]^2 \big/ 161,2 = 0.9422$
  - donde $\sum_{i=1}^{13} X_{(i)}^2 = (-3,8)^2 + (-3,6)^2 + (-3,4)^2 + (-3,4)^2 + \cdots + (4,2)^2 + (5,2)^2 = 161,2$
- Finalmente, al usar la tabla de valores críticos Shapiro Wilks al nivel $\alpha = 0.05$ seleccionado, se observa que $W_{1-\alpha,n} = W_{0.95;25} = 0.918 < 0.9422$.

Prueba de Shapiro-Wilks para la normalidad

- Otra expresión para el cálculo de W viene dado por
- $W = \left[\sum_{i=1}^{13} a_{(i),25} \times \left(X_{(25-i+1)} - X_{(i)} \right) \right]^2 \Big/ \sum_{i=1}^{13} X_{(i)}^2 = [12.32462]^2 \big/ 161,2 = 0.9422$
  - donde $\sum_{i=1}^{13} X_{(i)}^2 = (-3,8)^2 + (-3,6)^2 + (-3,4)^2 + (-3,4)^2 + \cdots + (4,2)^2 + (5,2)^2 = 161,2$
- Finalmente, al usar la tabla de valores críticos Shapiro Wilks al nivel $\alpha = 0.05$ seleccionado, se observa que $W_{1-\alpha,n} = W_{0.95;25} = 0.918 < 0.9422$.
- Tests para la Normalidad, usando Statgraphics. Estadístico W de $Shapiro - Wilks = 0,942656$. P-valor $= 0{,}179142$.

Modelo de efectos aleatorios con un sólo factor

- Los modelos lineales estudiados hasta ahora corresponden a *Modelos de Efectos Fijos*, ya que los niveles del factor en estudio eran los únicos de interés para el investigador, donde el experimento se podía repetir con los mismos tratamientos y las inferencias obtenidas son referidas al mismo conjunto de tratamientos estudiados.

Modelo de efectos aleatorios con un sólo factor

- Los modelos lineales estudiados hasta ahora corresponden a *Modelos de Efectos Fijos*, ya que los niveles del factor en estudio eran los únicos de interés para el investigador, donde el experimento se podía repetir con los mismos tratamientos y las inferencias obtenidas son referidas al mismo conjunto de tratamientos estudiados.
- Frecuentemente, al investigador se interesa por un Factor que tiene un gran número de posibles niveles.

Modelo de efectos aleatorios con un sólo factor

- Los modelos lineales estudiados hasta ahora corresponden a *Modelos de Efectos Fijos*, ya que los niveles del factor en estudio eran los únicos de interés para el investigador, donde el experimento se podía repetir con los mismos tratamientos y las inferencias obtenidas son referidas al mismo conjunto de tratamientos estudiados.
- Frecuentemente, al investigador se interesa por un Factor que tiene un gran número de posibles niveles.
- Se dice que éste *Factor es aleatorio* si los niveles que se estudian son seleccionados aleatoriamente de la población.

Modelo de efectos aleatorios con un sólo factor

- Los modelos lineales estudiados hasta ahora corresponden a *Modelos de Efectos Fijos*, ya que los niveles del factor en estudio eran los únicos de interés para el investigador, donde el experimento se podía repetir con los mismos tratamientos y las inferencias obtenidas son referidas al mismo conjunto de tratamientos estudiados.
- Frecuentemente, al investigador se interesa por un Factor que tiene un gran número de posibles niveles.
- Se dice que éste *Factor es aleatorio* si los niveles que se estudian son seleccionados aleatoriamente de la población.
- Las conclusiones se obtienen para toda la población de niveles del factor; este tipo de diseño se llama *Modelo de Efectos Aleatorios.*

Modelo de efectos aleatorios con un sólo factor

- Los modelos lineales estudiados hasta ahora corresponden a *Modelos de Efectos Fijos*, ya que los niveles del factor en estudio eran los únicos de interés para el investigador, donde el experimento se podía repetir con los mismos tratamientos y las inferencias obtenidas son referidas al mismo conjunto de tratamientos estudiados.
- Frecuentemente, al investigador se interesa por un Factor que tiene un gran número de posibles niveles.
- Se dice que éste *Factor es aleatorio* si los niveles que se estudian son seleccionados aleatoriamente de la población.
- Las conclusiones se obtienen para toda la población de niveles del factor; este tipo de diseño se llama *Modelo de Efectos Aleatorios.*
- Para el *Modelo de Efectos Aleatorios* una repetición del experimento producirá un nuevo conjunto de tratamientos de la misma población y por lo tanto, el interés del investigador estará en la variabilidad de los tratamientos.

Modelo de efectos aleatorios con un sólo factor

- Supongamos que tenemos un factor A con k niveles, con n repeticiones por tratamiento. Se mide la variable respuesta, la cual puede describirse mediante el siguiente modelo de efectos:

$$Y_{ij} = \mu + A_i + \varepsilon_{ij};\ i = 1, ..., k;\ j = 1, ..., n$$

donde A_i y ε_{ij} son variables aleatorias independientes, $A_i \sim N\left(0, \sigma_A^2\right)$ y $\varepsilon_{ij} \sim N\left(0, \sigma^2\right)$

Modelo de efectos aleatorios con un sólo factor

- Supongamos que tenemos un factor A con k niveles, con n repeticiones por tratamiento. Se mide la variable respuesta, la cual puede describirse mediante el siguiente modelo de efectos:

$$Y_{ij} = \mu + A_i + \varepsilon_{ij};\ \ i = 1, ..., k;\ \ j = 1, ..., n$$

  donde A_i y ε_{ij} son variables aleatorias independientes, $A_i \sim N\left(0, \sigma_A^2\right)$ y $\varepsilon_{ij} \sim N\left(0, \sigma^2\right)$
- La varianza de Y es la suma de dos varianzas:

$$V\left(Y\right) = \sigma_A^2 + \sigma^2$$

Modelo de efectos aleatorios con un sólo factor

- Supongamos que tenemos un factor A con k niveles, con n repeticiones por tratamiento. Se mide la variable respuesta, la cual puede describirse mediante el siguiente modelo de efectos:

$$Y_{ij} = \mu + A_i + \varepsilon_{ij};\ i = 1, ..., k;\ j = 1, ..., n$$

  donde A_i y ε_{ij} son variables aleatorias independientes, $A_i \sim N\left(0, \sigma_A^2\right)$ y $\varepsilon_{ij} \sim N\left(0, \sigma^2\right)$
- La varianza de Y es la suma de dos varianzas:

$$V\left(Y\right) = \sigma_A^2 + \sigma^2$$

  - Las varianzas σ_A^2, σ^2 se conocen como componentes de varianza y se denomina *modelo de componentes de varianzas o de efectos aleatorios*.

Modelo de efectos aleatorios con un sólo factor

- Supongamos que tenemos un factor A con k niveles, con n repeticiones por tratamiento. Se mide la variable respuesta, la cual puede describirse mediante el siguiente modelo de efectos:

$$Y_{ij} = \mu + A_i + \varepsilon_{ij}; \;\; i = 1, ..., k; \;\; j = 1, ..., n$$

  donde A_i y ε_{ij} son variables aleatorias independientes, $A_i \sim N\left(0, \sigma_A^2\right)$ y $\varepsilon_{ij} \sim N\left(0, \sigma^2\right)$
- La varianza de Y es la suma de dos varianzas:

$$V\left(Y\right) = \sigma_A^2 + \sigma^2$$

  - Las varianzas σ_A^2, σ^2 se conocen como componentes de varianza y se denomina *modelo de componentes de varianzas o de efectos aleatorios*.

- La descomposición de la suma de Cuadrados Total es válida:

$$SCT = SCTr + SCE$$

- Deben probarse:

$$\begin{cases} H_0 : \sigma_A^2 = 0 \\ H_1 : \sigma_A^2 \neq 0 \end{cases}$$

- Deben probarse:

$$\begin{cases} H_0 : \sigma_A^2 = 0 \\ H_1 : \sigma_A^2 \neq 0 \end{cases}$$

  - Si $\sigma_A^2 = 0$, entonces todos los efectos de los grupos son iguales, pero si $\sigma_A^2 > 0$, existe variabilidad entre los efectos de los grupos.

- Deben probarse:

$$\begin{cases} H_0 : \sigma_A^2 = 0 \\ H_1 : \sigma_A^2 \neq 0 \end{cases}$$

  - Si $\sigma_A^2 = 0$, entonces todos los efectos de los grupos son iguales, pero si $\sigma_A^2 > 0$, existe variabilidad entre los efectos de los grupos.
  - Para probar estas hipótesis se procede igual al caso de efectos fijos y se utiliza el mismo estadístico. Para este modelo,

$$\begin{aligned} E\left(CMTr\right) &= n\sigma_A^2 + \sigma^2 \\ E\left(CME\right) &= \sigma^2 \end{aligned}$$

Modelo de efectos aleatorios con un sólo factor

Prueba de Hipótesis

- Deben probarse:

$$\begin{cases} H_0 : \sigma_A^2 = 0 \\ H_1 : \sigma_A^2 \neq 0 \end{cases}$$

  - Si $\sigma_A^2 = 0$, entonces todos los efectos de los grupos son iguales, pero si $\sigma_A^2 > 0$, existe variabilidad entre los efectos de los grupos.
  - Para probar estas hipótesis se procede igual al caso de efectos fijos y se utiliza el mismo estadístico. Para este modelo,

$$\begin{aligned} E\left(CMTr\right) &= n\sigma_A^2 + \sigma^2 \\ E\left(CME\right) &= \sigma^2 \end{aligned}$$

  - se observa que si H_0 es verdadera, $CMTr$ y CME son estimadores insesgados de σ^2, mientras que si H_1 es verdadera, el valor esperado del numerador en mayor que el del denominador. Por lo tanto se debe rechazar H_0 para valores grandes del estadístico $F_0 = \frac{CMTr}{CME}$ y esto implica una región crítica unilateral derecha.

- El procedimiento y la tabla de análisis de la varianza para el modelo es idéntico al de efectos fijos, sin embargo la conclusiones son diferentes ya que se aplican a toda la población de tratamientos.

Tabla ANOVA para el DCA de efectos aleatorios					
Variab	SC	g.l	CM	F valor	
Entre	SCTr	k-1	CMTr=$\frac{SCTr}{k-1}$	$F_0=\frac{CMTr}{CME}$	*valor p*
Dentro	SCE	N-k	CME=$\frac{SCE}{N-k}$		
Total	SCT	N-1			

- Usualmente interesa calcular las componentes de la varianza. Para encontrar sus estimadores, planteamos las ecuaciones:

$$\begin{aligned} \widehat{\sigma}^2 &= CME \\ \widehat{\sigma}^2 + n\widehat{\sigma}^2_A &= CMTr \end{aligned}$$

- Usualmente interesa calcular las componentes de la varianza. Para encontrar sus estimadores, planteamos las ecuaciones:

$$\begin{aligned} \widehat{\sigma}^2 &= CME \\ \widehat{\sigma}^2 + n\widehat{\sigma}_A^2 &= CMTr \end{aligned}$$

- Esto es:

$$\begin{aligned} \widehat{\sigma}^2 &= CME \\ \widehat{\sigma}_A^2 &= \frac{CMTr - CME}{n} \end{aligned}$$

- Usualmente interesa calcular las componentes de la varianza. Para encontrar sus estimadores, planteamos las ecuaciones:

$$\begin{aligned} \widehat{\sigma}^2 &= CME \\ \widehat{\sigma}^2 + n\widehat{\sigma}_A^2 &= CMTr \end{aligned}$$

- Esto es:

$$\begin{aligned} \widehat{\sigma}^2 &= CME \\ \widehat{\sigma}_A^2 &= \frac{CMTr - CME}{n} \end{aligned}$$

- Si tenemos tamaños de muestras desiguales, n se reemplaza por

$$n_0 = \frac{1}{k-1}\left[\sum_{i=1}^{k} n_i - \frac{\sum_{i=1}^{k} n_i^2}{\sum_{i=1}^{k} n_i}\right]$$

- El modelo de efectos aleatorios se ve más afectado por la falta de normalidad, que en el de efectos fijos.

- El modelo de efectos aleatorios se ve más afectado por la falta de normalidad, que en el de efectos fijos.
- En particular, si se calculan intervalos de confianza para las componentes se ven afectados los niveles reales de confianza.

- El modelo de efectos aleatorios se ve más afectado por la falta de normalidad, que en el de efectos fijos.
- En particular, si se calculan intervalos de confianza para las componentes se ven afectados los niveles reales de confianza.
- La falta de homogeneidad de varianzas de los errores puede perturbar significativamente las inferencias respecto de las componentes de varianzas, aunque se tenga un experimento con datos balanceados.

- El modelo de efectos aleatorios se ve más afectado por la falta de normalidad, que en el de efectos fijos.
- En particular, si se calculan intervalos de confianza para las componentes se ven afectados los niveles reales de confianza.
- La falta de homogeneidad de varianzas de los errores puede perturbar significativamente las inferencias respecto de las componentes de varianzas, aunque se tenga un experimento con datos balanceados.
- Los supuestos se prueban a partir de los residuos, igual que en el Modelo I

- MONTGOMERY C. DOUGLAS. Diseño y Análisis de Experimentos. Segunda Edición. LIMUSA WILEY
- GUTIERREZ H y DE LA VARA R. Análisis y Diseño de Experimentos. Segunda Edición. Mc Graw Hill.
- KUEHL ROBERT. Diseño de Experimentos. Segunda Edición. Thomson.
- VICENTE M, GIRON P, NIETO C, PÉREZ T. Diseño de Experimentos. Pearson Prentice Hall.
- DEVORE JAY. Probabilidad y Estadística para Ingeniería y Ciencias. Sexta Edición.
- WALPOLE MYERS. Probabilidad y Estadística. Cuarta Edición. Mc Graw Hill.