

DISEÑO DE EXPERIMENTOS (PARTE I)

Módulo 16

APUNTES DE CLASE

Profesor: Arturo Ruiz-Falcó Rojas

Madrid, Mayo 2009

ÍNDICE

1.	DISEÑOS FACTORIALES A DOS NIVELES 2K.....	3
1.1.	UN PROCESO DE MEJORA MÁS SOFISTICADO: ¿CÓMO PUEDE SE MEJORAR UN PROCESO?	3
1.2.	COMPARACIÓN CON EL MÉTODO CLÁSICO DE “VARIAR UN FACTOR CADA VEZ”	3
1.3.	DESCRIPCIÓN DE LOS DISEÑOS FACTORIALES COMPLETOS A DOS NIVELES.....	7
1.3.1.	TERMINOLOGÍA	7
1.3.2.	CREACIÓN DE UN DISEÑO FACTORIAL COMPLETO A 2 NIVELES7	
1.3.3.	CÁLCULO DE LOS EFECTOS PRINCIPALES Y LAS INTERACCIONES.....	10
1.3.4.	DETERMINACIÓN DE LOS EFECTOS E INTERACCIONES SIGNIFICATIVAS.....	16
1.3.5.	MODELADO EMPÍRICO DEL PROCESO	19
2.	CASO DE DATOS NO NORMALES. TRANSFORMACIONES.....	24

1. DISEÑOS FACTORIALES A DOS NIVELES 2K

1.1. UN PROCESO DE MEJORA MÁS SOFISTICADO: ¿CÓMO PUEDE SE MEJORAR UN PROCESO?

El proceso de mejora visto hasta ahora se limita a “comprobar si se ha producido una mejora” en una alternativa, pero no soluciona el problema de diseñar esa alternativa.

En la industria habitualmente se deben resolver problemas como los que siguen:

- ¿Cómo se debe modificar el producto para aumentar su resistencia mecánica? ¿Qué factores del proceso hay que modificar?
- ¿Cómo se puede reducir los defectos de fabricación?

Este proceso de adquisición de conocimiento es en esencia similar al progreso científico. El progreso científico se basa en definir modelos teóricos, realizar observaciones o experimentos que corroboren el modelo supuesto o que sean incompatibles con él, etc.

Por esta razón, la experimentación puede ser también muy útil para la resolución de problemas industriales como los indicados anteriormente.

1.2. COMPARACIÓN CON EL MÉTODO CLÁSICO DE “VARIAR UN FACTOR CADA VEZ”

La empresa **NOVOMAT** necesita mejorar las características de calidad de su producto. En la actualidad el producto apenas supera las 200 uf¹, mientras que los

¹ Uf = unidades ficticias.

DISEÑO DE EXPERIMENTOS (PARTE I)

clientes requieren que dicha característica supere claramente ese valor. Para ello nombra un grupo de trabajo con la finalidad de que estudien el proceso de fabricación y definan las modificaciones a introducir en el mismo. Este grupo de trabajo estaba integrado por personal muy experto en el proceso y en las tecnologías relacionadas con el mismo. En la primera sesión de trabajo el grupo recolectó la siguiente información:

- Los factores que tienen una incidencia más directa son:
 - Temperatura del proceso, medida en °C.
 - Concentración del producto base (en %).
 - Tipo de catalizador utilizado, existen dos tipos, el A y el B.
- Los rangos en los que pueden operar los factores anteriores son:
 - Temperatura: entre 150 °C y 190 °C.
 - Concentración: entre 15% y 45%.
- El proceso esta operando actualmente en:
 - Temperatura: 170 °C.
 - Concentración: 30%.
 - Catalizador: tipo A.

Sin embargo no existía acuerdo en los siguientes puntos:

- **Grado de influencia de cada uno de los factores en la calidad del producto.** Las opiniones estaban muy divididas, un sector apostaba por la temperatura como factor clave, otros le daban más importancia a la concentración, y finalmente había quien apostaba por el catalizador B como la solución de todos los problemas.
- **Punto de operación del proceso.** Los partidarios de la temperatura creían que la solución pasaba por subir la temperatura de operación del proceso (pese al incremento de coste que conlleva). Por otra parte los partidarios de la concentración apostaban por un aumento de la misma. Finalmente los

DISEÑO DE EXPERIMENTOS (PARTE I)

partidarios del cambio de catalizador pensaban que los actuales puntos de operación eran correctos y que la solución consiste simplemente en cambiar al catalizador B.

Las posturas eran enconadas y nadie daba su brazo a torcer. Por fin alguien tuvo la idea... *“¿por qué no se lo preguntamos al proceso?”*, Antes de que los demás le acusaran de loco, aclaró que lo que quería decir es que la única solución pasa por experimentar y aprender todos de los resultados. Enseguida estuvieron todos de acuerdo, el problema surgió cuando alguien *preguntó “¿qué ensayos hacemos?”* El plan de ensayos propuesto fue el siguiente:

- Probar primero con un catalizador, fijar un de los dos factores hasta alcanzar el máximo de la respuesta; fijar este valor para el primer factor y variar el segundo hasta encontrar el máximo.
- Repetir con el segundo catalizador.
- Seleccionar el catalizador y los correspondientes puntos de operación que maximicen la calidad.

Antes de que se comenzara a discutir si se variaba primero la concentración o la temperatura, alguien señaló *“¿seguro que así encontraremos el máximo? ¿Qué pasa si nos confundimos de montaña?”* (Ver Figura 1), *“Además,...”* añadió *“¿cómo vamos a diferenciar en el experimento qué variaciones se deben al azar y cuáles a los nuevos puntos de operación?”*

DISEÑO DE EXPERIMENTOS (PARTE I)

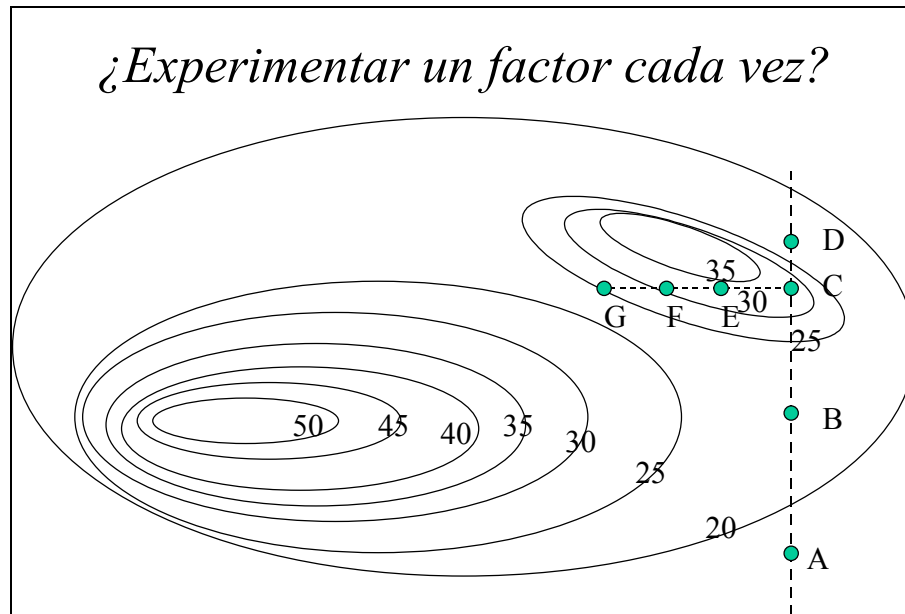


Figura 1: Estrategia experimental “variar un factor cada vez”

Cuando comenzó a cundir el desánimo, alguien se acordó de algo que le habían contado sobre el “*diseño estadístico de experimentos*”. No veían muy claro cómo se pueden extraer conclusiones de un ensayo en el que se modifique más de un factor a la vez. Por esta razón acordaron documentarse sobre el tema y reanudar la reunión unos días más tarde.

1.3. DESCRIPCIÓN DE LOS DISEÑOS FACTORIALES COMPLETOS A DOS NIVELES

1.3.1. TERMINOLOGÍA

Para resolver el problema planteado, se pueden aplicar los diseños factoriales, pero antes de seguir adelante conviene fijar los siguientes conceptos:

- **Respuesta:** Es el resultado de cada uno de los experimentos. En el caso de **NOVOMAT** es la característica de calidad del producto.
- **Factor o Tratamiento:** Son las variables de las que se considera que depende la respuesta. En el caso de **NOVOMAT** los factores son la temperatura, la concentración y el catalizador. A su vez pueden ser:
 - **Cuantitativos:** Cuando pueden variar en una escala continua. Por ejemplo, la temperatura y la concentración.
 - **Cualitativos:** Cuando pueden variar en tipos. Por ejemplo, el catalizador. En el caso anterior, si el proceso solo pudiera operar a dos presiones distintas, por ejemplo, si actuara una sola bomba o dos bombas, ésta sería un factor cualitativo.
- **Nivel:** Son los valores que toma un factor en un experimento dado.
- **Matriz del diseño:** Es una matriz en la que las filas hacen referencia a los distintos ensayos a realizar, las columnas a los distintos factores que intervienen y las celdas indican los niveles en los que hay que situar cada uno de los factores en los distintos experimentos.

1.3.2. CREACIÓN DE UN DISEÑO FACTORIAL COMPLETO A 2 NIVELES

Los diseños más sencillos (y desde luego los utilizados más frecuentemente) son los diseños a "*factoriales completos a dos niveles*". Se llaman "*a dos niveles*" porque cada factor puede tomar dos niveles, y se llaman "*completos*" porque se ensayan todos los niveles de los factores. Por lo tanto, si el diseño tiene k factores, para

DISEÑO DE EXPERIMENTOS (PARTE I)

cubrir todas las condiciones experimentales se requieren 2^k ensayos. En el caso del proceso de **NOVOMAT**, con tres factores se requerirían $2^3=8$ ensayos.

Cada nivel se codifica “-1” el inferior y “+1” el superior, simplemente “-” y “+”. Otra manera habitual de codificarlos es “0” y “1”.

Uno de los aspectos más difíciles es la fijación de los niveles de los factores cuantitativos. Para ello es conveniente tener un buen conocimiento del objeto de la experimentación. En general debe recomendarse que un primer ensayo debe realizarse con los niveles suficientemente separados como para que se pueda apreciar el efecto del factor, pero no demasiado de modo que haya una falta de continuidad en el fenómeno. El equipo de trabajo de **NOVOMAT** ha fijado los niveles de los factores de la manera siguiente:

FACTORES		NIVELES	
		BAJO	ALTO
T	Temperatura	160 °C	180 °C
C	Concentración	20%	40%
K	Catalizador	A	B

Tabla 1: Fijación de los niveles de los factores

La matriz de diseño sería:

Exp.	Factores		
	T	C	K
1	160	20	A
2	180	20	A
3	160	40	A
4	180	40	A
5	160	20	B
6	180	20	B
7	160	40	B
8	180	40	B

Tabla 2: Matriz del diseño

Obsérvese que en la matriz anterior están todas las combinaciones posibles y ninguna está repetida. La forma codificada de escribir la matriz es:

DISEÑO DE EXPERIMENTOS (PARTE I)

	Factores		
Exp.	T	C	K
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

Tabla 3: Matriz del diseño en forma codificada

Este orden de escribir los ensayos en la matrices se denomina “*forma estándar*” y es muy útil para la operativa. Todas las columnas empiezan en “-”. La primera columna alterna “-” y “+”, la segunda alterna “- -” y “+ +”, al tercera alterna “- - -” y “+ + +” y esta regla seguiría si hubiese más columnas. Este orden en el que se escriben los ensayos no presupone que los ensayos tengan que realizarse en ese orden. Todo lo contrario, el orden de los ensayos debe ser aleatorizado, de modo que se proteja al ensayo de posibles cambios sucedidos durante el mismo. Cuando sea preciso, se debe construir bloques aleatorios entre ellos.

APLICACIÓN MINITAB

Minitab dispone de la opción **Stat->DOE->Factorial->Create Factorial design** que simplifica enormemente las tareas anteriores. Con la opción **Stat->DOE->Factorial->Modify design** se puede personalizar los nombres y niveles de los factores. Con la opción **Stat->DOE->Factorial->Display design** se puede optar por presentar la matriz de diseño en orden estándar o en orden de ensayo, así como representar los niveles en unidades reales o codificadas.

1.3.3. CÁLCULO DE LOS EFECTOS PRINCIPALES Y LAS INTERACCIONES

Una vez realizados los ensayos, los resultados obtenidos han sido:

Exp.	T	C	K	Resp.
1	-1	-1	-1	60
2	1	-1	-1	72
3	-1	1	-1	54
4	1	1	-1	68
5	-1	-1	1	52
6	1	-1	1	83
7	-1	1	1	45
8	1	1	1	80

Tabla 4

Los datos se encuentran en el fichero NOVOMAT1.MTW

Representemos gráficamente estos valores de respuesta:

DISEÑO DE EXPERIMENTOS (PARTE I)

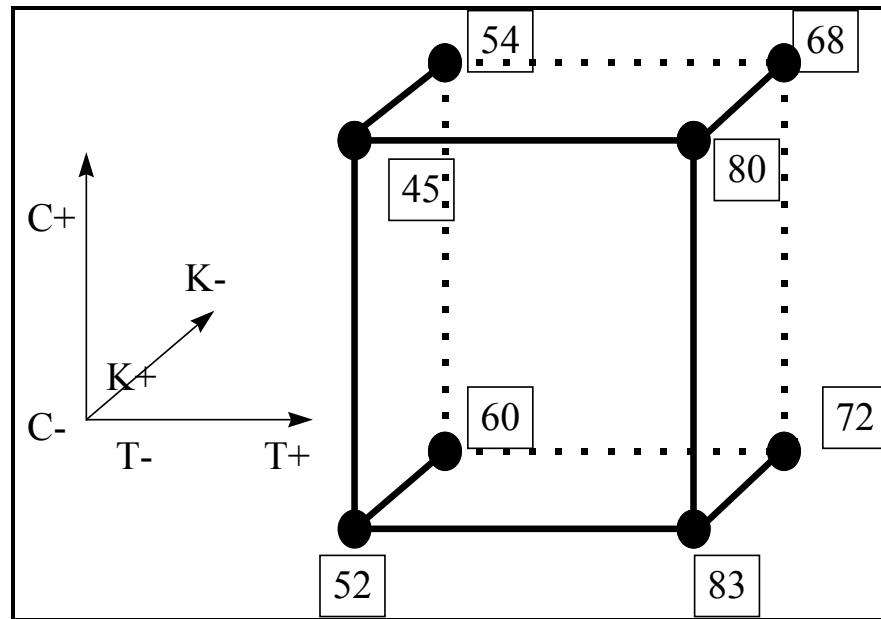


Figura 2: Diagrama de "Cubo"

¿Cuál es el efecto del cambio de temperatura a la respuesta del proceso? Observemos los ensayos 1 y 2. La única diferencia en ellos es la temperatura, luego la diferencia de respuesta deberá explicarse por el efecto de la temperatura más un factor aleatorio. Repitiendo el razonamiento en el resto de los ensayos, concluimos que una buena estimación del efecto de la temperatura es el promedio de la diferencia de respuesta en la que los otros factores no alteran su nivel. Es decir:

$$T = \frac{1}{4} [(72 - 60) + (68 - 54) + (83 - 52) + (80 - 45)] = 23$$

Es fácil comprobar que el efecto de un factor se puede calcular sin más que sumar la columna de los valores de respuesta, afectados por el signo del nivel del factor y divididos entre 4 (número de ensayos dividido entre 2). Por esta razón resulta adecuada la codificación elegida para los niveles.

Observando la Figura 2 puede verse que con el catalizador A obtiene mejores resultados a baja temperatura mientras que el B es más eficaz a alta temperatura;

es decir, existe una *interacción* entre el factor catalizador y el factor temperatura. Esta interacción puede verse de un modo más gráfico en la Figura 3.

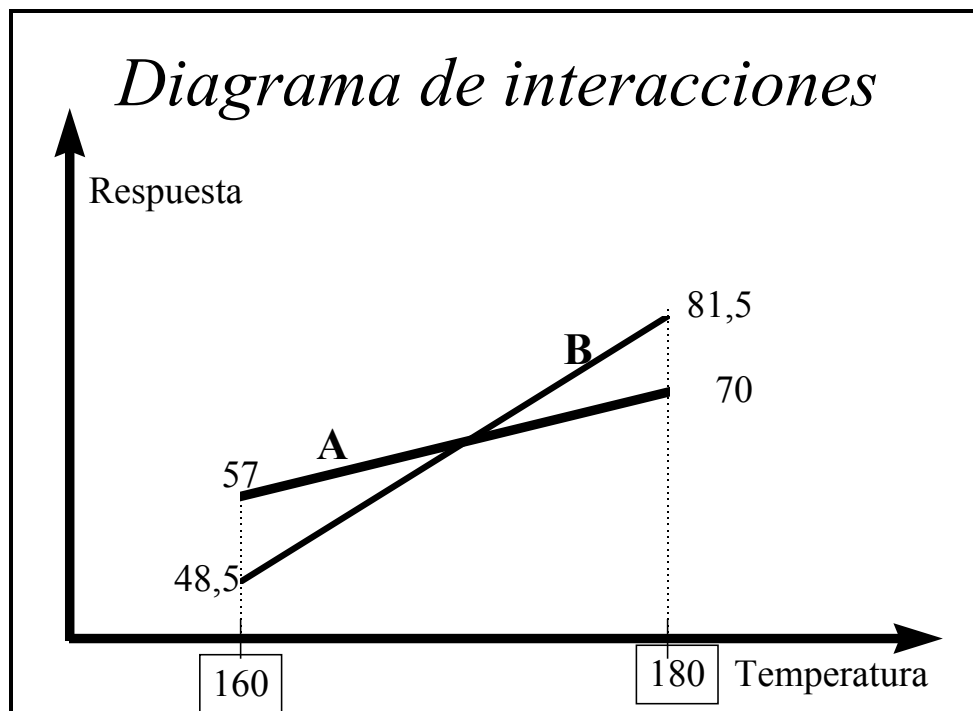


Figura 3: Diagrama de interacciones

Si el catalizador no interaccionara con la temperatura, la estimación del efecto de la temperatura teniendo sólo en cuenta la cara del cubo que corresponde a una catalizador, debería ser bastante similar a la estimación realizada en la cara correspondiente al otro catalizador. En nuestro caso:

Catalizador A:

$$TK_{C_+} = \frac{1}{2} [(72 - 60) + (68 - 54)] = 13$$

Catalizador B:

$$TK_{C_-} = \frac{1}{2} [(83 - 52) + (80 - 45)] = 33$$

DISEÑO DE EXPERIMENTOS (PARTE I)

Una estimación de la interacción entre la temperatura y el catalizador es:

$$TK = \frac{1}{2}(33 - 13) = 10$$

Puede comprobarse que el efecto de la interacción de dos factores se puede calcular sin más que sumar la columna de los valores de respuesta, afectados por el signo del producto de las columnas de niveles de los factores, dividido entre 4 (número de ensayos dividido entre 2). A continuación se representa la tabla completa, con las columnas correspondientes a todas las interacciones y el cálculo de los efectos:

Exp	T	C	K	TC	TK	CK	TCK	Resp.	Efecto	* ²
1	-1	-1	-1	1	1	1	-1	60	64,25	Media
2	1	-1	-1	-1	-1	1	1	72	23	T
3	-1	1	-1	-1	1	-1	1	54	-5	C
4	1	1	-1	1	-1	-1	-1	68	1,5	TC
5	-1	-1	1	1	-1	-1	1	52	1,5	K
6	1	-1	1	-1	1	-1	-1	83	10	TK
7	-1	1	1	-1	-1	1	-1	45	0	CK
8	1	1	1	1	1	1	1	80	0,5	TCK

Tabla 5: Tabla de cálculo de los efectos

APLICACIÓN MINITAB

Minitab dispone de la opción **Stat->DOE->Factorial->Analyze Factorial design**. En el cuadro de diálogo es preciso introducir el modelo a analizar; por ejemplo, sólo efectos principales, también interacciones etc. Además de los diagramas de la Figura 7 y Figura 9, que se verán más adelante, la salida que da es la siguiente:

² Corresponde a los efectos que tengan signo "+".

DISEÑO DE EXPERIMENTOS (PARTE I)

Estimated Effects and Coefficients for Respuest (coded units)

Term	Effect	Coef
Constant		64.250
T	9.000	4.500
C	-3.500	-1.750
K	16.000	8.000
T*C	17.000	8.500
T*K	1.500	0.750
C*K	-4.000	-2.000
T*C*K	1.500	0.750

Analysis of Variance for Respuest (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	698.50	698.500	232.833	*	*
2-Way Interactions	3	614.50	614.500	204.833	*	*
3-Way Interactions	1	4.50	4.500	4.500	*	*
Residual Error	0	0.00	0.000	0.000		
Total	7	1317.50				

Como se puede comprobar, los efectos coinciden con los calculados anteriormente. Además proporciona una tabla de análisis de la varianza, cuya utilidad se verá más adelante.

APLICACIÓN MINITAB

Minitab dispone de varias opciones gráficas para ayudar a visualizar los datos y los efectos. La opción **Stat->DOE->Factorial->Factorials Plot** permite dibujar los diagramas de las figuras siguientes.

Obsérvese en el diagrama de interacciones (ver Figura 6) la fuerte interacción que existe entre la concentración y la temperatura.

DISEÑO DE EXPERIMENTOS (PARTE I)

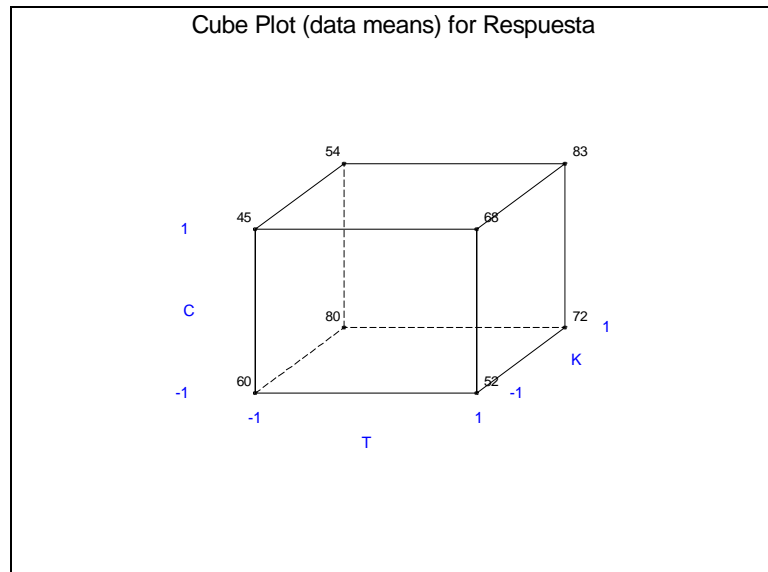


Figura 4: Diagrama de Cubo. Opción *Stat->DOE->Factorial->Factorials Plot*

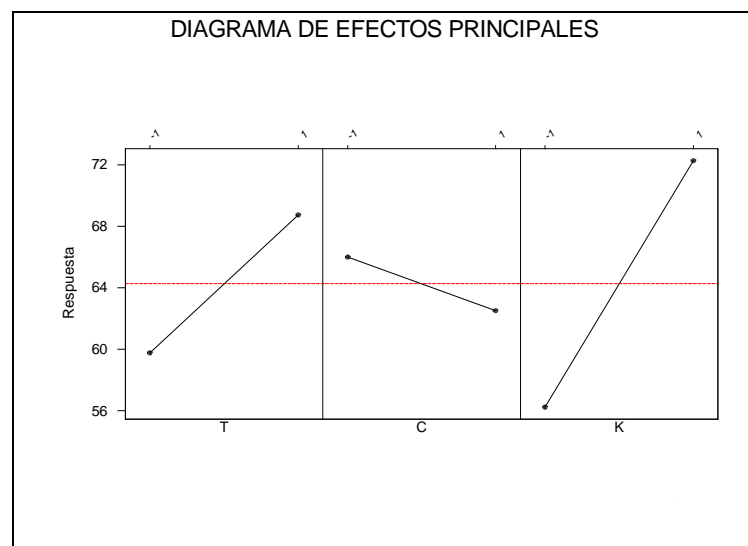


Figura 5: Diagrama de Efectos Principales. Opción *Stat->DOE->Factorial->Factorials Plot*

DISEÑO DE EXPERIMENTOS (PARTE I)

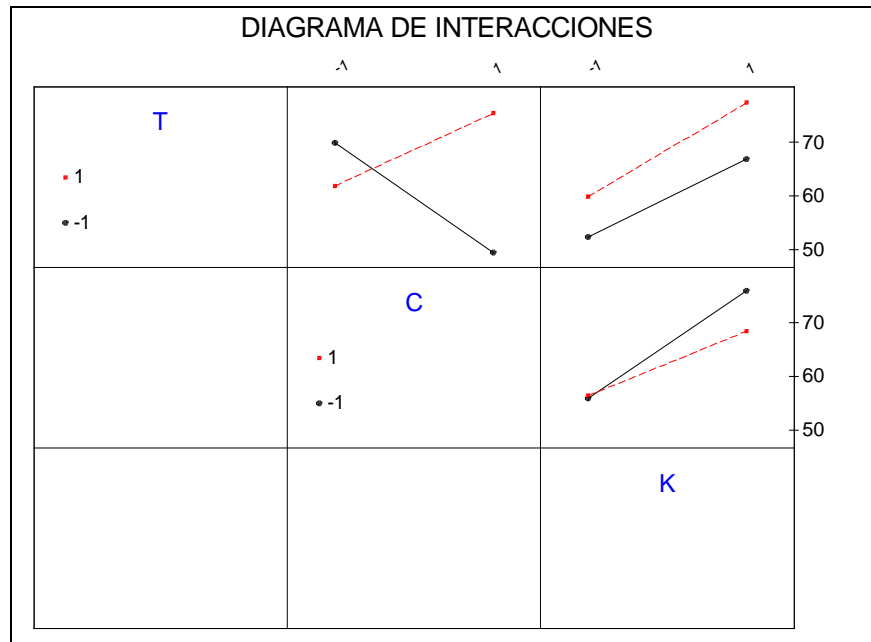


Figura 6: Diagrama de Interacciones. Opción *Stat->DOE->Factorial->Factorials Plot*

1.3.4. DETERMINACIÓN DE LOS EFECTOS E INTERACCIONES

SIGNIFICATIVAS

El siguiente paso es discriminar si los efectos calculados son fruto del propio factor (o interacción de factores, según corresponda), o del azar. Para ello hay tres métodos distintos.

MÉTODO DEL PAPEL PROBABILÍSTICO

El primero de ellos consiste en utilizar el papel probabilístico, de modo que aquellos efectos que no se deban al azar se alejarán de la recta. En la Figura 7 se ha representado el diagrama dibujado por MINITAB. Este diagrama dibuja la recta de regresión y debe ser corregido por la recta que pasa por los puntos próximos a 0 (pequeños y por tanto no significativos. Estos puntos son los que marcan la orientación de la recta. En la Figura 8 se muestra la recta corregida y puede verse que son significativos los efectos T, C y la interacción TK.

DISEÑO DE EXPERIMENTOS (PARTE I)

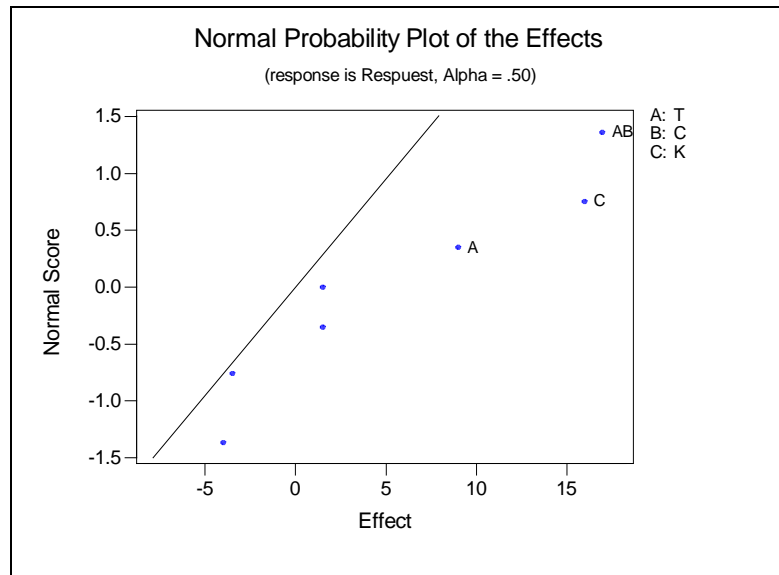


Figura 7: Papel probabilístico para ver qué efectos son significativos. opción *Stat->DOE->Factorial->Analyze Factorial design*

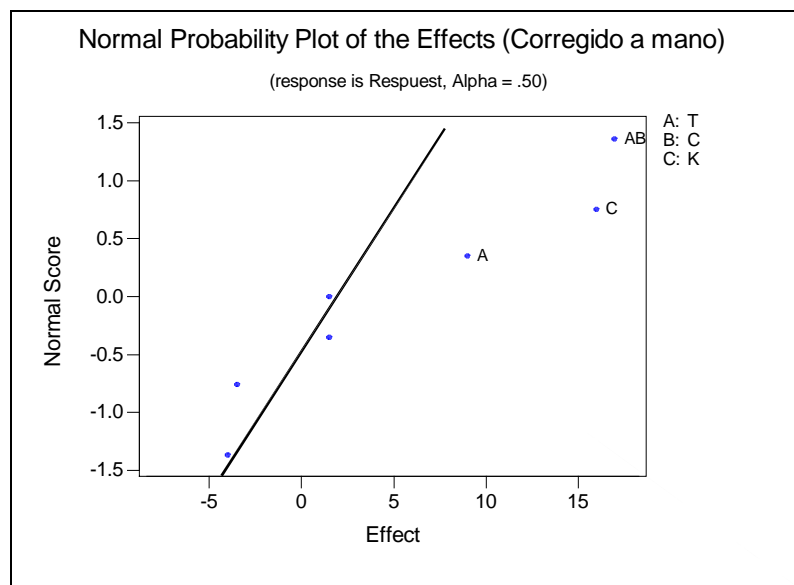


Figura 8: Papel probabilístico con la recta corregida a mano

MÉTODO DE LAS INTERACCIONES DE ORDEN SUPERIOR

Este método se basa en suponer que las interacciones de orden superior (orden tres en este caso) no son significativas y por tanto pueden realizar contrastes con la suma de cuadrados correspondientes en la tabla ANOVA anterior. Estos contrastes indicarán aquellos efectos o interacciones que son significativas. Los resultados pueden representarse en un diagrama de Pareto como el indicado en la Figura 9

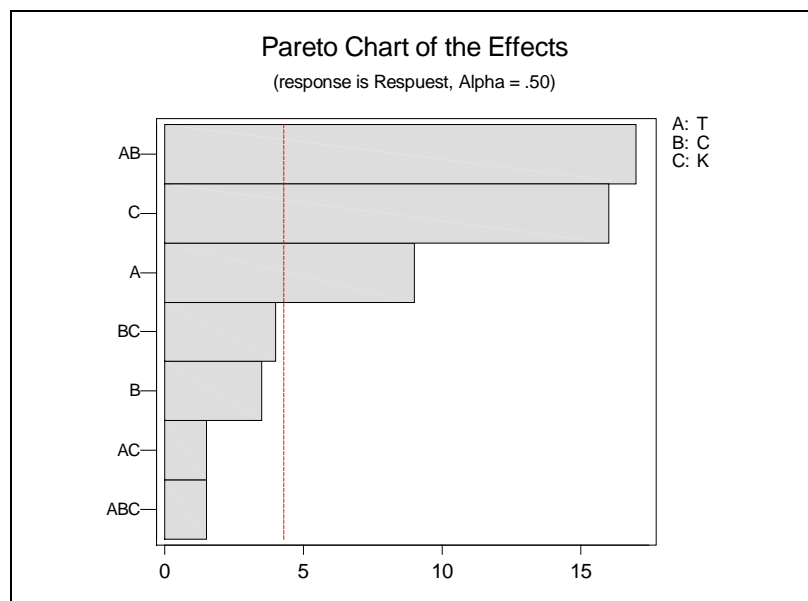


Figura 9: Diagrama de Pareto para ver qué efectos son significativos. opción Stat->DOE->Factorial->Analyze Factorial design

MÉTODO DE LAS RÉPLICAS

El segundo método consiste en *replicar* todos a algunos ensayos, de modo que se pueda realizar una estimación de la variabilidad propia del ensayo. Si σ es la estimación de la variabilidad, la desviación típica a la que está sometida el efecto del factor (por ser suma de cuatro variables aleatorias independientes, cada una de ellas con una desviación típica σ):

$$\sigma_{Ef.}^2 = \frac{4}{N} \sigma^2$$

donde N es el número de ensayos. Una vez conocido σ_{EF} se puede ver el nivel de significación de la magnitud del efecto.

1.3.5. MODELADO EMPÍRICO DEL PROCESO

¿Cómo se pueden interpretar estos resultados? Los efectos calculados son estimaciones del polinomio cuadrático:

$$y = \beta_0 + \beta_T T + \beta_C C + \beta_K K + \beta_{TC} TC + \beta_{TK} TK + \beta_{CK} CK + \beta_{TCK} TCK$$

En efecto, β_0 es la media y los restantes coeficientes son la mitad del efecto correspondiente. En nuestro caso:

$$y = 64,25 + \frac{23}{2} T - \frac{5}{2} C + \frac{10}{2} TK$$

Es preciso observar que los valores de T y C están codificados; por ejemplo, si se quisiese realizar una predicción para C= x %, sería preciso reemplazar C por:

$$L_x = \frac{x\% - 30}{20}$$

De esta manera es posible saber cuales son los factores e interacciones que influyen en la respuesta y en qué dirección hay que modificarlos para mejorar esa respuesta.

APLICACIÓN MINITAB

Ya se ha visto cómo Minitab calcula los efectos y coeficientes. Con la opción **Stat->DOE->Factorial->Display design** se puede optar por representar los niveles en unidades reales. En este caso los coeficientes que da Minitab no coinciden con los conseguidos con el cambio de unidades a los efectos e interacciones significativas, como se ha descrito anteriormente, porque Minitab realiza el cambio para todos los términos, significativo y no significativos. Con la opción **Stat->DOE->Factorial->Contour / Surface (Wireframe) Plots** se puede representar gráficamente el modelo (ver figuras siguientes).

DISEÑO DE EXPERIMENTOS (PARTE I)

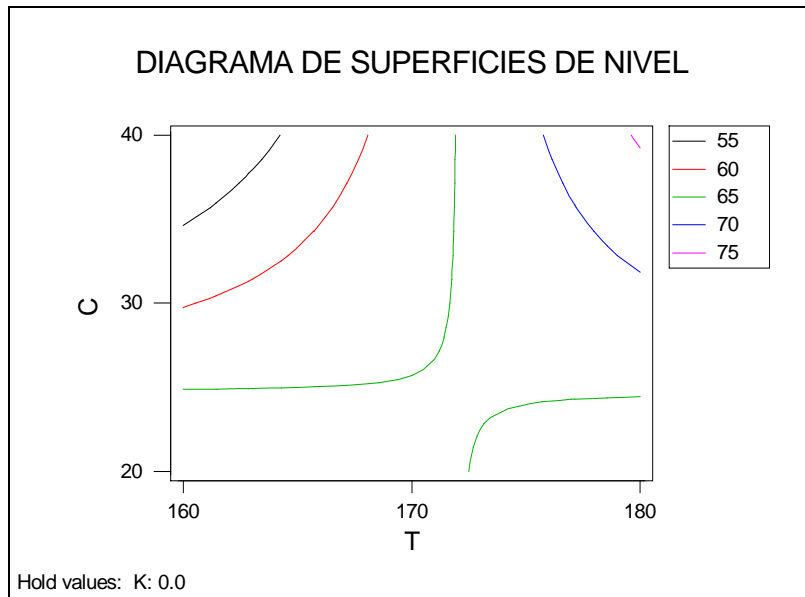


Figura 10: Diagrama de superficies de nivel. Opción *Stat->DOE->Factorial->Contour / Surface (Wireframe) Plots*

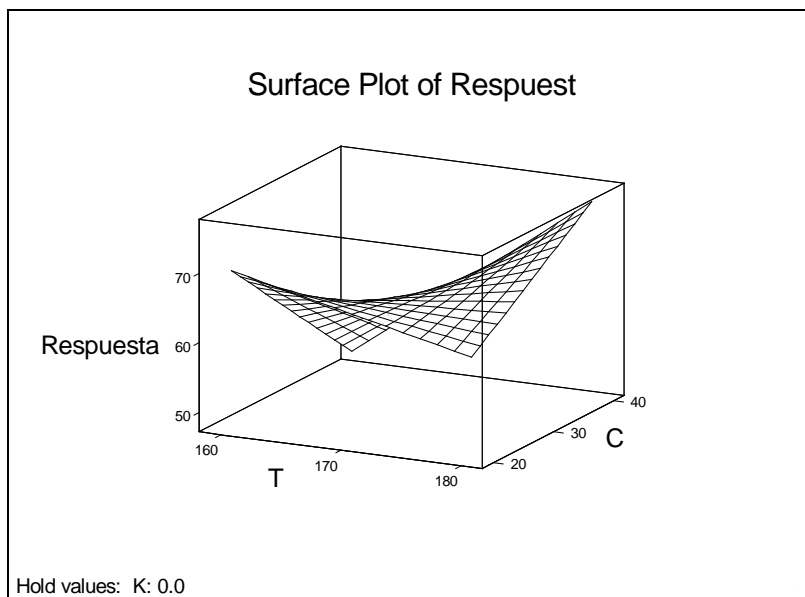


Figura 11: Diagrama de superficie. Opción *Stat->DOE->Factorial->Contour / Surface (Wireframe) Plots*

EJEMPLO 1. EJEMPLO DE ESTIMACIÓN DE LOS EFECTOS SIGNIFICATIVOS A PARTIR DE RÉPLICAS

LACADOS, S.A está poniendo a punto un nuevo proceso de acabado superficial en madera. El proceso consiste esencialmente en una cinta transportadora que arrastra las piezas a una cabina dónde se le chorrea un producto químico. Hasta el momento las pruebas no eran totalmente satisfactorias debido a la falta de uniformidad de la capa depositada. Para mejorar este proceso constituyó un grupo de trabajo formado por un ingeniero de proceso, un técnico de laboratorio y un ingeniero de calidad. Éste último sugirió realizar algunos ensayos para caracterizar el comportamiento del proceso. Al planificar los ensayos, el grupo de trabajo se enfrentó a dos problemas:

- Cómo medir la uniformidad de la capa depositada de una manera sencilla.
- Cuáles son los factores que afectan al proceso.

Para resolver el primer punto, el técnico de laboratorio ideó un método basado en la cantidad de luz reflejada en una dirección determinada por una probeta tratada por el proceso de protección superficial, y compararla con la reflejada por un espejo. Esto se basa en la idea de que cuanto más rugosa es la capa depositada más disperso es el haz de luz reflejado.

El segundo punto lo resolvió el ingeniero de proceso quien identificó los factores indicados en la tabla. También indicó el rango de valores en los que de acuerdo con su experiencia podrían situarse los factores y el valor actual de operación.

Factor	Valor actual	Máximo	Mínimo
Cantidad de producto en g/min.	100 g/min.	130 g /min.	80 g /min.
Velocidad de avance de la cinta	5 m / min.	8 m/min.	3 m/min.

Tabla 6

DISEÑO DE EXPERIMENTOS (PARTE I)

Además el ingeniero de proceso atribuyó el problema al tipo de producto utilizado (A) y sugería la utilización del (B) pese a que este es bastante más caro.

Con esta información se pretende planificar los ensayos para mejorar el proceso.

Ejemplo 4: Parte 1

El grupo de trabajo acordó realizar un ensayo 2^3 sin replicar. Los puntos de ensayo elegidos fueron:

Factor	Alto	Bajo
Cantidad depositada en g/min.	110 g /min.	90 g /min.
Velocidad de avance de la cinta	6 m/min.	4 m/min.

A continuación se reproduce la matriz de diseño con los valores obtenidos:

Los datos se encuentran en el fichero LACADOS1.MTW

Ensayo número	Orden	Cantidad de pintura: (C)	Velocidad de avance: (V)	Tipo de pintura: (P)	CV	CP	VP	CV P	Respuesta	Efecto	
1	2	-1	-1	-1	1	1	1	-1	40	37,13	Mediana
2	4	1	-1	-1	-1	-1	1	1	25	1,75	C
3	5	-1	1	-1	-1	1	-1	1	30	6,75	V
4	1	1	1	-1	1	-1	-1	-1	50	19,25	CV
5	8	-1	-1	1	1	-1	-1	1	45	1,75	P
6	6	1	-1	1	-1	1	-1	-1	25	-0,75	CP
7	7	-1	1	1	-1	-1	1	-1	30	-0,75	VP
8	3	1	1	1	1	1	1	1	52	1,75	CVP

Tabla 7

A) ¿QUÉ EFECTOS SON SIGNIFICATIVOS?.

B) ¿ESTÁ JUSTIFICADO EL CAMBIO DE PINTURA?.

C) ¿HACIA DONDE SUGIERE CAMBIAR EL PUNTO DE OPERACIÓN DEL PROCESO?.

DISEÑO DE EXPERIMENTOS (PARTE I)

EJEMPLO 4: PARTE 2

Los efectos significativos serían la velocidad V y la interacción entre la cantidad de producto y la velocidad VC . La ecuación que modeliza el proceso es:

$$\text{Calidad de acabado} = 37,13 + \frac{6,75}{2}V + \frac{19,25}{2}VC$$

Donde V y C está en unidades codificadas ($V=-1$ equivale a 4 m/min, $V = 1$ equivale a 6 m/min, etc.). En unidades convencionales sería preciso reemplazar en la ecuación anterior V por $\frac{V-5}{2}$ y C por $\frac{C-100}{110-90}$ respectivamente.

Las predicciones con el modelo serían:

Predic.	Real	Residuo
43,375	40	3,375
24,125	25	-0,875
30,875	30	0,875
50,125	50	0,125
43,375	45	-1,625
24,125	25	-0,875
30,875	30	0,875
50,125	52	-1,875

Para confirmar la significación de los efectos y dado que se aprecia una falta de ajuste del modelo, se realiza una segunda réplica de los ensayos con los siguientes resultados:

DISEÑO DE EXPERIMENTOS (PARTE I)

Ejemplo 4: Parte 3

Los resultados de la segunda réplica junto con los nuevos cálculos se incluyen en la tabla siguiente:

Ensayo número	Orden	Cantidad de pintura: (C)	velocidad de avance: (V)	Tipo de pintura: (P)	CV	CP	VP	CVP	R1	R2	Respuesta	Efecto		$(r_1-r_2)^2$
1	2	-1	-1	-1	1	1	1	-1	40	36	38	37,00	Media	16
2	4	1	-1	-1	-1	-1	1	1	25	28	26,5	1,75	C	9
3	5	-1	1	-1	-1	1	-1	1	30	32	31	6,75	V	4
4	1	1	1	-1	1	-1	-1	-1	50	48	49	19,25	CV	4
5	8	-1	-1	1	1	-1	-1	1	45	43	44	1,75	P	4
6	6	1	-1	1	-1	1	-1	-1	25	30	27,5	-0,75	CP	25
7	7	-1	1	1	-1	-1	1	-1	30	29	29,5	-0,75	VP	1
8	3	1	1	1	1	1	1	1	52	49	50,5	1,75	CVP	9
SUMA														72

$$Var = \frac{72}{8} = 9 \quad Var \text{ efecto} = \left(\frac{1}{8} + \frac{1}{8} \right) 9 = 2,25; \sigma_{Efecto} = 1,50$$

**2. CASO DE DATOS NO NORMALES.
TRANSFORMACIONES**

La técnica de normalización de datos más eficaz es la transformación de Box-Cox. Esta transformación es la siguiente:

$$x^{(\lambda)} = \begin{cases} \frac{(x+m)^\lambda - 1}{\lambda} & (\lambda \neq 0) \quad (\forall x \quad x > -m) \\ \ln(x+m) & (\lambda = 0) \quad (m > 0) \end{cases}$$

Es continua en λ , puesto que:

DISEÑO DE EXPERIMENTOS (PARTE I)

$$\lim_{\lambda \rightarrow 0} x^{(\lambda)} = \ln(x + m)$$

La función de verosimilitud de los datos normalizados es:

$$L(\lambda) = -\frac{1}{n} \ln(\hat{\sigma}^2(\lambda)) + (\lambda - 1) \sum_{i=1}^n \ln(x_i)$$

Donde:

$$\hat{\sigma}^2(\lambda) = \frac{1}{n} \sum_{i=1}^n (x^{(\lambda)} - \bar{x}^{(\lambda)})^2$$

La técnica consiste en hallar el valor de λ que maximiza la verosimilitud (ver Figura 12).

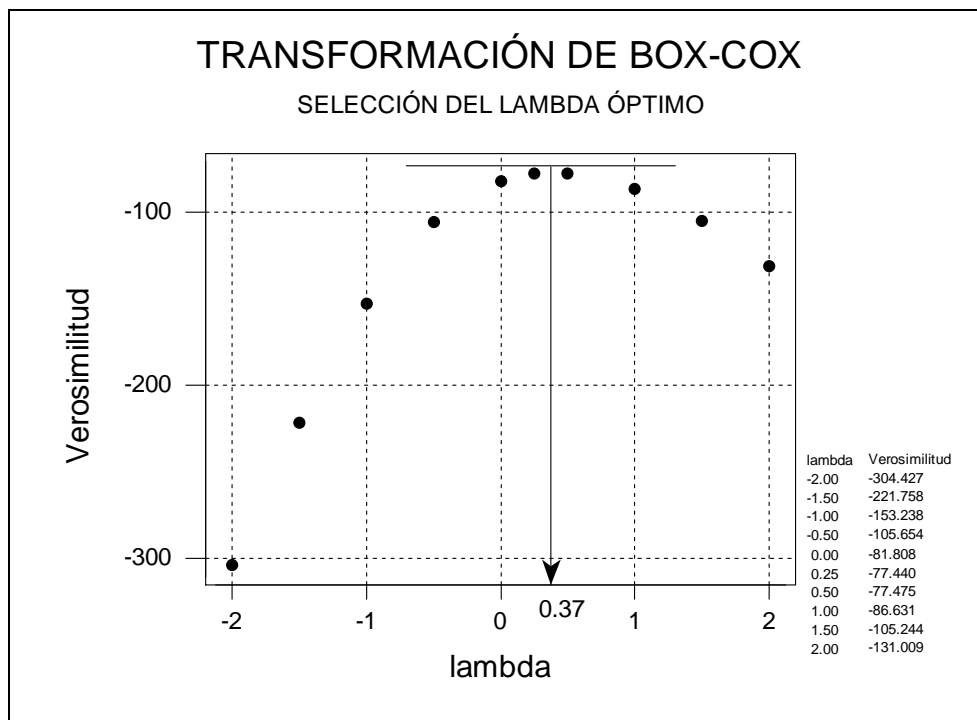


Figura 12: Transformación de Box-Cox para normalizar los datos

Según sea el origen de los datos, en algunos casos se puede esperar heterocedasticidad de los mismos (varianza no constante). Según esté relacionada la varianza y la media, resultan apropiadas las transformaciones indicadas en la Tabla 8.

Relación media-varianza	λ	Transformación estabilizadora de la varianza	Ejemplos típicos de aplicación
$\sigma \propto \eta^2$	-1	Inversa	
$\sigma \propto \eta^{1.5}$	-1/2	Inversa de la raíz cuadrada	
$\sigma \propto \eta$	0	Logaritmo	Análisis de varianzas muestrales, gráficos de control de series cortas en el que se conjuguen familias de varios espesores.
$\sigma \propto \eta^{0.5}$	1/2	Raíz cuadrada	Datos procedentes de una distribución de Poisson.
$\sigma \propto const$	1	No se transforma	

Tabla 8: Transformaciones de datos para eliminar la heterocedasticidad

EJEMPLO 2. NORMALIZACIÓN DE DATOS DE UN PROCESO NO NORMAL

En la Figura 13 se ha representado el histograma de un proceso de protección superficial. A la vista del citado histograma, se concluye la no normalidad de los datos. Se desea realizar una transformación que los normalice.

APLICACIÓN MINITAB

Menú **Stat->Control Charts->Box-Cox Transformation**. Los datos se encuentran en el fichero DatosNoNormales.MTW. La respuesta proporcionada por MINITAB es la indicada en la Figura 14, en la que se indica $\hat{\lambda}=0.113$ como el valor óptimo. Una vez transformados los datos, se puede comprobar su normalidad mediante un histograma (ver Figura 15) o un papel probabilístico, de modo análogo al hecho anteriormente (ver Figura 16).

DISEÑO DE EXPERIMENTOS (PARTE I)

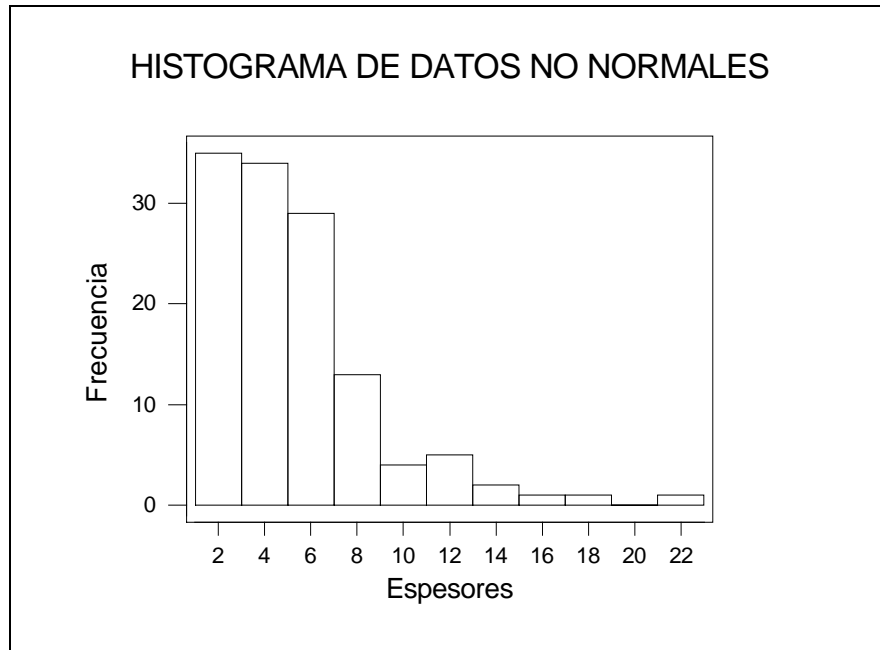


Figura 13: Histograma de datos de un proceso no normal

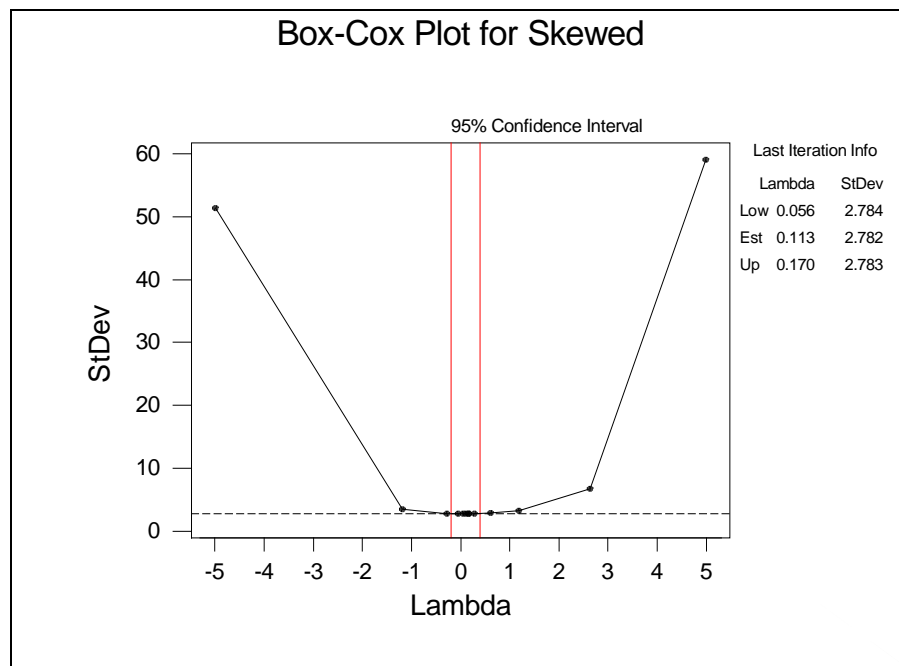


Figura 14: Transformación de Box-Cox (Stat->Control Charts->Box-Cox Transformation)

DISEÑO DE EXPERIMENTOS (PARTE I)

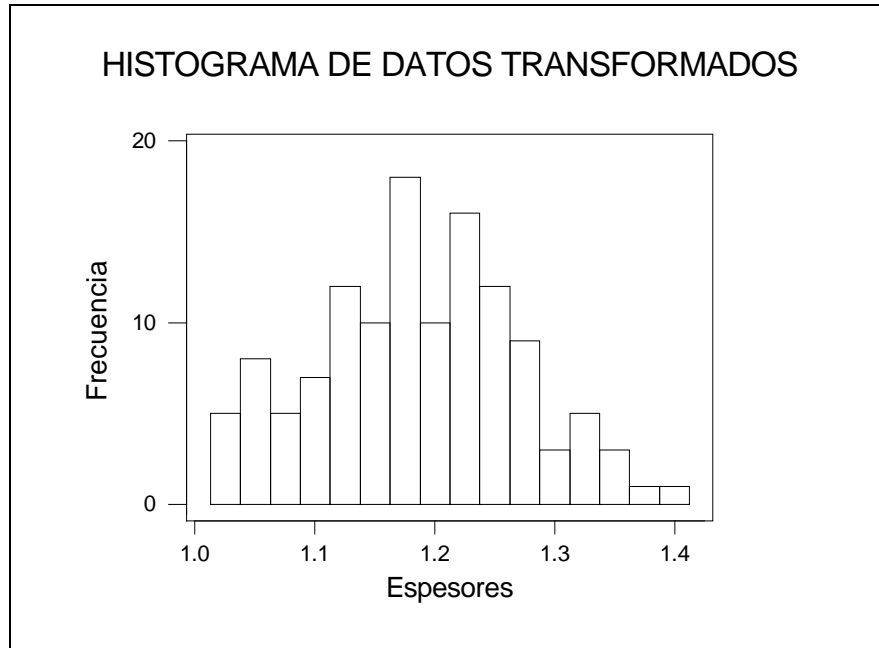


Figura 15: Histograma de los datos transformados (Graph->Histogram)

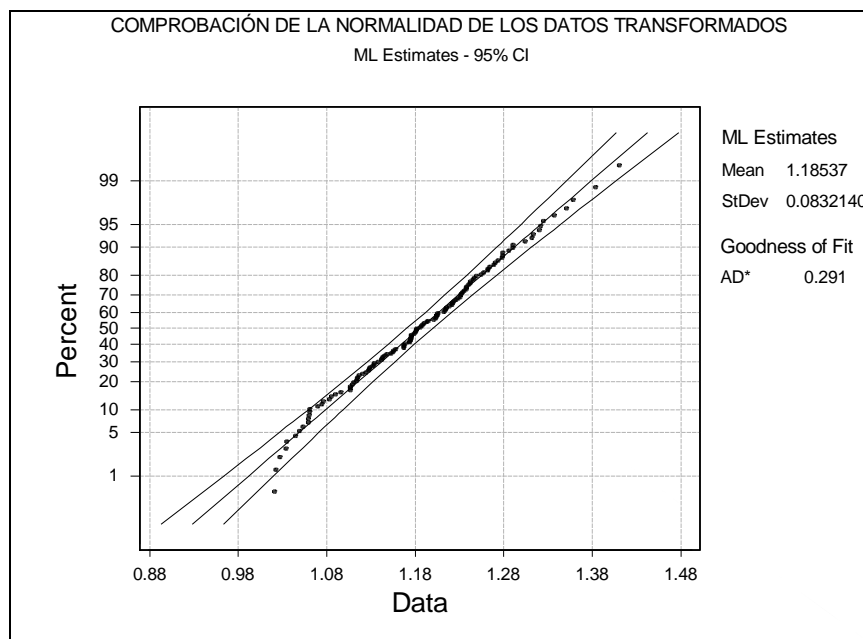


Figura 16: Comprobación de la normalidad de los datos transformados