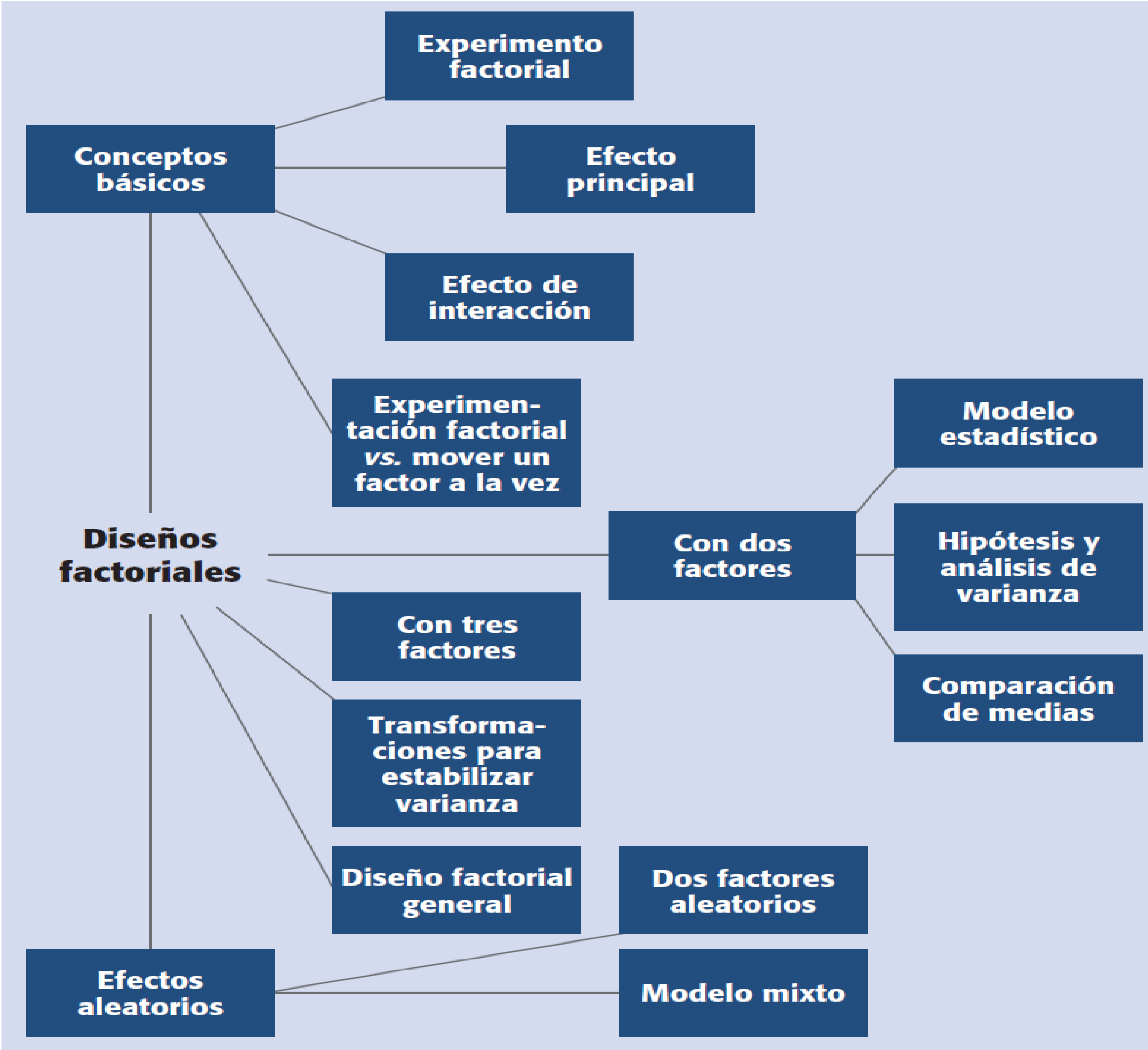

Experimento factorial
Conceptos básicos
Efecto principal
Efecto de interacción
Experimentación factorial *vs.* mover un factor a la vez
Modelo estadístico
Diseños factoriales
Con dos factores
Hipótesis y análisis de varianza
Con tres factores
Comparación de medias
Transformaciones para estabilizar varianza
Diseño factorial general
Dos factores aleatorios
Efectos aleatorios
Modelo mixto

- El objetivo de un *diseño factorial* es estudiar el efecto de varios factores sobre una o varias respuestas, **cuando se tiene el mismo interés sobre todos los factores.**
- Los factores pueden ser de tipo cualitativo (máquinas, tipos de material, operador, la presencia o ausencia de una operación previa, etc.), o de tipo cuantitativo (temperatura, humedad, velocidad, presión, pH, tiempo, etc.).
- Para estudiar la manera en que influye cada factor sobre la variable de respuesta es necesario elegir **al menos dos niveles** de prueba para cada uno de ellos.
- Con el **diseño factorial completo** se corren aleatoriamente todas las posibles combinaciones que pueden formarse con los niveles de los factores a investigar.

- En general, la familia de diseños factoriales 2^k consiste en “k” factores, todos con dos niveles de prueba;
- Y la familia de diseños factoriales 3^k consiste en “k” factores cada uno con tres niveles de prueba.

Diseño factorial 2^2. Supongamos que en un proceso de fermentación tequilera, se tienen dos factores A: *tipo de levadura* y B: *temperatura*, cada uno con dos niveles

Tabla 5.1 Diseño factorial 2^2.

A: Levadura	*B*: Temperatura	*Y*: Rendimiento
$A_1 = 1$ (-1)	$B_1 = 22$ (-1)	28
$A_2 = 2$ (1)	$B_1 = 22$ (-1)	41
$A_1 = 1$ (-1)	$B_2 = 30$ (1)	63
$A_2 = 2$ (1)	$B_2 = 30$ (1)	45

El modelo aditivo lineal de un arreglo factorial con 2 factores conducido en un DCA es:

$$Y_{ijk} = \underbrace{\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}}_{\mu_{ij}} + \varepsilon_{ijk}\,;\; i = 1,\ldots,p\,[p=2]\,;\; j = 1,\ldots,q\,[q=3]\,;\; k = 1,\ldots,r_{ij}\,[r_{ij} = r = 4]$$

Y_{ijk} = Cantidad de ácido ascórbico obtenida con la i-ésima marca de concentrado de jugo y el j-ésimo período en la k-ésima repetición.

μ = Efecto de la cantidad de ácido ascórbico media general.

α_i = Efecto de la i-ésima marca.

β_j = Efecto del j-ésimo periodo.

$(\alpha\beta)_{ij}$ = Efecto de la interacción entre la i-ésima marca y el j-ésimo periodo.

μ_{ij} = Efecto de la media de la combinación (marca-período) ij.

ε_{ijk} = Efecto del error experimental obtenida con la i-ésima marca de concentrado de jugo y el j-ésimo período en la k-ésima repetición.

El primer subíndice indica el nivel del primer factor, el segundo el nivel del segundo factor y el tercero la observación dentro de la muestra. Los factores pueden ser ambos de efectos fijos (se habla entonces de *modelo I*), de efectos aleatorios *(modelo II)* o uno de efectos fijos y el otro de efectos aleatorios *(modelo mixto)*. El modelo matemático de este análisis es:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$ modelo I

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk}$$ modelo II

$$Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + \varepsilon_{ijk}$$ modelo mixto

donde μ es la media global, α_i o A_i el efecto del nivel *i* del 1º factor, β_j o B_j el efecto del nivel *j* del 2º factor y ε_{ijk} las desviaciones aleatorias alrededor de las medias, que también se asume que están normalmente distribuidas, son independientes y tienen media 0 y varianza σ^2.

Efecto principal y efecto de interacción.

- El efecto de un factor se define como el cambio observado en la variable de respuesta debido a un cambio de nivel de tal factor.
- En particular, los efectos principales son los cambios en la media de la variable de respuesta que se deben a la acción individual de cada factor.
- En términos matemáticos, el ***efecto principal*** de un factor con dos niveles es la diferencia entre la respuesta media observada cuando tal factor estuvo en su primer nivel, y la respuesta media observada cuando el factor estuvo en su segundo nivel.

Efecto principal y efecto de interacción.

- Por otra parte, se dice que dos factores interactúan entre sí o tienen un efecto de interacción sobre la variable de respuesta, cuando el efecto de un factor depende del nivel en que se encuentra el otro.

- Por ejemplo, los factores A y B interactúan si el efecto de A es muy diferente en cada nivel de B, o viceversa.

Efecto principal y efecto de interacción.

- Los valores absolutos (sin importar el signo) de los efectos principales y del efecto de interacción son una medida de importancia de su efecto sobre la variable de respuesta.

- Sin embargo, como se tienen estimaciones muestrales, para saber si los efectos son estadísticamente significativos (diferentes de cero) se requiere el análisis de varianza (ANOVA).

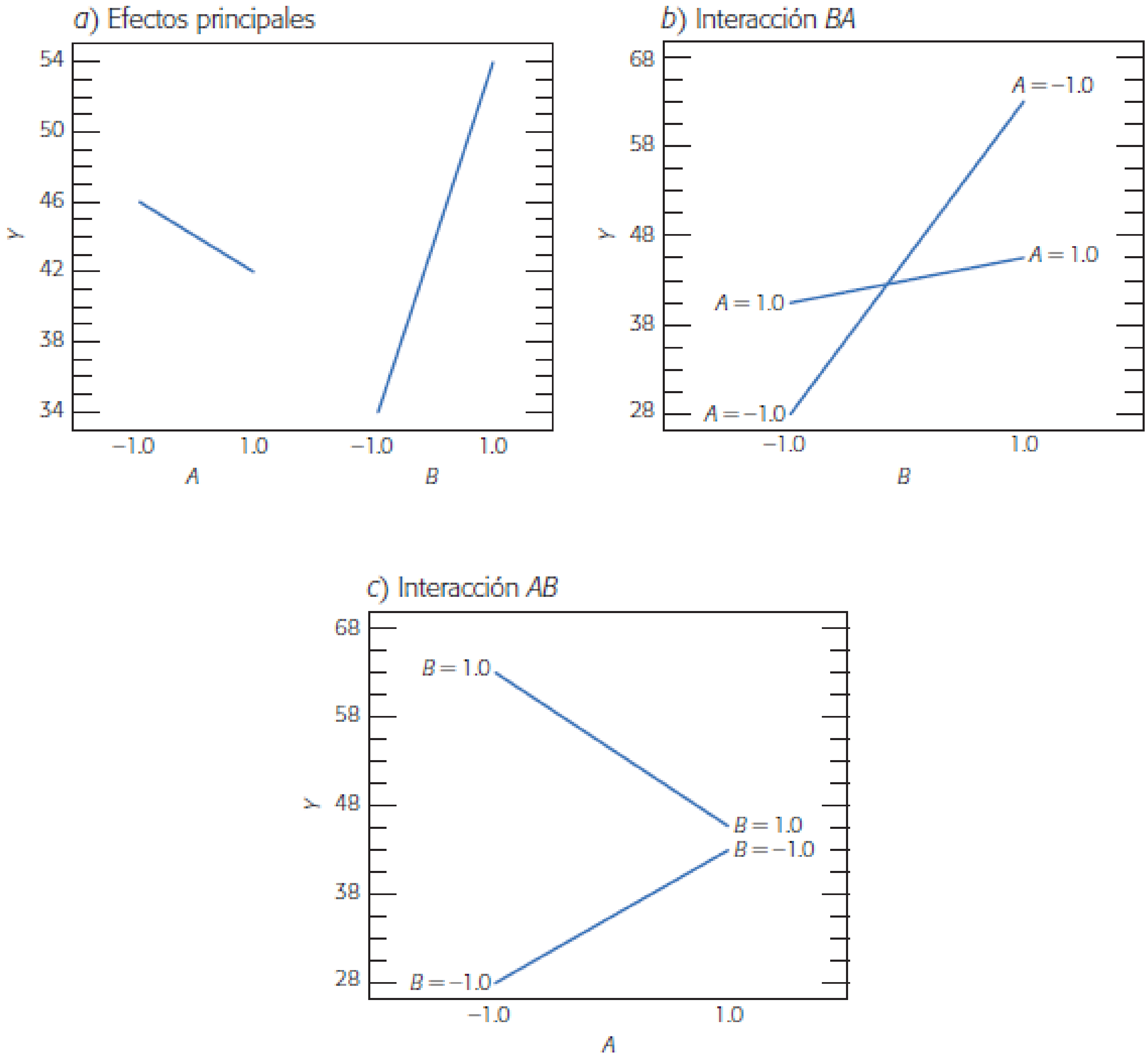
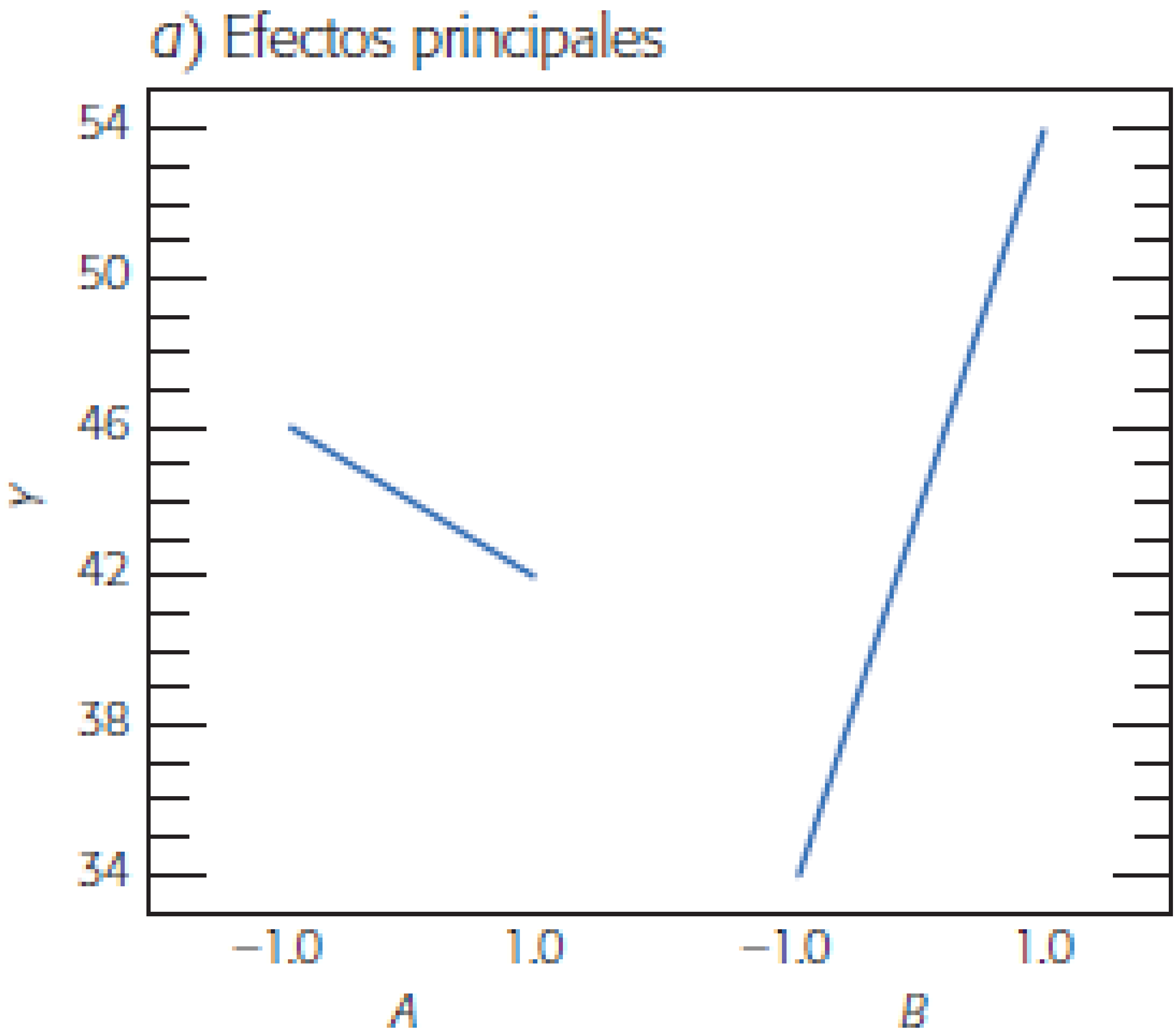


Figura 5.1 Interacción para datos tabla 5.1. Sí existe interacción: el efecto del incremento de B sobre Y es diferente dependiendo del nivel de A, y viceversa.


a) Efectos principales
54
50
46
42
38
34
y
−1.0
1.0
−1.0
1.0
A
B

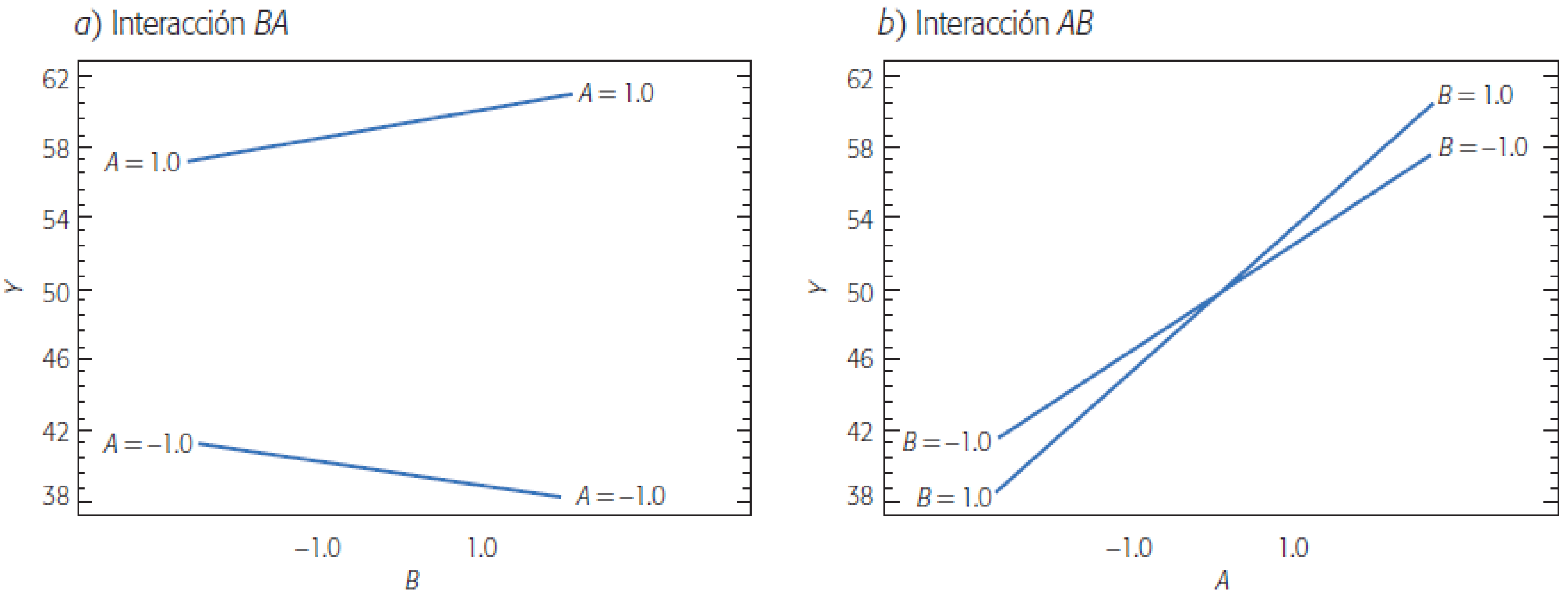


Figura 5.2 No hay efecto de interacción. En *b*) se aprecia que si *A* se aumenta, *Y* aumenta, independientemente del valor de *B*.

Ventajas de los diseños factoriales

1. Permiten estudiar el efecto individual y de interacción de los distintos factores.

2. Son diseños que se pueden aumentar para formar diseños compuestos en caso de que se requiera una exploración más completa. Por ejemplo, es útil aumentar el diseño si el comportamiento de la respuesta no es lineal en los factores controlados.

3. Se pueden correr **fracciones de diseños factoriales**, las cuales son de gran utilidad en las primeras etapas de una investigación que involucra a muchos factores, cuando interesa descartar de manera económica los que no son importantes, antes de hacer un estudio más detallado con los factores que sí son importantes.

4. Pueden utilizarse en combinación con diseños de bloques en situaciones en las que no puede correrse todo el diseño factorial bajo las mismas condiciones.
Por ejemplo, cuando cada lote de material sólo alcanza para correr la mitad del experimento, éste se puede realizar en dos bloques (dos lotes), lo cual implica repartir las pruebas en los dos lotes de la manera más conveniente posible.

5. La interpretación y el cálculo de los efectos en los experimentos factoriales se puede hacer con aritmética elemental, en particular cuando cada factor se prueba en dos niveles.

Modelo estadístico e hipótesis de interés. El modelo estadístico de efectos para este tipo de diseño está dado por:

$$\begin{gathered} Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}; \\ i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, b;\ k = 1, 2, \ldots, n \end{gathered} \tag{5.1}$$

donde μ es la media general, α_i es el efecto debido al i-ésimo nivel del factor A, β_j es el efecto del j-ésimo nivel del factor B, $(\alpha\beta)_{ij}$ representa al efecto de interacción en la combinación ij y ε_{ijk} es el error aleatorio que se supone sigue una distribución normal con media cero y varianza constante σ^2 ($N(0, \sigma^2)$) y son independientes entre sí. Para que la estimación de los parámetros en este modelo sea única, se introducen las restricciones $\sum_{i=1}^{a}\alpha_i = 0$, $\sum_{j=1}^{b}\beta_j = 0$ y $\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij} = 0$. Es decir, los efectos dados en el modelo son desviaciones relacionadas con la media global.

$$H_0: \ \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$$
$$H_A: \ \alpha_i \neq 0 \text{ para algún } i$$

$$H_0: \ \beta_1 = \beta_2 = \ldots = \beta_b = 0$$
$$H_A: \ \beta_j \neq 0 \text{ para algún } i$$

$$H_0: \ (\alpha\beta)_{ij} = 0 \text{ para todo } ij$$
$$H_A: \ (\alpha\beta)_{ij} \neq 0 \text{ para algún } ij$$

Estas hipótesis se prueban mediante la técnica de análisis de varianza, que para un diseño factorial $a \times b$ con n réplicas resulta de descomponer la variación total como,

$$SC_T = SC_A + SC_B + SC_{AB} + SC_E$$

Tabla 5.4 ANOVA para el diseño factorial $a \times b$.

FV	*SC*	*GL*	*CM*	F_0	Valor-*p*
Efecto A	SC_A	$a - 1$	CM_A	CM_A/CM_E	$P(F > F_0^A)$
Efecto B	SC_B	$b - 1$	CM_B	CM_B/CM_E	$P(F > F_0^B)$
Efecto AB	SC_{AB}	$(a - 1)(b - 1)$	CM_{AB}	CM_{AB}/CM_E	$P(F > F_0^{AB})$
Error	SC_E	$ab(n - 1)$	CM_E		
Total	SC_T	$abn - 1$			

Tabla 5.3 Datos del experimento factorial 4 × 3 (ejemplo 5.2).

		***B*: velocidad**			
		0.20	**0.25**	**0.30**	**Total $Y_{i..}$**
A: profundidad	0.15	74 64 $\boxed{198}$ 60	92 86 $\boxed{266}$ 88	99 98 $\boxed{299}$ 102	763
	0.18	79 68 $\boxed{220}$ 73	98 104 $\boxed{290}$ 88	104 99 $\boxed{298}$ 95	808
	0.21	82 88 $\boxed{262}$ 92	99 108 $\boxed{302}$ 95	108 110 $\boxed{317}$ 99	881
	0.24	99 104 $\boxed{299}$ 96	104 110 $\boxed{313}$ 99	114 111 $\boxed{332}$ 107	944
	Total $Y_{\bullet j \bullet}$	979	1 171	1 246	$Y_{...} = 3\ 396$

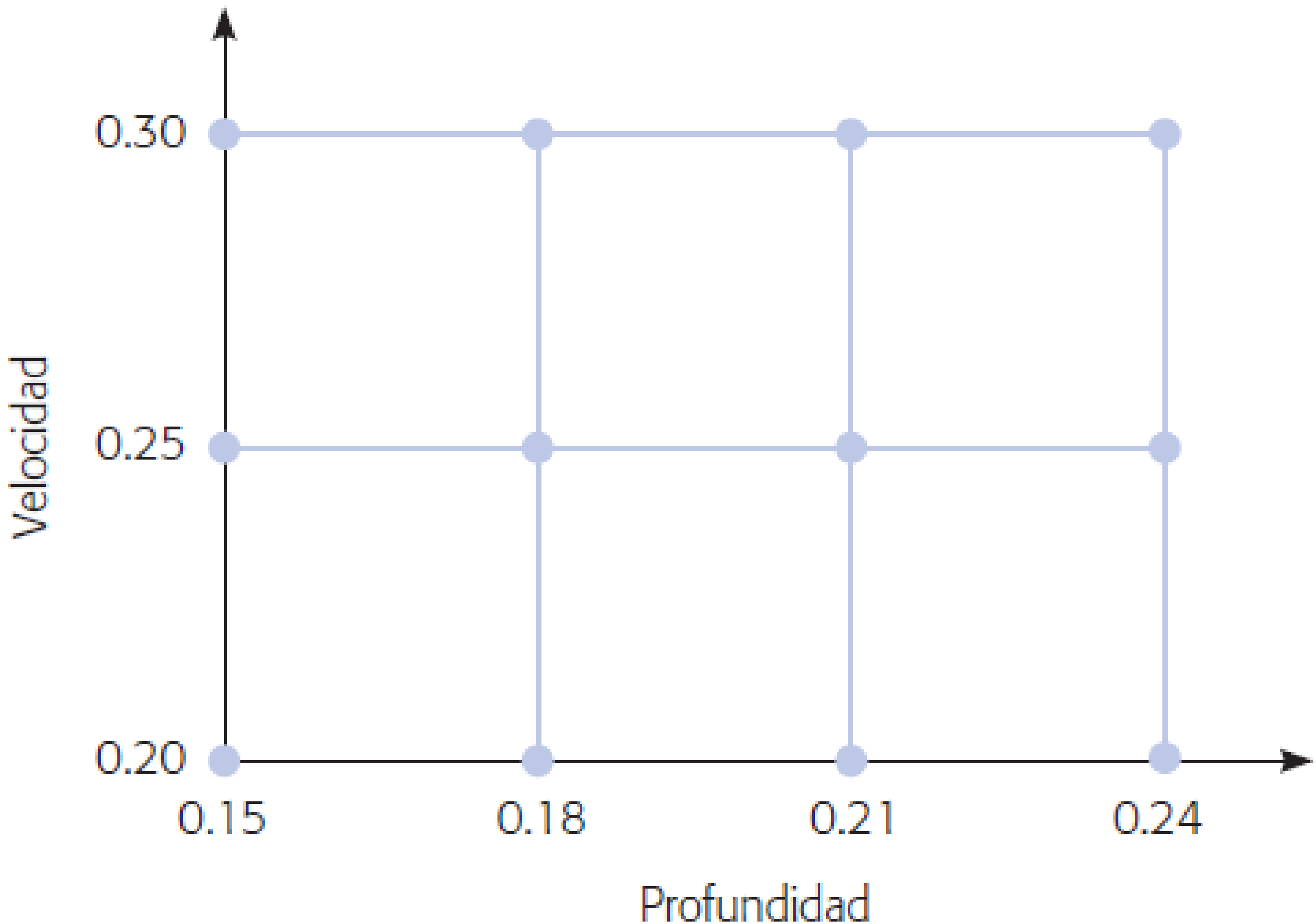


Figura 5.3 Representación del diseño factorial 4 × 3.

$$H_0 : \text{Efecto de profundidad } (A) = 0$$
$$H_A : \text{Efecto de profundidad } (A) \neq 0$$

$$H_0 : \text{Efecto de velocidad } (B) = 0$$
$$H_A : \text{Efecto de velocidad } (B) \neq 0$$

$$H_0 : \text{Profundidad} \times \text{velocidad } (AB) = 0$$
$$H_A : \text{Profundidad} \times \text{velocidad } (AB) \neq 0$$

Tabla 5.5 ANOVA para el ejemplo 5.2.

FV	*SC*	*GL*	*CM*	F_0	Valor-*p*
B: veloc	3 160.5	2	1 580.25	55.02	0.0000
A: prof	2 125.10	3	708.37	24.66	0.0000
AB	557.07	6	92.84	3.23	0.0180
Error	689.33	24	28.72		
Total	6 532.0	35			

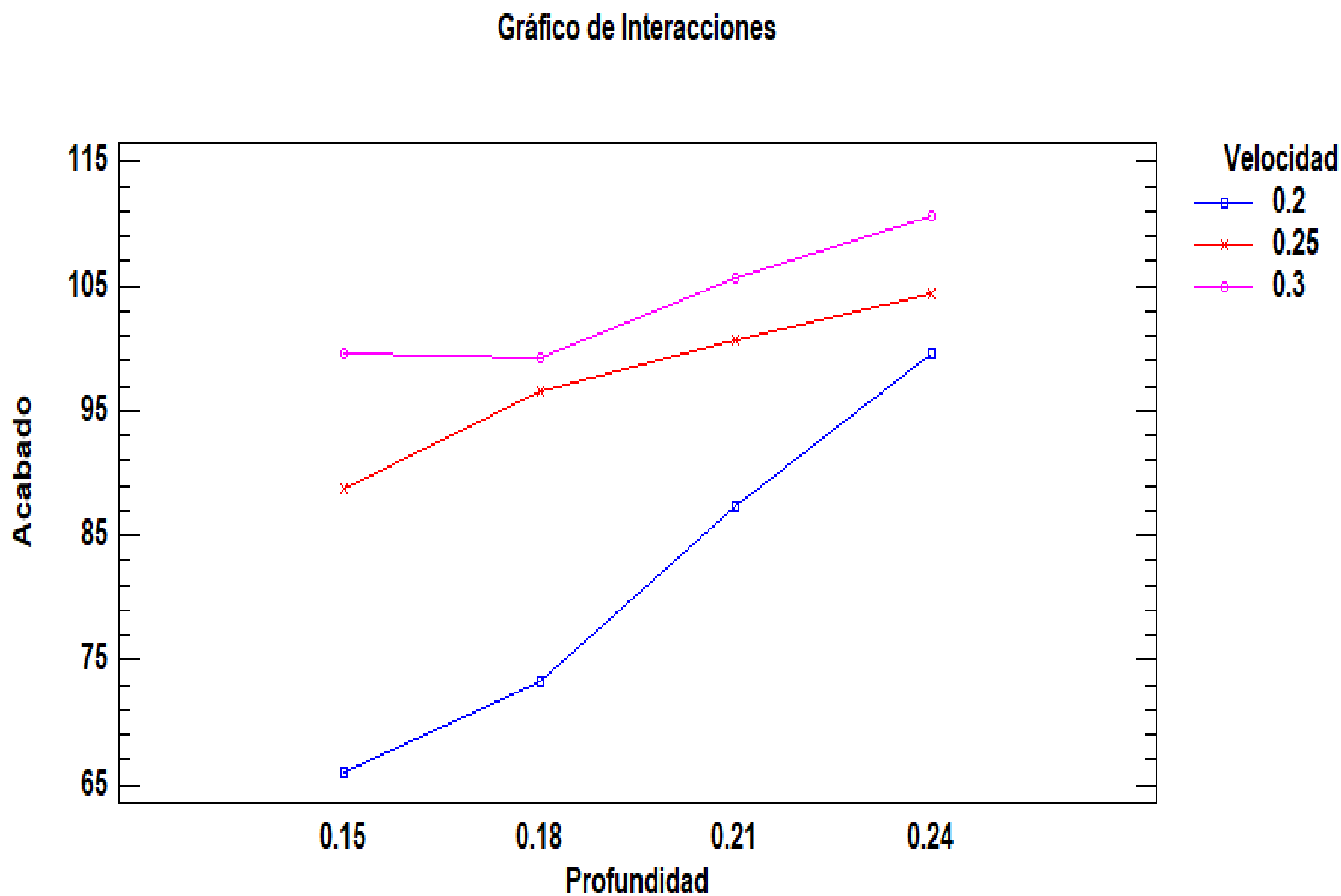


Gráfico de Interacciones
Velocidad
0.2
0.25
0.3
115
105
95
85
75
65
Acabado
0.15
0.18
0.21
0.24
Profundidad

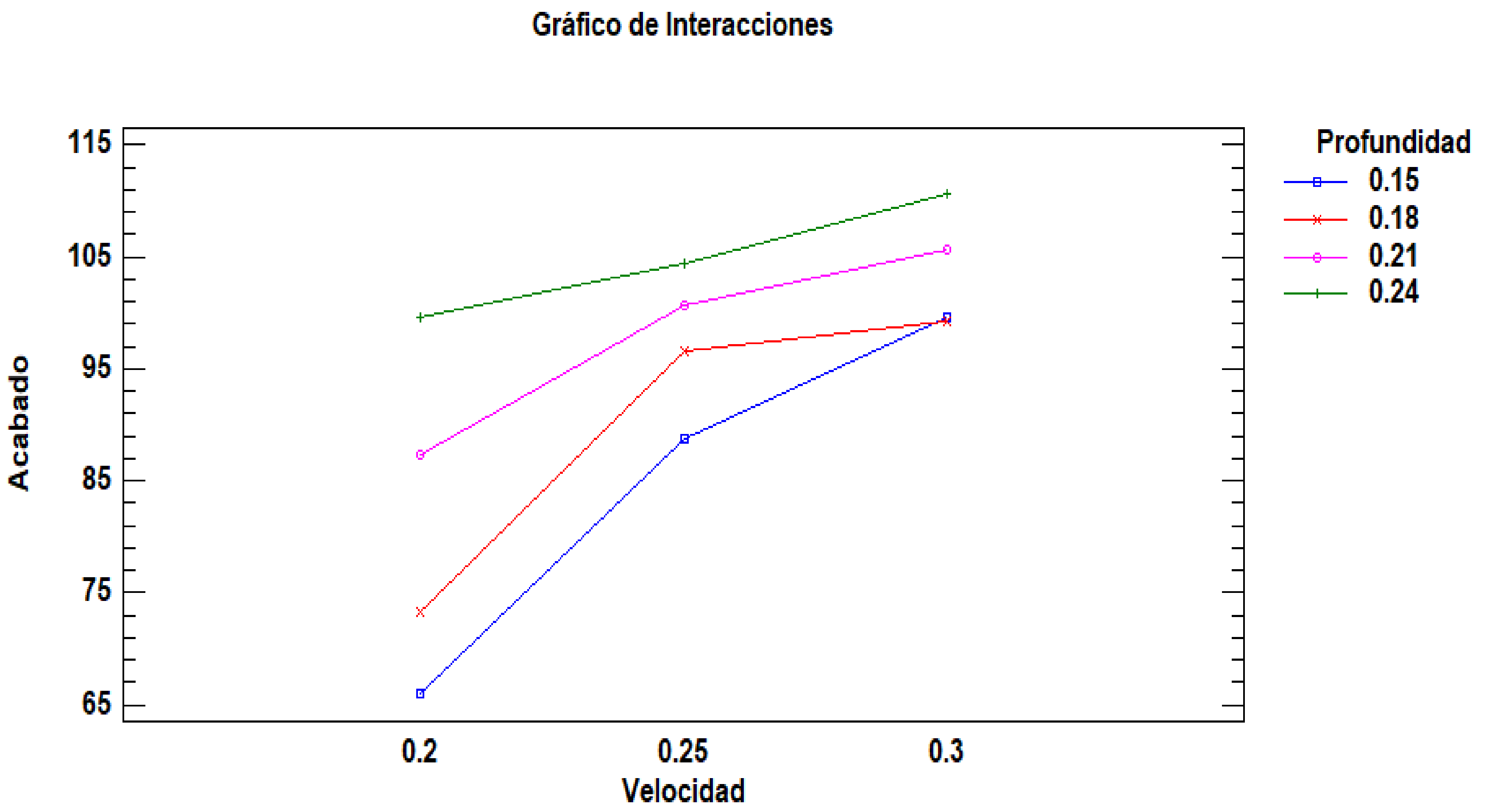


Gráfico de Interacciones
Profundidad
0.15
0.18
0.21
0.24
Acabado
115
105
95
85
75
65
0.2
0.25
0.3
Velocidad

ANOVA Multifactorial - Acabado

Análisis de Varianza para Acabado - Suma de Cuadrados Tipo III

Fuente	*Suma de Cuadrados*	*Gl*	*Cuadrado Medio*	*Razón-F*	*Valor-P*
EFECTOS PRINCIPALES					
A:Profundidad	2125.11	3	708.37	24.66	0.0000
B:Velocidad	3160.5	2	1580.25	55.02	0.0000
INTERACCIONES					
AB	557.056	6	92.8426	3.23	0.0180
RESIDUOS	689.333	24	28.7222		
TOTAL (CORREGIDO)	6532.0	35			

Todas las razones-F se basan en el cuadrado medio del error residual

El StatAdvisor

La tabla ANOVA descompone la variabilidad de Acabado en contribuciones debidas a varios factores. Puesto que se ha escogido la suma de cuadrados Tipo III (por omisión), la contribución de cada factor se mide eliminando los efectos de los demás factores. Los valores-P prueban la significancia estadística de cada uno de los factores. Puesto que 3 valores-P son menores que 0.05, estos factores tienen un efecto estadísticamente significativo sobre Acabado con un 95.0% de nivel de confianza.

Pruebas de Múltiple Rangos para Acabado por Profundidad

Método: 95.0 porcentaje Tukey HSD

Profundidad	*Casos*	*Media LS*	*Sigma LS*	*Grupos Homogéneos*
0.15	9	84.7778	1.78644	X
0.18	9	89.7778	1.78644	X
0.21	9	97.8889	1.78644	X
0.24	9	104.889	1.78644	X

Contraste	*Sig.*	*Diferencia*	*+/- Límites*
0.15 - 0.18		-5.0	6.97114
0.15 - 0.21	*	-13.1111	6.97114
0.15 - 0.24	*	-20.1111	6.97114
0.18 - 0.21	*	-8.11111	6.97114
0.18 - 0.24	*	-15.1111	6.97114
0.21 - 0.24	*	-7.0	6.97114

* indica una diferencia significativa.

El StatAdvisor

Esta tabla aplica un procedimiento de comparación multiple para determinar cuáles medias son significativamente diferentes de otras. La mitad inferior de la salida muestra las diferencias estimadas entre cada par de medias. El asterisco que se encuentra al lado de los 5 pares indica que estos pares muestran diferencias estadisticamente significativas con un nivel del 95.0% de confianza. En la parte superior de la página, se han identificado 3 grupos homogéneos según la alineación de las X's en columnas. No existen diferencias estadisticamente significativas entre aquellos niveles que compartan una misma columna de X's. El método empleado actualmente para discriminar entre las medias es el procedimiento de diferencia honestamente significativa (HSD) de Tukey. Con este método hay un riesgo del 5.0% al decir que uno o más pares son significativamente diferentes, cuando la diferencia real es igual a 0.

Pruebas de Múltiple Rangos para Acabado por Velocidad

Método: 95.0 porcentaje Tukey HSD

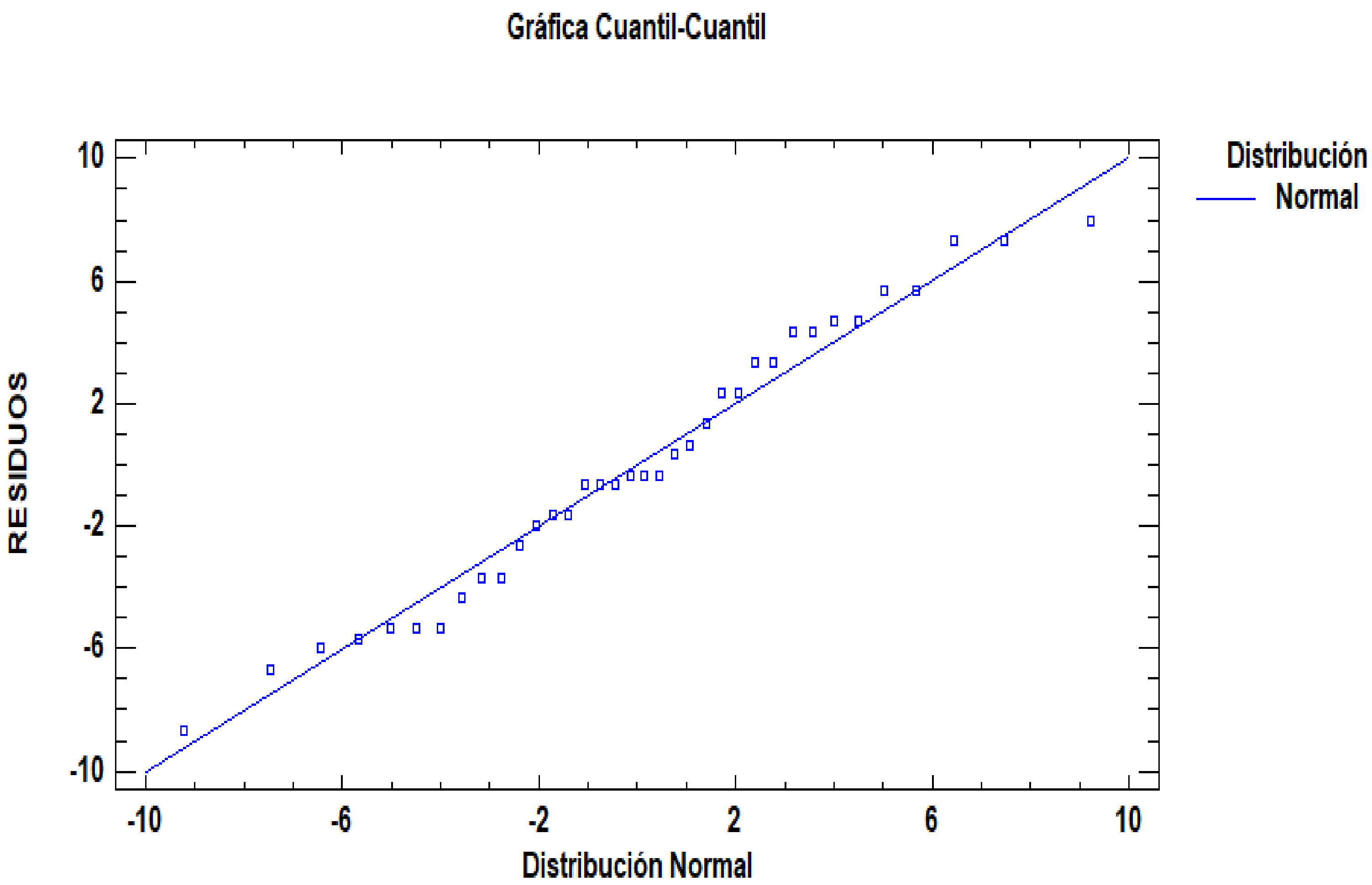
Velocidad	*Casos*	*Media LS*	*Sigma LS*	*Grupos Homogéneos*
0.2	12	81.5833	1.5471	X
0.25	12	97.5833	1.5471	X
0.3	12	103.833	1.5471	X

Contraste	*Sig.*	*Diferencia*	*+/- Límites*
0.2 - 0.25	*	-16.0	5.46563
0.2 - 0.3	*	-22.25	5.46563
0.25 - 0.3	*	-6.25	5.46563

* indica una diferencia significativa.

El StatAdvisor

Esta tabla aplica un procedimiento de comparación multiple para determinar cuáles medias son significativamente diferentes de otras. La mitad inferior de la salida muestra las diferencias estimadas entre cada par de medias. El asterisco que se encuentra al lado de los 3 pares indica que estos pares muestran diferencias estadísticamente significativas con un nivel del 95.0% de confianza. En la parte superior de la página, se han identificado 3 grupos homogéneos según la alineación de las X's en columnas. No existen diferencias estadísticamente significativas entre aquellos niveles que compartan una misma columna de X's. El método empleado actualmente para discriminar entre las medias es el procedimiento de diferencia honestamente significativa (HSD) de Tukey. Con este método hay un riesgo del 5.0% al decir que uno o más pares son significativamente diferentes, cuando la diferencia real es igual a 0.



Gráfica Cuantil-Cuantil
Distribución
Normal
RESIDUOS
10
6
2
-2
-6
-10
-10
-6
-2
2
6
10
Distribución Normal

Ajuste de Datos No Censurados - RESIDUOS

Pruebas de Normalidad para RESIDUOS

Prueba	*Estadístico*	*Valor-P*
Estadístico W de Shapiro-Wilk	0.96462	0.373697

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si RESIDUOS puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es mayor ó igual a 0.05, no se puede rechazar la idea de que RESIDUOS proviene de una distribución normal con 95% de confianza.

Pruebas de Bondad-de-Ajuste para RESIDUOS

Prueba de Kolmogorov-Smirnov

	Normal
DMAS	0.0854946
DMENOS	0.0855751
DN	0.0855751
Valor-P	0.954684

D de Kolmogorov-Smirnov Modificada

	Normal
D	0.0855751
Forma Modificada	0.524718
Valor-P	>=0.10*

Kuiper V

	Normal
V	0.17107
Forma Modificada	1.05835
Valor-P	>=0.10*

Cramer-Von Mises W^2

	Normal
W^2	0.0469699
Forma Modificada	0.0476223
Valor-P	0.54575*

Watson U^2

	Normal
U^2	0.0468527

Forma Modificada	1.05835
Valor-P	>=0.10*

Cramer-Von Mises W^2

	Normal
W^2	0.0469699
Forma Modificada	0.0476223
Valor-P	0.54575*

Watson U^2

	Normal
U^2	0.0468527
Forma Modificada	0.0475035
Valor-P	0.505108*

Anderson-Darling A^2

	Normal
A^2	0.325885
Forma Modificada	0.33324
Valor-P	0.509686*

*Indica que el Valor-P se ha comparado con tablas de valores críticos especialmente construidas para ajustar la distribución seleccionada. Otros valores-P están basados en tablas generales y pueden ser muy conservadores (excepto para la Prueba de Chi-Cuadrada).

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si RESIDUOS puede modelarse adecuadamente con una distribución normal.

Debido a que el valor-P más pequeño de las pruebas realizadas es mayor ó igual a 0.05, no se puede rechazar la idea de que RESIDUOS proviene de una distribución normal con 95% de confianza.

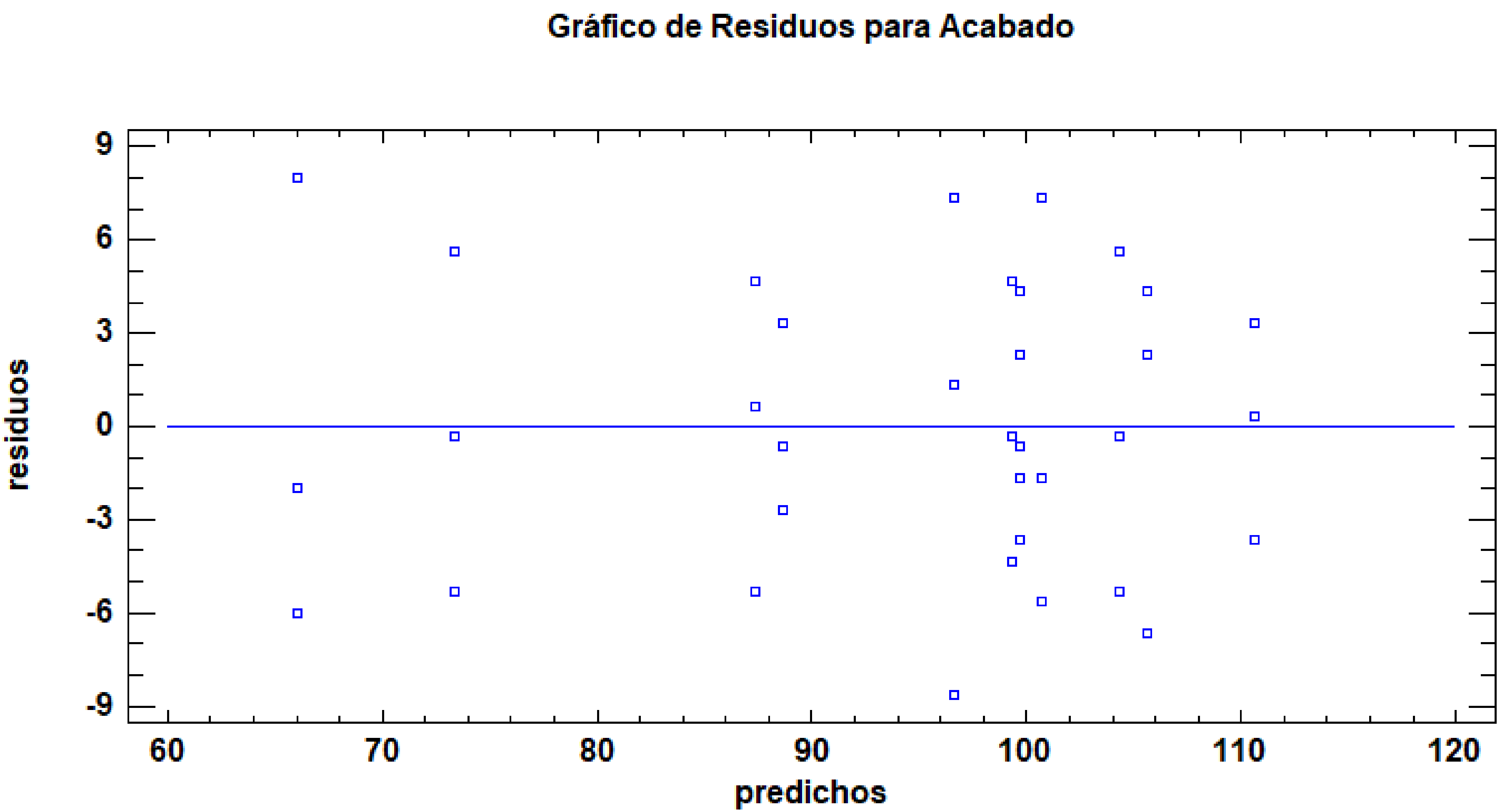


Gráfico de Residuos para Acabado
residuos
9
6
3
0
-3
-6
-9
60
70
80
90
100
110
120
predichos

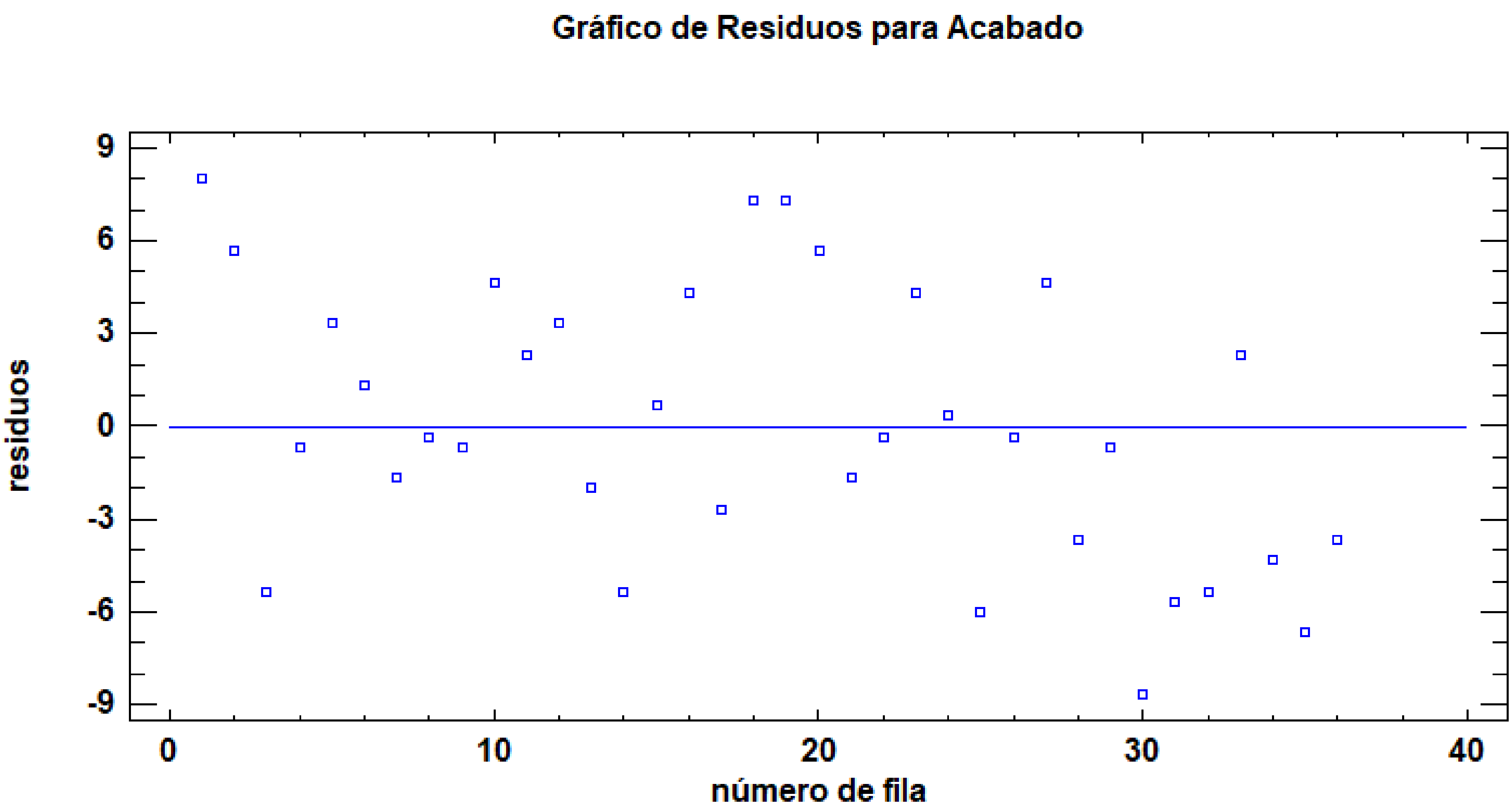


Gráfico de Residuos para Acabado
9
6
3
0
-3
-6
-9
residuos
0
10
20
30
40
número de fila

Diseños factoriales con tres factores

Modelo estadístico

En un diseño factorial $a \times b \times c$ como el del ejemplo, se supone que el comportamiento de la respuesta Y puede describirse mediante el modelo de efectos dado por:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl};$$

$$i = 1, 2, \ldots, a; \; j = 1, 2, \ldots, b; \; k = 1, 2, \ldots, c; \; l = 1, 2, \ldots, n$$

donde μ es la media general, α_i es el efecto del nivel i-ésimo del factor A, β_j es el efecto del nivel j del factor B y γ_k es el efecto del nivel k en el factor C; $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$ y $(\beta\gamma)_{jk}$ representan efectos de interacción dobles (de dos factores) en los niveles ij, ik, jk, respectivamente, y $(\alpha\beta\gamma)_{ijk}$ es el efecto de interacción triple en la combinación o punto ijk; ε_{ijkl} representa el error aleatorio en la combinación $ijkl$ y l son las repeticiones o réplicas del experimento. Todos los efectos cumplen la restricción de sumar cero, es decir, son desviaciones relacionadas con la media general μ. De manera alternativa, se tiene el modelo de regresión dado por

$$Y_{ijkl} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2j} + \beta_3 X_{3k} + \beta_{12} X_{1i} X_{2j} + \beta_{13} X_{1i} X_{3k} + \beta_{23} X_{2j} X_{3k} + \beta_{123} X_{1i} X_{2j} X_{3k} + \varepsilon_{ijkl};$$

$$i = 1, 2, \ldots, a; \; j = 1, 2, \ldots, b; \; k = 1, 2, \ldots, c; \; l = 1, 2, \ldots, n$$

Calculo de contrastes en arreglos factoriales

Métodos para calcular contrastes.

- La *tabla de signos* se construye a partir de la matriz de diseño, multiplicando las columnas que intervienen en la interacción que se quiera calcular.
- Por ejemplo, si se quiere obtener el contraste de la interacción doble *AB*, se multiplica la columna de signos *A* por la columna *B*, y el resultado son los signos de contraste *AB*.
- Esto se muestra en la siguiente tabla de signos para el diseño factorial 2^2.

A	*B*	*AB*	Yates
−	−	+	(1)
+	−	−	*a*
−	+	−	*b*
+	+	+	*ab*

Tabla 2. Diseño factorial completo 2^3 y constraste ABC

A	B	C	ABC
-1	-1	-1	$-$
1	-1	-1	$+$
-1	1	-1	$+$
1	1	-1	$-$
-1	-1	1	$+$
1	-1	1	$-$
-1	1	1	$-$
1	1	1	$+$

Tabla 3. Dos posibles diseños fraccionados 2^{3-1}							
Fracción 1 $(I = +ABC)$				Fracción 2 $(I = -ABC)$			
A	B	C		A	B	C	
1	-1	-1	a	-1	-1	-1	(1)
-1	1	-1	b	1	1	-1	ab
-1	-1	1	c	1	-1	1	ac
1	1	1	abc	-1	1	1	bc

Tabla 8.2 Diseño factorial completo 2^3 y contraste *ABC*.

A	*B*	*C*	⇒	*ABC*
–1	–1	–1		–
1	–1	–1		+
–1	1	–1		+
1	1	–1		–
–1	–1	1		+
1	–1	1		–
–1	1	1		–
1	1	1		+

Estructura de alias del diseño 2^{3-1} con $I = ABC$. Al estimar los efectos potencialmente importantes con cualquiera de las fracciones dadas en la tabla 8.3, resulta que cada efecto estimado tiene un alias. Consideremos, por ejemplo, la fracción 1 de la tabla 8.3. Este diseño se generó con $I = +ABC$, que en este caso también es

Tabla 8.3 Dos posibles diseños fraccionados 2^{3-1}.

Fracción 1 ($I = +ABC$)				Fracción 2 ($I = -ABC$)			
A	*B*	*C*		*A*	*B*	*C*	
1	–1	–1	*a*	–1	–1	–1	(1)
–1	1	–1	*b*	1	1	–1	*ab*
–1	–1	1	*c*	1	–1	1	*ac*
1	1	1	*abc*	–1	1	1	*bc*