Planejamento e Análise Estatística de Experimentos fatoriais 2^k: análise de dados de experimentos completamente aleatorizados

Prof. Caio Azevedo

Contexto

- Vimos como delinear e analisar experimentos (completamente aleatorizados e em blocos) envolvendo 2 ou mais fatores.
- Em muitos casos, pode-se ter interesse em muitos fatores (4, 5 ou mais).
- Quanto maior o número de níveis por fator, maior o número de tratamentos e consequentemente, mais unidades experimentais são necessárias e mais complexa se torna a análise.

Cont.

- Em muitos desse casos pode-se, primeiramente, considerar apenas dois níveis de cada fator (os principais níveis ou simplesmente a ausência e presença de cada fator), para se ter uma idéia de quais fatores são relevantes.
- Em outros casos, simplestemente tem-se interesse em apenas dois níveis de cada fator (dois que são considerados mais importantes ou ausência e presença do fator).
- Essencialmente, todas as metodologias vistas para análise de experimentos fatoriais, podem ser aplicadas.

Cont.

- Contudo, algumas facilidade (particularidades) surgem quando se considera apenas dois níveis por fator.
- Os fatores podem ser qualitativos ou quantitativos.

Exemplo 11: efeitos de concentração de reagente e quantidade de catalisador

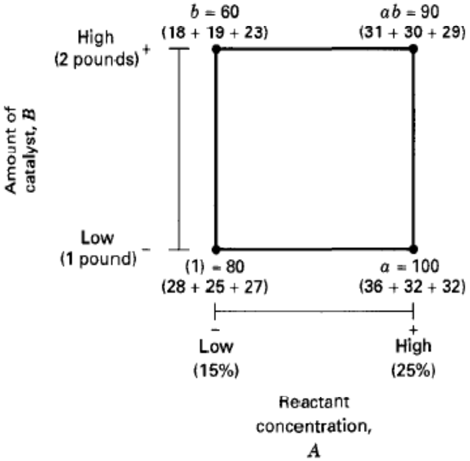
- Suponha que um pesquisador esteja interessado em estudar o efeitos das concentrações de um reagente e da quantidade de catalisador em um determinado processo químico.
- A resposta é a “produção relativa à esse processo químico”.
- Fator A (reagente): dois níveis, 15% e 25%.
- Fator B (catalisador): dois níveis, 1 libra e 2 libras.
- Foram feitas três repetições por tratamento.

Resultados relativos ao experimento (exemplo 11)

Fator		Tratamento	Repetição			Total
A	**B**		**1**	**2**	**3**	
-	-	A 15, B 1	28	25	27	80
+	-	A 25, B 1	36	32	32	100
-	+	A 15, B 2	18	19	23	60
+	+	A 25, B 2	31	30	29	90

Os sinais "-" e "+" denotam, respectivamente, os menores e maiores níveis de cada fator.

Representação gráfica do experimento do exemplo 11



Análise do experimento (exemplo 11)

- Denotaremos por A, B e AB, os efeitos, respectivamente, do fator A, do fator B e da interação. Seja n o número de repetições por tratamento.
- Os efeitos serão definidos pelas diferenças entre médias (como anteriormente).
- Por exemplo, para o fator A:

  $A = \overline{y}_{A^+} - \overline{y}_{A^-} = \frac{1}{2n}[ab + a - b - (1)]$
- Para o fator B:

  $A = \overline{y}_{B^+} - \overline{y}_{B^-} = \frac{1}{2n}[ab + (1) - a - b]$

Análise do experimento (exemplo 11), cont.

- No caso da interação, como antes, temos:

$$AB = (\overline{y}_{A^+,B^+} - \overline{y}_{A^-,B^+}) - (\overline{y}_{A^+,B^-} - \overline{y}_{A^-,B^-}) = \frac{1}{2n}[ab + (1) - a - b]$$

- Voltando ao exemplo, temos:

$$\begin{aligned} A &= \frac{1}{2(3)}[90 + 100 - 60 - 80] = 8,33 \\ B &= \frac{1}{2(3)}[90 + 60 - 100 - 80] = -5,00 \\ AB &= \frac{1}{2(3)}[90 + 80 - 100 - 60] = 1,67 \end{aligned}$$

Análise do experimento (exemplo 11), cont.

- Neste caso, vemos que o fator A tem um efeito crescente (a resposta média aumenta de 15% para 25%), o fator B tem um efeito decrescente (a resposta média diminui de 1 libra para 2 libras) e que a interação apresenta um "impacto" menor quando comparada com os efeitos principais.
- A análise de variância (feita de modo usual) pode ser usada para confirmar a existência de interação como também de efeito dos fatores principais.
- Apesar das fórmulas relacionadas à ANOVA (soma de quadrados, quadrados médios etc) poderem ser utilizadas, há fórmulas mais simples.

Análise do experimento (exemplo 11), cont.

- Note que para o cálculo dos efeitos de cada fator, foram empregados contrastes. Por exemplo, para o fato A, temos $ab + a - b - (1)$
- Em geral, chamamos de contraste do efeito total do fator A.
- O mesmo vale para os outros fatores.
- Pode-se então usar as fórmulas das somas de quadrados dos constrastes como definido para os testes de comparação múltipla $\left(SQ(.) = \frac{\left(\sum_{i=1}^{a} c_i \overline{y}_i\right)^2}{n \sum_{i=1}^{a} c_i^2}\right)$

Análise do experimento (exemplo 11), cont.

- Assim, tem-se

$$
\begin{aligned}
SQ_A &= \frac{1}{4n}[ab + a - b - (1)]^2 = 208,33 \\
SQ_B &= \frac{1}{4n}[ab + b - a - (1)]^2 = 75,00 \\
SQ_{AB} &= \frac{1}{4n}[ab + (1) - a - b]^2 = 8,33
\end{aligned}
$$

- Tais fórmulas podem ser úteis quando se analisa bancos de dados com milhares de informações.
- Pode-se calcular SQT e SQR das forma usuais.

Análise do experimento (exemplo 11), cont.

- Em relação aos contrastes, temos a seguinte tabela:

Efeito	**(1)**	**a**	**b**	**ab**
A:	-1	+ 1	-1	+1
B:	-1	- 1	+1	+1
AB:	+1	- 1	-1	+1

Contribuição de cada fator na explicação da var. dos dados

Fator	**Estimativa do efeito**	SQ	Contribuição %
A	8,33	208,33	64,44
B	-5,00	75,00	23,22
AB	1,67	8,33	2,58
Resíduo		31,33	
Total		323,00	

Exercício: construir a Tabela ANOVA e verificar que a interação não é significativa mas há efeitos dos fatores principais. Fazer a análise residual do modelo utilizado.

Modelo de regressão

- Como os efeitos dos fatores são significativos e ambos são quantitivos podemos ajustar um modelo de regressão.
- Como a interação não foi significativa, podemos considerar um modelo com intercepto e duas covariáveis:

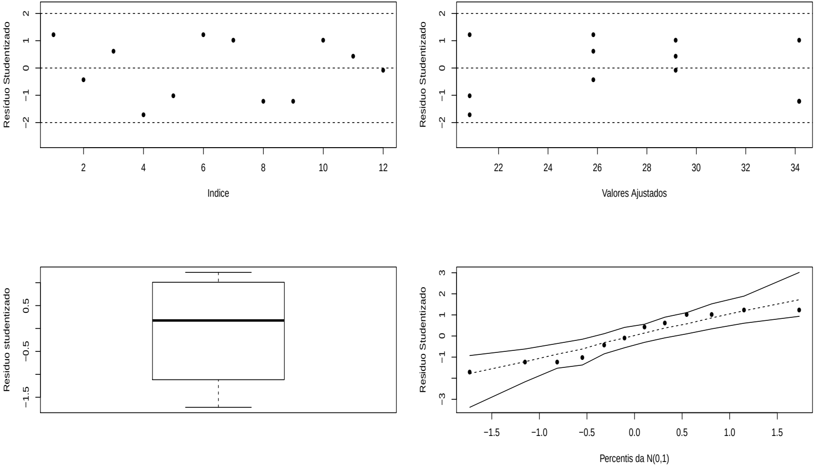
  $$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \xi_i$$

  com as suposições usuais, em que

  $$x_1 = \frac{reagente - (reagente_{baixo} + reagente_{alto})/2}{(reagente_{alto} - reagente_{baixo})/2} = \frac{reagente - 20}{5}$$

  $$x_2 = \frac{catalisador - (catalisador_{baixo} + catalisador_{alto})/2}{(catalisador_{alto} - catalisador_{baixo})/2} = \frac{catalisador - 1{,}5}{0{,}5}$$

Análise residual do modelo de regressão



Modelo de regressão (resultados)

- As estimativas com os erros-padrão são $(\beta_0, \beta_1, \beta_2)$: 27,50 (0,61);4,17 (0,61); -2,50 (0,61)
- Vamos graficar o modelo de regressão em termos das variáveis originais, ou seja:

$$
\begin{aligned}
\widetilde{y} &= 27,50 + 4,17 * \left(\frac{reagente - 20}{5}\right) - 2,50\left(\frac{catalisador - 1,50}{0,50}\right) \\
\widetilde{y} &= 18,33 + 0,83 reagente - 5,00 catalisador
\end{aligned}
$$

Superfície produzida pelo modelo de regressão

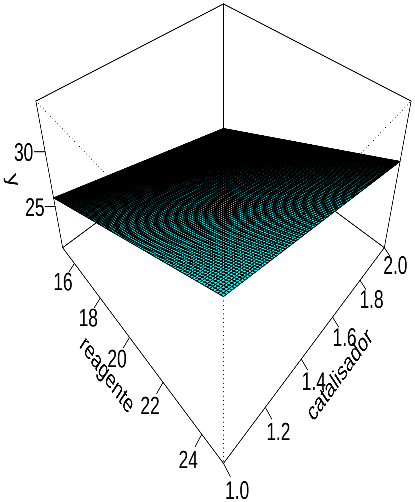
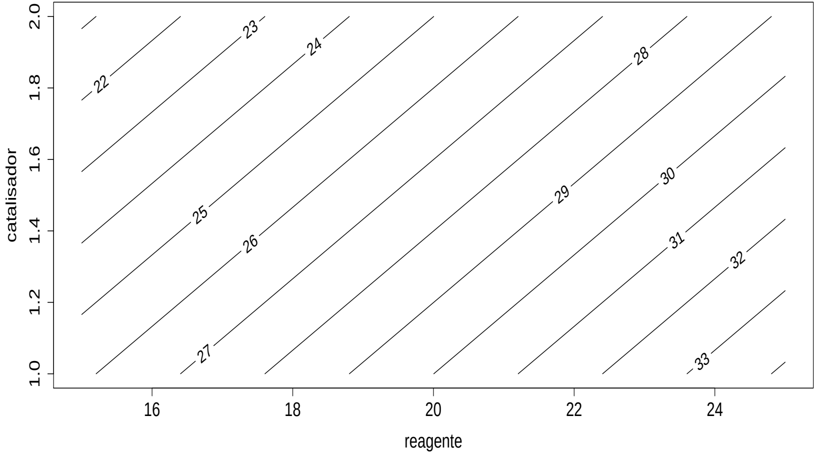


Gráfico de contorno produzida pelo modelo de regressão



Exemplo 12: exemplo engarrafamento de refrigerante (somente dois níveis de cada fator)

- Uma empresa está interessada que a quantidade de refrigerante colocada em cada garrafa seja mais uniforme entre os vasilhames.
- Fatores de interesse:
  - Percentual de carbonatação (CARB): 10% e 12% .
  - Pressão de operação no enchimento (PRE): 25 e 30 psi.
  - Velocidade na linha de produção (VELOC): 200 e 250 bpm.

Exemplo 12: cont.)

- Para cada combinação dos fatores (temos um total de $2 \times 2 \times 3 = 12$ tratamentos) foram medidas as diferenças entre o a quantidade de refrigerante inserida no vasilhame menos o valor padrão de dois refrigerantes escolhidos ao acaso.

Resultados relativos ao experimento (exemplo 12)

Fator			Repetição		Total
A	**B**	**C**	**1**	**2**	
10	25	200	-3	-1	-4 = (1)
10	25	250	-1	0	-1 = c
10	30	200	-1	0	-1 = b
10	30	250	1	1	2 = bc
20	25	200	0	1	1 = a
20	25	250	2	1	3 = ac
20	30	200	2	3	5 = ab
20	30	250	6	5	11 = abc

Os sinais "-" e "+" denotam, respectivamente, os menores e maiores níveis de cada fator.

Representação gráfica do experimento do exemplo 12

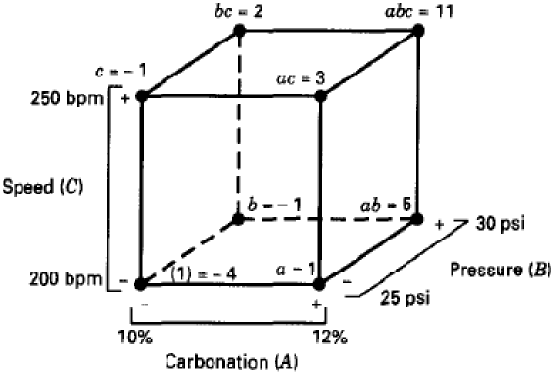


Figure 6-6 The 2^3 design for the fill height deviation experiment for Example 6-1.

Análise do experimento (exemplo 12), cont.

- Em relação aos contrastes, temos a seguinte tabela:

Tratamento				**Coeficientes**				
	I	**A**	**B**	**AB**	**C**	**AC**	**BC**	**ABC**
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

- Com excessão da 1a. coluna, todas as demais tem o mesmo número de sinais positivos e negativos ;
- A soma dos produtos dos sinais em quaisquer duas colunas é zero;
- A coluna I multiplicada por qualquer outra coluna deixa esta inalterada, isto é, a coluna I é um elemento identidade.
- O produto de qualquer duas colunas produz uma coluna da tabela, por exemplo, A x B = AB, e AB x B = AB^2 = A

Estimação dos efeitos dos fatores

- Fator A:

$$\begin{aligned} A &= \overline{y}_{A^+} - \overline{y}_{A^-} = \frac{1}{4n}[a - (1) + ab - b + ac - c + abc - bc] \\ &= \frac{1}{8}[1 - (-4) + 5 - (-1) + 3(-1) + 11 - 2] = 3,00 \end{aligned}$$

- De maneira análoga, obtemos os seguintes valores para os demais efeitos
  $B = 2,25$; $C = 1,75$; $AB = 0,75$; $AC = 0,25$; $BC = 0,50$; $ABC = 0,50$

Somas de Quadrados

- De maneira análoga ao exemplo 11, as somas de quadrados dos fatores e interações podem ser calculadas como: $SQ(.) = \frac{(contraste)^2}{8n}$
- Dessa forma, temos: $SQ_A = 36,00; SQ_B = 20,25; SQ_C = 12,25; SQ_{AB} = 2,25; SQ_{AC} = 0,25; SQ_{BC} = 1,00; SQ_{ABC} = 1,00$

Contribuição de cada fator na explicação da var. dos dados

Fator	Efeito	SQ	Contribuição %
A	3,00	36,00	46,15
B	2,25	20,25	25,96
C	2,75	12,25	15,71
AB	0,75	2,25	2,88
AC	0,25	0,25	0,32
BC	0,50	1,00	1,28
ABC	0,50	1,00	1,28
Resíduo		5,00	6,41
Total		78,00	

Continuação

Exercício: construir a Tabela ANOVA e verificar que a interação não é significativa mas há efeitos dos fatores principais. Fazer a análise residual do modelo utilizado. Verificar que a interação de segunda ordem e as interações de primeira ordem (AC) e (BC) não são significativas.

Modelo de regressão

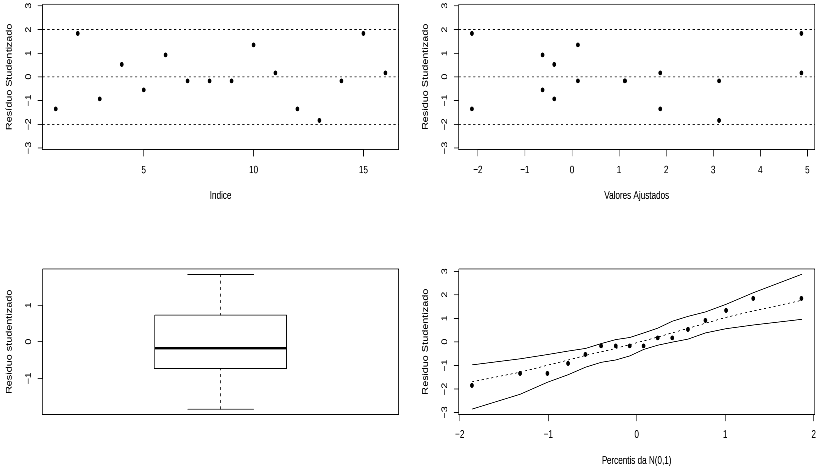
- Como os efeitos dos fatores são significativos e quantitivos podemos ajustar um modelo de regressão.
- Como a interação não foi significativa, podemos considerar o seguinte modelo:

  $$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i1}\xi_{i2} + \xi_i$$

  com as suposições usuais, em que

  $$x_1 = \frac{carb - (carb_{baixo} + carb_{alto})/2}{(carb_{alto} - carb_{baixo})/2} = \frac{carb - 11}{1}$$
  $$x_2 = \frac{pre - (pre_{baixo} + pre_{alto})/2}{(pre_{alto} - pre_{baixo})/2} = \frac{pre - 27{,}50}{2{,}5}$$
  $$x_3 = \frac{veloc - (veloc_{baixo} + veloc_{alto})/2}{(veloc_{alto} - veloc_{baixo})/2} = \frac{veloc - 225}{25}$$

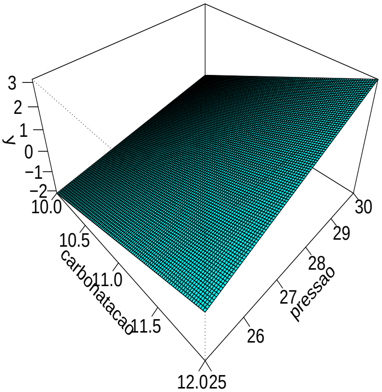
Análise residual do modelo de regressão

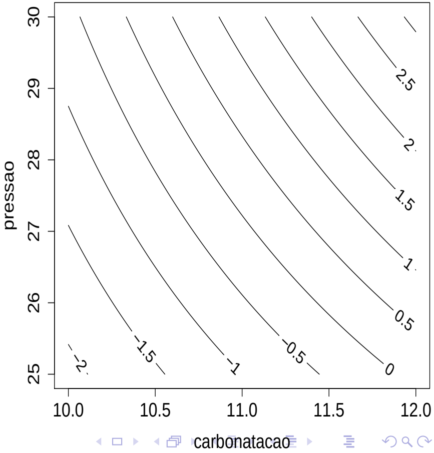


Modelo de regressão (resultados)

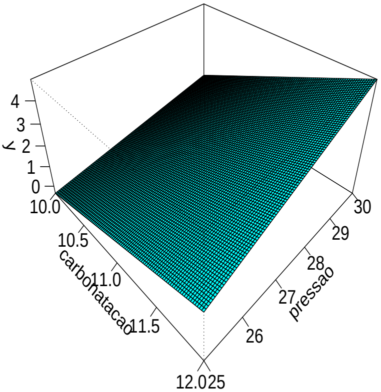
- As estimativas com os erros-padrão são $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$: 1,00 (0,20); 1,50 (0,20); 1,13 (0,20); 0,88 (0,20); 0,38 (0,2).
- Vamos graficar o modelo de regressão em termos das variáveis originais. (exercício: escrever o modelo em termos das variáveis originais).

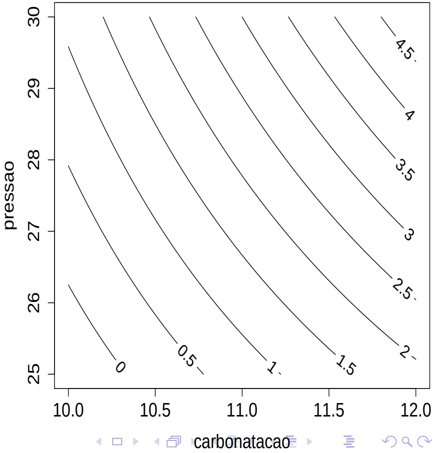
Superfície e gráficos de contorno para Velocidade = 200 bpm





Superfície e gráficos de contorno para Velocidade = 250 bpm





Comentários

- Ver em detalhes a Figura 6-5, página 229, do livro do Montgomery.

Experimentos 2^k com uma única réplica por tratamento

- Mesmo em experimentos do tipo 2^k, quando k aumenta, o número de tratamentos tende a aumentar exponencialmente.
- Por exemplo, para k = 4, tem-se 16 tratamentos, k = 5, tem -se 32 e k = 6, tem-se 64 tratamentos.
- Praticamente impossível avaliar a signficância dos efeitos (fatores principais e interação), através do modelo (ANOVA) sem ter pelo menos duas repetições por tratamento.
- Com apenas uma observação por tratamento, em ajustando-se o modelo completo restará 0 graus de liberdade para o resíduo.

Experimentos 2^k com uma única réplica por tratamento (cont.)

- Alternativas
  - O pesquisador assume a responsabilidade de desconsiderar certos efeitos (por exemplo, algumas interações) do modelo. O problema é que tais efeitos serão contabilizados nos resíduos e, dessa forma, conclusões errôneas podem ser obtidas.
  - Usar alguma estatística (como a estimativa dos efeitos visto anteriormente) para que, de um modo descritivo, se avalie a magnitude dos efeitos.
  - Usar outros modelos (várias opções).

Experimentos 2^k com uma única réplica por tratamento

- Daniel, em 1959, desenvolveu uma técnica para se avaliar a magnitude dos efeitos.
- Dada as suposições do modelo, espera-se que os estimadores de cada efeito (baseados em constrastes) tenham distribuição normal com uma certa média e variância σ^2.
- Isto se deve ao fato de que os estimadores são combinações lineares de médias.
- Assim, se um efeito não é significativo, espera-se que a distribuição do estimador acima tenha média próxima à zero.

Exemplo 13: Produção de um produto químico num recipiente sob pressão.

- Um produto químico é produzido num recipiente sob pressão. Esse experimento foi realizado com fatores que provavelmente influenciam a taxa de filtração do produto. Os quatro fatores considerados em estudo foram: A (temperatura), B (pressão), C (concentração de formaldeído) e D (taxa de agitação).

Exemplo 13: Produção de um produto químico num recipiente sob pressão (cont.).

- Os 16 experimentos (um para cada tratamento) foram realizados em ordem aleatória. O engenheiro está interessado em maximizar a taxa de filtração. O processo atual apresenta uma taxa de filtração em torno de 75 gal/h. O processo também utiliza o fator C no nível alto. Deseja-se reduzir a concentração de formaldeído tanto quanto possível. Porém, isso causa uma diminuição na taxa de filtração.

Experimento	Fator				"Label"	Taxa de filtração (gal/h)
	A	B	C	D		
1	-	-	-	-	(1)	45
2	+	-	-	-	a	71
3	-	+	-	-	b	48
4	+	+	-	-	ab	65
5	-	-	+	-	c	68
6	+	-	+	-	ac	60
7	-	+	+	-	bc	80
8	+	+	+	-	abc	65
9	-	-	-	+	d	43
10	+	-	-	+	ad	100
11	-	+	-	+	bd	45
12	+	+	-	+	abd	104
13	-	-	+	+	cd	75
14	+	-	+	+	acd	86
15	-	+	+	+	bcd	70
16	+	+	+	+	abcd	96

Representação gráfica do experimento do exemplo 13

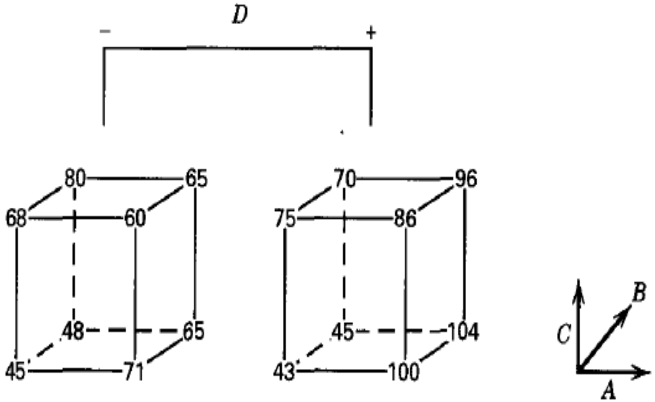


Tabela de contrastes do fatorial 2^4

	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
(1)	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+
a	+	-	-	-	-	+	+	-	-	+	+	+	+	-	-
b	-	+	-	-	+	-	+	-	+	-	+	+	-	+	-
ab	+	+	+	-	-	-	-	-	-	-	-	+	+	+	+
c	-	-	+	+	-	-	+	-	+	+	-	-	+	+	-
ac	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+
bc	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
abc	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
d	-	-	+	-	+	+	-	+	-	-	+	-	+	+	-
ad	+	-	-	-	-	+	+	+	+	-	-	-	-	+	+
bd	-	+	-	-	+	-	+	+	-	+	-	-	+	-	+
abd	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-
cd	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+
acd	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-
bcd	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Fator	Efeito	SQ	Contribuição %
A	21,63	1870,56	32,64
B	3,13	39,06	0,68
C	9,88	390,06	6,80
D	14,63	855,56	14,93
AB	0,13	0,06	$<$0,01
AC	-18,13	1314,06	22,29
AD	16,63	1105,56	19,29
BC	2,38	22,56	0,39
BD	-0,38	0,56	$<$0,01
CD	-1,13	5,06	0,09
ABC	1,88	14,06	0,24
ABD	4,13	68,06	1,19
ACD	1,63	10,56	0,18
BCD	-2,63	27,56	0,48
ABCD	1,38	7,56	0,13

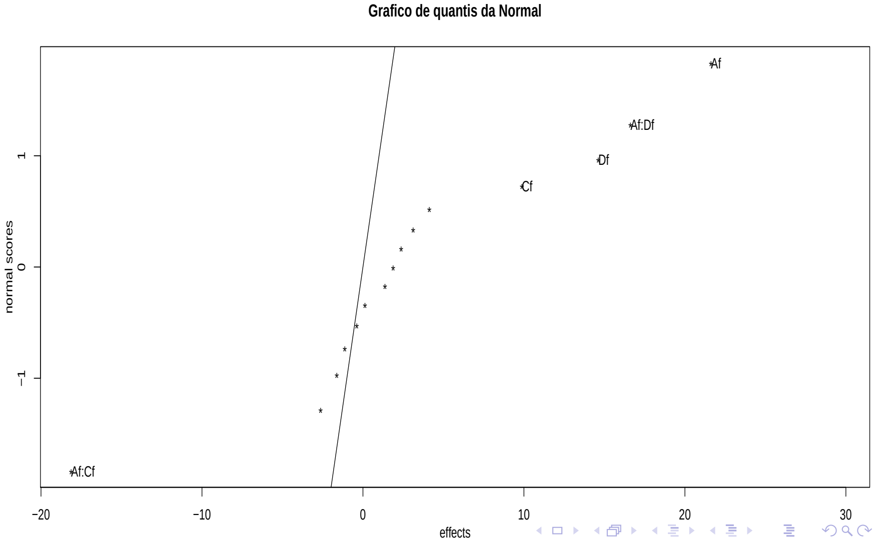


Grafico de quantis da Normal
normal scores
effects
Af
Af:Df
Df
Cf
Af:Cf
1
0
−1
−20
−10
0
10
20
30

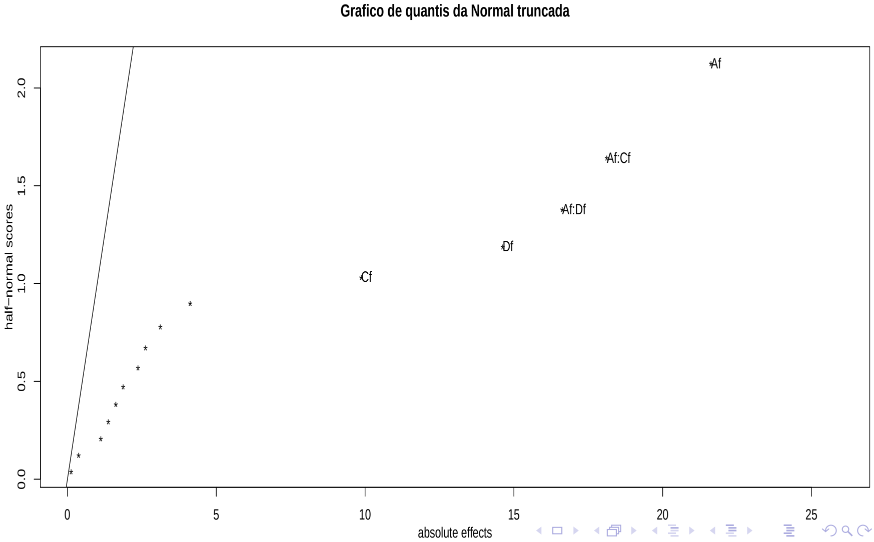


Grafico de quantis da Normal truncada
half−normal scores
absolute effects
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
25
Af
Af:Cf
Af:Df
Df
Cf

Exercício

- Fazer a análise completa (como nas listas de exercícios) considerando somente os fatores significativos (incluindo os gráficos de superfície).
- Fatores significativos (A, C, D, AC, AD).
- Ler o resto do Capítulo 6 do livro do Montgomery.