Projeto com fatorial 2^kr e fatorial fracionado 2^{k-p}

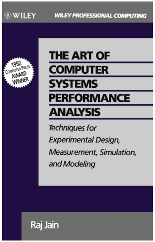
André Gustavo dos Santos

Departamento de Informática
Universidade Federal de Viçosa

INF222 - 2022/2

– Fonte do material

O conteúdo é baseado no livro texto da disciplina (cap. 18 e 19):



Jain, Raj.
The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling.
John Wiley & Sons, 1990

Tópicos da aula

1. Projeto com fatorial $2^2 r$
   - Estimativa de erros experimentais
   - Intervalos de confiança
2. Projeto com fatorial $2^k r$
3. Projeto com fatorial fracionado 2^{k-p}
   - Confounding

Projeto com fatorial 2^kr

- Na aula anterior vimos conceitos e métodos para projeto fatorial 2^k
- Um problema com o fatorial 2^k é que não é possível estimar erros experimentais
- Isso ocorre porque os experimentos não são repetidos
- Então também não é possível calcular intervalos de confiança
- Erros experimentais podem ser estimados repetindo-se as medições para uma mesma combinação de níveis de fatores
- Se cada um dos 2^k experimentos é replicado r vezes, teremos 2^kr observações
- Esse é o projeto 2^kr, fatorial com replicações

Projeto com fatorial $2^2 r$

- Assim cono no 2^k, vamos primeiro estudar o $2^2 r$, depois generalizamos para $2^k r$
- 2 fatores, com 2 níveis, repetidos r vezes para isolar o erros experimentais
- Podemos acrescentar um termo para o erro (e) (como antes, q são os efeitos)

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e$$

- O cálculo dos efeitos (coeficientes) é feito da mesma forma que no projeto 2^k, usando o método da tabela de sinais, mas usando-se a média as observações

Projeto com fatorial 2^2r – computação de efeitos

- Como exemplo do projeto fatorial 2^2 analisamos os efeitos dos fatores memória (A) e cache (B) no desempenho em MIPS de um sistema:

I	A	B	AB	y
$+1$	-1	-1	$+1$	15
$+1$	$+1$	-1	-1	45
$+1$	-1	$+1$	-1	25
$+1$	$+1$	$+1$	$+1$	75
160	80	40	20	Total
40	20	10	5	Total/4
q_0	q_A	q_B	q_{AB}	

- Foi usado o resultado (y) de apenas um experimento por combinação de fatores

- Para 2^2r, com r experimentos, basta usar a média $\bar{y}$ das observações

I	A	B	AB	y	$\bar{y}$
$+1$	-1	-1	$+1$	$(15, 18, 12)$	15
$+1$	$+1$	-1	-1	$(45, 48, 51)$	48
$+1$	-1	$+1$	-1	$(25, 28, 19)$	24
$+1$	$+1$	$+1$	$+1$	$(75, 75, 81)$	77
164	86	38	20		Total
41	21.5	9.5	5		Total/4
q_0	q_A	q_B	q_{AB}		

Projeto com fatorial $2^2 r$ – estimativa de erros experimentais

- Podemos calcular o erro em cada observação:

$$e_{ij} = y_{ij} - \bar{y}_i$$

dos $i = 1, 2, 3, 4$ experimentos e $j = 1, 2, 3$ replicações

Média $\bar{y}$	Observações y_{i1}	y_{i2}	y_{i3}	Erros e_{i1}	e_{i2}	e_{i3}
15	15	18	12	0	3	-3
48	45	48	51	-3	0	3
24	25	28	19	1	4	-5
77	75	75	81	-2	-2	4

- A soma dos erros em cada experimento é zero, portanto a soma total é zero
- Já a soma dos quadrados dos erros (SSE) pode ser usada para estimar a variância dos erros e intervalos de confiança para os efeitos

$$\begin{aligned}\text{SSE} &= \sum_{i=1}^{2^2} \sum_{j=1}^{r} e_{ij}^2 \\ &= (0)^2 + (3)^2 + (-3)^2 + (-3)^2 + (0)^2 + (3)^2 + (1)^2 + (4)^2 + (-5)^2 + (-2)^2 + (-2)^2 + (4)^2 \\ &= 102\end{aligned}$$

UFV

Projeto com fatorial 2^2r – Alocação de variação

- SST no projeto fatorial 2^2r pode ser dividida em quatro termos:

$$\begin{aligned}\text{SST} &= \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSE} \\ &= 2^2rq_A^2 + 2^2rq_B^2 + 2^2rq_{AB}^2 + \sum_{i=1}^{2^2}\sum_{j=1}^{r} e_{ij}^2\end{aligned}$$

- No exemplo:

$$\begin{aligned}\text{SSA} &= 2^2rq_A^2 = 12(21.5)^2 = 5547 \\ \text{SSB} &= 2^2rq_B^2 = 12(9.5)^2 = 1083 \\ \text{SSAB} &= 2^2rq_{AB}^2 = 12(5)^2 = 300 \\ \text{SSE} &= \sum_{i=1}^{2^2}\sum_{j=1}^{r} e_{ij}^2 = 102 \\ \text{SST} &= \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSE} = 5547 + 1083 + 300 + 102 = 7032\end{aligned}$$

- Fator A explica $5547/7032 = 78.88\%$ da variação; o fator B explica $1083/7032 = 15.40\%$ e a interação AB explica $300/7032 = 4.27\%$
- Os restantes 1.45% são inexplicáveis e atribuídos a erros experimentais $(102/7032)$

Projeto com fatorial 2^2r – Intervalos de confiança

- A variância dos erros (MSE - *Mean Square of Errors*) pode ser estimada por

$$s_e^2 = \frac{\text{SSE}}{2^2(r-1)}$$

- O desvio padrão dos efeitos por

$$s_{q_0} = s_{q_A} = s_{q_B} = s_{q_{AB}} = \frac{s_e}{\sqrt{2^2 r}}$$

- E o intervalo de confiança dos efeitos por

$$q_i = \mp t_{[1-\alpha, 2^2(r-1)]} s_{q_i}$$

- $2^2(r-1)$ é o total de graus de liberdade (em cada experimento a soma dos r erros é zero, então apenas $r-1$ podem ser escolhidos de forma independente)

Projeto com fatorial 2^2r – Intervalos de confiança

- No exemplo da memória-cache:

$$s_e = \sqrt{\frac{\text{SSE}}{2^2(r-1)}} = \sqrt{\frac{102}{8}} = 3.57$$

- Temos então:

$$s_{q_i} = \frac{s_e}{\sqrt{s^2 r}} = \frac{3.57}{\sqrt{12}} = 1.03$$

- $2^2(r-1) = 8$ graus de liberdade, confiança de 90%: valor-t da tabela 1.86
- Intervalos de confiança: $q_i \mp t s_{q_i}$

  $q_0 : (39.08, 42.91)$
  $q_A : (19.58, 23.41)$
  $q_B : (7.58, 11.41)$
  $q_{AB} : (3.08, 6.91)$

- Nenhum inclui 0, então todos são significantes, neste nivel de confiança

Projeto com fatorial 2^2r – Intervalos de confiança para contrastes

- Podemos calcular variância e intervalos de confiança para qualquer contraste
- Um contraste é uma combinação linear cujos coeficientes têm soma zero
- A variância de $\sum h_i q_i$ com $\sum h_i = 0$ é

$$s^2_{\sum h_i q_i} = \frac{s_e^2 \sum h_i^2}{2^2 r}$$

- E o intervalo de confiança pode ser calculado usando $t_{[1-\alpha/2, 2^2 r]}$

Projeto com fatorial 2^2r – Estimativa de resposta

- Queremos uma estimativa de resposta (futura) de m repetições de um experimento (mesma combinação de fatores)
- A média da resposta pode ser estimada facilmente, independente de m

$$\bar{y} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B$$

- Quando m aumenta, é esperado que a média amostral se aproxime da estimativa
- Em outras palavras, o desvio padrão da estimativa decresce com m
- Pode ser mostrado que

$$s_{\hat{y}_m} = s_e \sqrt{\frac{1}{n_{eff}} + \frac{1}{m}}$$

- onde

$$n_{eff} = \frac{\text{quantidade total de observações}}{1 + \text{soma dos graus de liberdade dos parâmetros usados em } \hat{y}}$$

- no caso de projeto 2^2r

$$n_{eff} = \frac{2^2 r}{1 + 4}$$

Projeto com fatorial 2^2r – Estimativa de resposta

- Intervalo de confiança

$$\bar{y} \mp t_{[1-\alpha/2, 2^2(r-1)]} s_{\bar{y}_m}$$

sendo

$$s_{\bar{y}_m} = s_e \sqrt{\frac{5}{2^2 r} + \frac{1}{m}}$$

- Casos especiais de interesse:

$m = 1$ $\quad s_{\bar{y}_m} = s_e \sqrt{\frac{5}{2^2 r} + 1}$ (estimativa para um experimento)

$m = \infty$ $\quad s_{\bar{y}_m} = s_e \sqrt{\frac{5}{2^2 r}}$ (estimativa da média populacional)

Projeto com fatorial 2^2r – Estimativa de resposta

Exemplos de intervalos de confiança para resposta média com $x_A = -1$ e $x_B = -1$ no exemplo da memória-cache

- Estimativa de resposta para UM experimento futuro de confirmação
  $\hat{y}_1 = q_0 - q_A - q_B + q_{AB} = 41 - 21.5 - 9.5 + 5 = 15$
  $s_{\hat{y}_1} = s_e\sqrt{\frac{5}{2^2r} + 1} = 3.57\sqrt{\frac{5}{12} + 1} = 4.25$
  $t(0.95, 8) = 1.86$, intervalo de confiança de 90%:
  $15 \mp 1.86 \times 4.25 = (7.09, 22.91)$
- Estimativa de resposta para 5 experimentos futuros
  $\hat{y}_1 = q_0 - q_A - q_B + q_{AB} = 41 - 21.5 - 9.5 + 5 = 15$
  $s_{\hat{y}_1} = s_e\sqrt{\frac{5}{2^2r} + \frac{1}{m}} = 3.57\sqrt{\frac{5}{12} + \frac{1}{5}} = 2.80$
  $t(0.95, 8) = 1.86$, intervalo de confiança de 90%:
  $15 \mp 1.86 \times 2.80 = (9.79, 20.21)$
- Estimativa de resposta para muitos experimentos futuros
  $\hat{y}_1 = q_0 - q_A - q_B + q_{AB} = 41 - 21.5 - 9.5 + 5 = 15$
  $s_{\hat{y}_1} = s_e\sqrt{\frac{5}{2^2r}} = 3.57\sqrt{\frac{5}{12}} = 2.30$
  $t(0.95, 8) = 1.86$, intervalo de confiança de 90%:
  $15 \mp 1.86 \times 2.30 = (10.71, 19.29)$
- Note que o intervalo para a média vai se estreitando com mais replicações

Projeto com fatorial 2^2r – Estimativa de resposta

- Resposta média atual (não é uma previsão)
  $\hat{y}_1 = q_0 - q_A - q_B + q_{AB} = 41 - 21.5 - 9.5 + 5 = 15$
  $s_{\hat{y}_1} = s_e\sqrt{\frac{\sum h_i^2}{2^2r}} = 3.57\sqrt{\frac{1+1+1+1}{12}} = 2.06$
  $t(0.95, 8) = 1.86$, intervalo de confiança de 90%:
  $15 \mp 1.86 \times 2.06 = (11.17, 18.83)$

- Note que o intervalo é mais estreito que os das estimativas de replicações futuras
- Isso ocorre porque, para estimativas, são adicionados os erros das replicações futuras

Projeto com fatorial $2^2 r$ – Premissas

As seguintes premissas foram feitas nas derivações anteriores:

- Erros estatisticamente independentes
- Erros são aditivos
- Erros são normalmente distribuídos
- Erros têm desvio padrão constante
- Efeitos dos fatores são aditivos

Elas indicam que as observações foram independentes, normalmente distribuídas e com variância constante.

- Podem ser verificadas por testes visuais

Modelos multiplicativos

- O modelo usado para analisar experimentos 2^2r foi um modelo aditivo

$$y_{ij} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + eij$$

- Ele assume que os efeitos, suas interações e os erros experimentais são aditivos
- Antes de usá-lo, um analista precisa validar se são realmente aditivos

Contra-exemplo: desempenho de processadores para cargas diferentes

Seja y_{ij} o tempo necessário para executar uma carga de w_j instruções em um processador que executa $1/v_i$ instruções por segundo. Então:

$$y_{ij} = v_i w_j$$

- Não podemos usar o modelo aditivo, porque os efeitos são multiplicativos
- Transformação: $\log(y_{ij}) = \log(v_i) + \log(w_j)$
- Modelo:

$$y'_{ij} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + eij$$

com $y'_{ij} = \log(y_{ij})$

Projeto com fatorial $2^k r$

Os métodos e expressões para projetos $2^k r$ são extensões do $2^2 r$

- Modelo: $y_{ij} = q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi} + \cdots + e_{ij}$
- Efeitos: $q_j = \frac{1}{2^k} \sum_{i=1}^{2^k} S_{ij} \bar{y}_i$, onde S_{ij} é o sinal na tabela
- Soma dos quadrados:
  - SSj $= 2^k r q_j^2$
  - SSE $= \sum_{i=1}^{2^k} \sum_{j=1}^{r} e_{ij}^2$
- Desvio padrão do erro: $s_e = \sqrt{\frac{\text{SSE}}{2^k(r-1)}}$
- Desvio padrão dos efeitos: $s_{q_0} = s_{q_A} = s_{q_B} = s_{q_{AB}} = \cdots = \frac{s_e}{\sqrt{2^k r}}$
- Variância do contraste $\sum h_i q_i$ com $\sum h_i = 0$: $s^2_{\sum h_i q_i} = \frac{s_e^2 \sum h_i^2}{2^k r}$
- Desvio padrão da média de m observações futuras: $s_{\bar{y}_m} = s_e \sqrt{\frac{1+2^k}{2^k r} + \frac{1}{m}}$
- Intervalos de confiança calculados usando $t_{[1-\alpha/2, 2^k(r-1)]}$

detalhes e exemplo podem ser consultados na seção 18.9 do livro texto

Projeto com fatorial fracionado 2^{k-p}

- Se o número de fatores for grande, um projeto fatorial pode ser caro, pois exige muitos experimentos
- Um projeto fatorial fracionado requer consideravelmente menos experimentos
- Um projeto 2^{k-p} analisa k fatores com 2^{k-p} experimentos
- Nesse caso, p é um valor escolhido adequadamente
- Por exemplo, um projeto 2^{k-1} realiza metade dos experimentos do 2^k
- E um projeto 2^{k-2} realiza apenas 1/4 dos experimentos

Projeto com fatorial fracionado 2^{k-p}

- Uma vantagem do projeto fatorial 2^k é a facilidade de se calcular os vários efeitos e suas contribuições na variação total
- Isso ocorre por causa da ortogonalidade mútua das colunas da tabela de sinais
  - A soma de cada coluna é zero: $\sum_i x_{ij} = 0, \forall j$
  - A soma do produto de quaisquer duas colunas é zero: $\sum_i x_{ij}x_{il} = 0, \forall j \neq l$
  - A soma dos quadrados de cada coluna é 2^k: $\sum_i x_{ij}^2 = 2^k, \forall j$
- Assim, efeitos e contribuições na resposta y são calculados por produtos internos

- No projeto fracionado 2^{k-p}, se montarmos a tabela de sinais cuidadosamente para manter a ortogonalidade, poderemos computar efeitos da mesma forma
- O problema é que não é possível isolar os efeitos de cada fator e suas interações
- Requer planejamento cuidadoso para minimizar o “confundimento” (*confounding*)

UFV

Projeto com fatorial fracionado 2^{k-p} – exemplo

- Um projeto 2^k com $k = 7$ fatores requer $2^7 = 128$ experimentos
- Um projeto 2^{7-4} requer apenas $2^3 = 8$ experimentos
- Considere a tabela abaixo, montada para estudar os fatores A, B, C, D, E, F, G

Experiment No.	*A*	*B*	*C*	*D*	*E*	*F*	*G*
1	–1	–1	–1	1	1	1	–1
2	1	–1	–1	–1	–1	1	1
3	–1	1	–1	–1	1	–1	1
4	1	1	–1	1	–1	–1	–1
5	–1	–1	1	1	–1	–1	1
6	1	–1	1	–1	1	–1	–1
7	–1	1	1	–1	–1	1	–1
8	1	1	1	1	1	1	1

- Note que as colunas são mutuamente ortogonais
- Pode ser usada para estudar o modelo

$$y = q_0 + q_A x_A + q_B x_B + q_C x_C + q_D x_D + q_E x_E + q_F x_F + q_G x_G$$

- por exemplo, $q_A = (-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8)/8$
- fórmulas de desvio padrão dos efeitos e intervalos de confiança são calculadas da mesma forma que no 2^k substituindo-se 2^k por 2^{k-p} (e vale com replicação)

Projeto com fatorial fracionado 2^{k-p} – exemplo

- Os $2^{7-4} = 8$ experimentos são executados, com os fatores nos níveis indicados

I	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*y*
1	–1	–1	–1	1	1	1	–1	20
1	1	–1	–1	–1	–1	1	1	35
1	–1	1	–1	–1	1	–1	1	7
1	1	1	–1	1	–1	–1	–1	42
1	–1	–1	1	1	–1	–1	1	36
1	1	–1	1	–1	1	–1	–1	50
1	–1	1	1	–1	–1	1	–1	45
1	1	1	1	1	1	1	1	82
317	101	35	109	43	1	47	3	Total
39.62	12.62	4.37	13.62	537	0.125	5.87	0.37	Total/8

- Os efeitos $q_A, q_B, \ldots, q_G$ são calculados como no fatorial 2^3
- A alocação de variação revela que os fatores A, B, C, D, E, F, G explicam respectivamente 37.26, 4.47, 43.40, 6.75, 0.00, 8.07 e 0.03% da variação
- Se for necessário conduzir mais experimentos, concentrar nos fatores A e C
- Vale lembrar que esse estudo não revela os efeitos da combinação de fatores... mas se a interação for pequena ou desprezível, economizamos muito trabalho

Projeto com fatorial fracionado 2^{k-p} – tabela

Passos para montar a tabela de sinais de um projeto 2^{k-p}

1. Escolha $k-p$ fatores e prepare uma tabela de sinais completa para $k-p$ fatores
   - isso resulta em um tabela com 2^{k-p} linhas e 2^{k-p} colunas
   - a primeira coluna, I, contém somente 1's
   - as próximas $k-p$ são marcadas com os $k-p$ fatores escolhidos
   - as restantes são produtos destes fatores
2. Das $2^{k-p}-(k-p)-1$ colunas da direita, escolha p e marque-as com os p fatores não escolhidos no passo 1

Experiment No.	*A*	*B*	*C*	*AB*	*AC*	*BC*	*ABC*
1	–1	–1	–1	1	1	1	–1
2	1	–1	–1	–1	–1	1	1
3	–1	1	–1	–1	1	–1	1
4	1	1	–1	1	–1	–1	–1
5	–1	–1	1	1	–1	–1	1
6	1	–1	1	–1	1	–1	–1
7	–1	1	1	–1	–1	1	–1
8	1	1	1	1	1	1	1

$\Rightarrow$

Experiment No.	*A*	*B*	*C*	*D*	*E*	*F*	*G*
1	–1	–1	–1	1	1	1	–1
2	1	–1	–1	–1	–1	1	1
3	–1	1	–1	–1	1	–1	1
4	1	1	–1	1	–1	–1	–1
5	–1	–1	1	1	–1	–1	1
6	1	–1	1	–1	1	–1	–1
7	–1	1	1	–1	–1	1	–1
8	1	1	1	1	1	1	1

Ex. de fatorial $2^{7}-4$: tabela para $k-p=3$ fatores A, B, C; em seguida substituição pelos $p=4$ fatores restantes

Projeto com fatorial fracionado 2^{k-p} – tabela

- Exemplo para 4 fatores (A, B, C, D), fatorial 2^{4-1}
- Tabela do fatorial 2^3 com fatores A, B, C
- Substituir uma das 4 últimas (no exemplo ABC) por D

Experiment No.	A	B	C	AB	AC	BC	ABC
1	–1	–1	–1	1	1	1	–1
2	1	–1	–1	–1	–1	1	1
3	–1	1	–1	–1	1	–1	1
4	1	1	–1	1	–1	–1	–1
5	–1	–1	1	1	–1	–1	1
6	1	–1	1	–1	1	–1	–1
7	–1	1	1	–1	–1	1	–1
8	1	1	1	1	1	1	1

$\Rightarrow$

Experiment No.	A	B	C	AB	AC	BC	D
1	–1	–1	–1	1	1	1	–1
2	1	–1	–1	–1	–1	1	1
3	–1	1	–1	–1	1	–1	1
4	1	1	–1	1	–1	–1	–1
5	–1	–1	1	1	–1	–1	1
6	1	–1	1	–1	1	–1	–1
7	–1	1	1	–1	–1	1	–1
8	1	1	1	1	1	1	1

- Permitiria analisar os 4 efeitos principais bem como as interações AB, AC e BC

Projeto com fatorial fracionado 2^{k-p} – *confounding*

- Um problema com o projeto fatorial fracionado é que nem todos os efeitos podem ser determinados
- Somente a influência combinada de dois ou mais efeitos é computada
- Este problema é conhecido como *confounding*
- Os efeitos cuja influência não pode ser separada são ditos *confounded*

- Exemplo
  - $q_D = \sum_i y_i x_{Di} = (-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8)/8$
  - $q_{ABC} = \sum_i y_i x_{Ai} x_{Bi} x_{Ci} = (-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8)/8$
  - As expressões para q_D e q_{ABC} são idênticas
  - De fato, o somatório não é nenhum deles, é a soma, seus efeitos são "confundidos"
  - $q_D + q_{ABC} = (-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8)/8$
  - Usamos a notação $D = ABC$
  - Sem fazer o projeto completo não é possível separar seus efeitos
  - Isto não é um problema se é sabido, a priori, que o feito da interação entre A, B e C é pequeno comparado ao de D; nesse caso a expressão indica majoritariamente o de D

Projeto com fatorial fracionado 2^{k-p} – *confounding*

- Não é apenas D e ABC que são confundidos

Experiment No.	*A*	*B*	*C*	*AB*	*AC*	*BC*	*D*
1	–1	–1	–1	1	1	1	–1
2	1	–1	–1	–1	–1	1	1
3	–1	1	–1	–1	1	–1	1
4	1	1	–1	1	–1	–1	–1
5	–1	–1	1	1	–1	–1	1
6	1	–1	1	–1	1	–1	–1
7	–1	1	1	–1	–1	1	–1
8	1	1	1	1	1	1	1

- Exemplo:
  - $q_{BCD} = \sum_i y_i x_{Bi} x_{Ci} x_{Di} = (-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8)/8$
  - Note que é a mesma expressão de q_A
  - Então, A e BCD também são confundidos: $A = BCD$
- Em um projeto 2^{4-1} somente 8 dos $2^4 = 16$ efeitos podem ser computados
- Cada coluna é, na verdade, a soma de 2 efeitos:

$$
\begin{array}{cccc}
A = BCD & B = ACD & C = BCD & AB = CD \\
AC = BD & BC = AD & D = ABC & I = ABCD
\end{array}
$$

- $I = ABCD$ indica que $ABCD$ confunde com a média dos resultados

Projeto com fatorial fracionado 2^{k-p} – *confounding*

- Considere o seguinte projeto 2^{4-1} alternativo

Experiment No.	*A*	*B*	*C*	*D*	*AC*	*BC*	*ABC*
1	–1	–1	–1	1	1	1	–1
2	1	–1	–1	–1	–1	1	1
3	–1	1	–1	–1	1	–1	1
4	1	1	–1	1	–1	–1	–1
5	–1	–1	1	1	–1	–1	1
6	1	–1	1	–1	1	–1	–1
7	–1	1	1	–1	–1	1	–1
8	1	1	1	1	1	1	1

- Quais *confounding* existem neste projeto?

$$I = ABD \qquad A = BD \qquad B = AD \qquad C = ABCD$$
$$D = AB \qquad AC = BCD \qquad BC = ACD \qquad ABC = CD$$

- Este projeto não é tão bom quanto o anterior...
  - No anterior, a média é confundida com a interação de 4ª ordem (ABCD)...
  - ... e cada efeito principal é confundido com interações de 3ª ordem
  - Nesse, a média é confundida com a interação de 3ª ordem...
  - ... e vários dos efeitos principais com interações de 2ª ordem
  - Geralmente, interações de maior ordem são menores que as de menor ordem

Projeto com fatorial fracionado 2^{k-p} – álgebra de *confounding*

- O primeiro projeto 2^{4-1} mostrado é chamado $I = ABCD$ e o segundo $I = ABD$
- Esses são os polinômios geradores dos projetos
- Dado apenas este *confounding* é possível gerar todos os demais por multiplicação de ambos os lados pelos diferentes termos e duas regras simples:
  - A média, I, é tratada como unidade (assim, IA é A)
  - Termos potência de 2 também, podem ser apagados (assim, AB^2C é AC)
- Exemplo no primeiro projeto, $I = ABCD$:
  - multiplicando por A, obtemos $IA = A^2BCD$, ou seja, $A = BCD$ (*confounded*)
  - multiplicando por B, obtemos $IB = AB^2CD$, ou seja, $B = ACD$ (*confounded*)
  - ...
  - multiplicando por BD, obtemos $IBD = AB^2CD^2$, ou seja, $BD = AC$ (*confounded*)
  - e assm por diante

Projeto com fatorial fracionado 2^{k-p} – álgebra de *confounding*

- Projeto 2^{k-p}, 2^p efeitos são *confounded* e o polinômio gerador tem 2^p termos
- Considere o exemplo de projeto 2^{7-4} dado anteriormente
  - As colunas AB, AC, BC, ABC foram substituídas por D, E, F, G respectivamente:

$$AB = D, \quad AC = E, \quad BC = F, \quad ABC = G$$

  - Multiplicando cada equação pelo seu lado esquerdo obtemos

$$I = ABD, \quad I = ACE, \quad I = BCF, \quad I = ABCG$$

  - Ou, de forma equivalente

$$I = ABD = ACE = BCF = ABCG$$

  - O produto de qualquer desses termos também é igual a I
  - O polinômio gerador completo é então

$$I = ABD = ACE = BCF = ABCG = BCDE = ACDF = CDG = ABEF = BEG \\ = AFG = DEF = ADEG = BDFG = CEFG = ABCDEFG$$

  - Outros *confoundings* podem ser obtidos multiplicando-se esta equação por A, B, ...
  - Por exemplo, multiplicando-se o polinômio gerador por A:

$$A = BD = CE = ABCF = BCG = ABCDE = CDF = ACDG = BEF = ABEG = \\ FG = ADEF = DEG = ABDFG = ACEFG = BCDEFG$$

UFV

Projeto com fatorial fracionado 2^{k-p} – exemplo

Exemplo 2^{6-1}

O tempo de CPU gasto por dois aplicativos de formatação de texto, LaTeX e troff, foi medido com arquivos de vários tamanhos e níveis de complexidade. Seis fatores, cada com com dois níveis, foram escolhidos para o estudo.
Os dois primeiros são o aplicativo (*program)* e o tamanho dos arquivos (*bytes*). Os demais são número de equações, de tabelas, *floats*[a] e notas de rodapé.

Foi feito um projeto de experimentos 2^{6-1} com polinômio gerador $I = BCDEF$. Abaixo, os fatores e níveis usados, e os maiores efeitos e interações computados pela tabela de sinais.

Symbol	Factor	Level –1	Level +1
A	Program	LaTeX	troff
B	Bytes	2100	25,000
C	Equations	0	10
D	Floats	0	10
E	Tables	0	10
F	Footnotes	0	10

Symbol	Factor	Effect	Percentage of Variation
B	Bytes	12.0	39.4
A	Program	9.4	24.4
C	Equations	7.5	15.6
AC	Program × Equations	7.2	14.4
E	Tables	3.5	3.4
F	Footnotes	1.6	0.7

Conclusões:

- Mais de 90% da variação é explicada por 3 fatores: aplicativo, tamanho dos arquivos e número de equações, e uma iteração de segunda-ordem
- Os níveis usados para tamanho dos arquivos foi muito diferente, tornando o efeito do tamanho maior que o dos aplicativos sendo comparados
- A interação de “Aplicativo x Tamanho” é baixa. Isto indica que mudar o tamanho do arquivo afeta ambos os aplicativos de forma similar

[a] elementos que não podem ser quebrados em páginas diferentes

Projeto com fatorial fracionado 2^{k-p} – exemplo

Exemplo 2^{6-1} (cont...)

Symbol	Factor	Level –1	Level +1
A	Program	LaTeX	troff
B	Bytes	2100	25,000
C	Equations	0	10
D	Floats	0	10
E	Tables	0	10
F	Footnotes	0	10

Symbol	Factor	Effect	Percentage of Variation
B	Bytes	12.0	39.4
A	Program	9.4	24.4
C	Equations	7.5	15.6
AC	Program × Equations	7.2	14.4
E	Tables	3.5	3.4
F	Footnotes	1.6	0.7

Conclusões (cont...):

- A alta porcentagem de variação explicada por “Aplicativo x Equações” indica que a escolha do aplicativo depende do número de equações no texto. Se considerarmos apenas esses dois fatores, o tempo de CPU relativo gasto nas várias combinações é o mostrado abaixo:

	Number of Equations	
Program	**–1(0)**	**1(10)**
–1(LaTeX)	–9.7	–9.1
1(troff)	–5.3	24.1

  Isso mostra que o aplicativo troff gasta bem mais tempo de CPU se houver equações no texto.

- Se possível, refazer os experimentos com menor variação no tamanho de arquivos para que os aplicativos (em vez do tamanho) influenciem mais. Alternativamente, testar mais níveis de tamanho do arquivo.