Experimento Fatorial Fracionado

Ítalo Cegatta e Francisco Oliveira

2022-07-28

Introdução

Em um delineamento em esquema fatorial tem como objetivo avaliar ou testear os níveis presentes em um experimento com mais de fator. No entanto, com o aumento de níveis e fatores, o experimento pode ficar inviável por conta da quantidade de parcelas a serem instaladas e avaliadas. O delineamento fatorial fracionário é uma alternativa quando se quer reduzir o número de parcelas/ensaios de experimentos com muitos fatores.

Neste documento, abordaremos os principais conceitos por tras do delineamento fatorial fracionado com a resolução de um exemplo prático.

Esquema fatorial

O esquema fatorial consiste na analise de um ou mais fatores de tratamentos e todas as possiveis combinações dos seus níveis. Por exemplo, se temos um fator A com dois níveis e um fator B com 4 níveis, teremos 2 x 4 = 8 combinações a serem testadas, assim as combinações passam a ser os tratamentos.

Considere, por exemplo, um processo de usinagem, onde desaja-se verificar o efeito do avanço, profundidade de corte e geometria da ferramenta na rugorisade superficial. Para realização do exeprimento, foi considerado três níveis para o avanço (A_1,A_2,A_3), 2 níveis para a profundidade do corte (B_1,B_2) e dois níveis para a geometria da ferramenta (C_1, C_2), Carpinet (2009, pag. 108).

Em um exeperimento fatorial completo são realizadas todas as possiveis combinações dos nívies dos fatores. No exemplo, temos 3 x 2 x 2 = 12 combinações de teste, conforme segue na tabela.

Table 1: Combinações de teste de um experimento fatorial 3x2

$A_1B_1C_1$	$A_1B_1C_2$
$A_1B_2C_1$	$A_1B_2C_2$
$A_2B_1C_1$	$A_2B_1C_2$
$A_2B_2C_1$	$A_2B_2C_2$
$A_3B_1C_1$	$A_3B_1C_2$
$A_3B_2C_1$	$A_3B_2C_2$

No geral, quando deseja-se estudar os efeitos dos fatores, planejamentos fatoriais são mais eficientes. Generalizando o exemplo anterior, considere agora que temos um fator A com a níveis, um fator B com b níveis e um fator C com c níveis, teremos abc combinações, sendo assim cada réplica ira conter todas as abc combinações de tratamentos. Em um esquema fatorial os fatores são ditos cruzados, quando são dispostos.

O efeito de um fator é definido como a mudança no valor observado da variável resposta quando o muda-se o nível do fator. Geralmente chamado de efeito principal, refere-se aos fatores primários de interesse no experimento.

Considerado como o esquema fatorial mais simples, envolvendo somente dois fatores A e B, tendo cada um desses dois fatores um número de níveis a ser testado, consideramos a níveis para o fator A e b níveis para o fator B. Assim, termos ab comibinações no experimento. Nesse tipo de experimento, podemos testar:

- a resposta é alterada significativamente quando o nível do fator A é alterado;
- a resposta é alterada significativamente quando o nível do fator B é alterado;
- a alteração na resposta quando da alteração do nível do fator A (ou B) é dependente do nível do fator B (ou A), ou seja, se existe efeito de interação entre os dois fatores.

Assim, considerando um delineamento inteiramente casualizado com n repetições, os dois fatores A com a níveis e B com b, o modelo estatístico é dado por:

$$y_{ikj} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ik} + \epsilon_{ikj}$$

onde:

- μ é a média geral
- α_i é o efeito o i-ésimo nível do fator A, i = 1,. . . ,a
- γ_j é o efeito do j-ésimo nível do fator B, j = 1,. . . ,b
- $(\alpha\gamma)_{ij}$ é o efeito da interação entres os fatores A e B
- ϵ_{ijk} é o erro experimental

Ou ainda em sua forma matricial:

$$\boldsymbol{y} = \boldsymbol{X}_G\mu + \boldsymbol{X}_A\alpha + \boldsymbol{X}_B\gamma + \boldsymbol{X}_{AB} + (\alpha\gamma) + \boldsymbol{\epsilon}$$

Para exemplificar o modelo na forma matricial, considere um experimento fatorial 2x2, feito em um delineamento inteiramente casualizado com 3 repetições para cada tratamento, seja o fator A com e o fator B cada um com 2 níveis, o modelo estatístico na forma matricial será dado por:

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{221} \\ y_{222} \\ y_{223} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} (\alpha\gamma)_{11} \\ (\alpha\gamma)_{12} \\ (\alpha\gamma)_{21} \\ (\alpha\gamma)_{22} \end{bmatrix}$$

Fatorial fracionado

A partir do esquema fatorial, temos o caso particular de experimentos no esquema fatorial fracionado. Este desenho experimental têm grande utilidade nos experimentos que antecedem ao experimento final, como no caso da seleção de fatores e de seus níveis (screening desing) e são exaustivamente utilizados no desenvolvimento e na otimização de processos e produtos.

São casos em que o pesquisador considera como desprezíveis os efeitos das interações de segunda ordem e superiores, ou seja, somente as informações de interesse são sobre os efeitos princiais e interações de primeira ordem, o que podem ser obtidas através de uma fração de um fatorial completo.

Um dos principais uso deste delineamento são em experimentos de triagem, experimentos nos quais muitos fatores são considerados e o objetivo é identifica aqueles fatores (se houver) que tenham grandes efeitos. Os experimentos de triagem geralmente são realizados nos estágios iniciais de um projeto, quando muitos dos fatores inicialmente considerados prováveis têm pouco ou nenhum efeito sobre a resposta. Os fatoes identiicados como importantes são então investigados mais detalhadamente em experimetnos subsequentes.

O uso bem-sucedido de experimentos fatoriais fracionários é baseado em três ideias-chave:

1. **O princípio da esparsidade dos efeitos**: Quando há várias variáveis, o sistema ou processo provavelmente será conduzido principalmente por alguns dos principais efeitos e interações de baixa ordem.

2. **A propriedade de projeção**. Os experimentos fatoriais fracionários podem ser projetados em experimentos mais fortes (maiores) no subconjunto de fatores significativos.

3. **Experimentação sequencial**. É possível combinar as execuções de dois (ou mais) fatoriais fracionários para montar sequencialmente um projeto maior para estimar os efeitos dos fatores e as interações de interesse.

Estrutura do modelo

Segundo Montgomery (2012), o modelo estatístico que será apresentado pelo uso da relação de definição completa, funciona bem em projetos simples, com as frações regulares, no entanto não funciona tão bem em configurações, onde as frações sejam não regulares. E ainda, existem alguns fatroias fracionários que não possuem relação definidora.

Como os experimentos fatoriais fracionados podem ter a ocorrencia de confudimento por não trabalhar com todas as possíveis combinações, sendo assim, para apresentar o modelo estatístico será utilizado a estrutura de aliases

Assim, o método usa a representação do modelo polinomial ou de regressão do modelo, ou seja,

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$$

onde $\boldsymbol{y}$ é um vetor das respostas, $\boldsymbol{X}_1$ é uma matrix nxp_1 contendo a matriz de planejamento que pode ser expandida para a forma do modelo que o pesquisador deseja ajustar, $\boldsymbol{\beta}_1$ é um vetor p_1x1 de parâmetros e $\boldsymbol{\epsilon}$ é um vetor $nx1$ de erros. A estimativa de mínimos quadrados de $\boldsymbol{\beta}_1$ é

$$\hat{\boldsymbol{\beta}}_1 = (\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{y}$$

Agora suponha que o modelo verdadeiro seja,

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$$

onde $\boldsymbol{X}_2$ é uma matriz nxp_2 contendo variáveis adicionais que não estão presentes no modelo ajustado e $\boldsymbol{\beta}_2$ é um vetor p_2x1 dos parâmetros associados a essas variáveis. Assim, podemos mostrar que

$$E(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + (\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 = \boldsymbol{\beta}_1 + \boldsymbol{A}\boldsymbol{\beta}_2$$

$\boldsymbol{A}$ é chamada de matriz de aliase. Os elemtnos desta matriz operão em $\boldsymbol{\beta}_2$ indentificando as relações de aliase para os parâmetos no vetor $\boldsymbol{\beta}_1$.

Para exepmlificar, suponha que em um experimento fatorial fracionado 2^{3-1} com a realação definidora $I = ABC$ ou $I = x_1x_2x_3$. O modelo que será ajustado terá somente a presença dos efeitos principais, ou seja

$$y = \mu + \alpha x_1 + \gamma x_2 + \tau x_3 + \epsilon$$

Na notação anterior

$$\boldsymbol{\beta}_1 = \begin{bmatrix} \mu \\ \alpha \\ \gamma \\ \tau \end{bmatrix} \; e \;\; \boldsymbol{X}_1 = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Agora vamos supor que o verdadeiro modelo contenha todas as interaçoes de seunga ordem dos fatores, desse modo,

$$y = \mu + \alpha x_1 + \gamma x_2 + \tau x_3 + (\alpha\gamma)x_1x_2 + (\alpha\tau)x_1x_3 + (\gamma\tau)x_2x_3 + \epsilon$$

Assim, nosso $\boldsymbol{\beta}_2$ e nossa $\boldsymbol{X}_2$ serão,

$$\boldsymbol{\beta}_2 = \begin{bmatrix} (\alpha\gamma) \\ (\alpha\tau) \\ (\gamma\tau) \end{bmatrix} \; e \;\; \boldsymbol{X_2} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

Logo,

$$(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1} = (1/4)\boldsymbol{I}_4 \;\; e \;\; \boldsymbol{X'}_1\boldsymbol{X}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{bmatrix}$$

Portanto,

$$E(\hat{\boldsymbol{\beta}}_{\boldsymbol{1}}) = \boldsymbol{\beta}_1 + \boldsymbol{A}\boldsymbol{\beta}_2 E \begin{bmatrix} \hat{\mu} \\ \hat{\alpha} \\ \hat{\gamma} \\ \hat{\tau} \end{bmatrix} = \begin{bmatrix} \mu \\ \alpha \\ \gamma \\ \tau \end{bmatrix} + (1/4)\boldsymbol{I}_4 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{bmatrix} \begin{bmatrix} (\alpha\gamma) \\ (\alpha\tau) \\ (\gamma\tau) \end{bmatrix} = \begin{bmatrix} \mu \\ \alpha \\ \gamma \\ \tau \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} (\alpha\gamma) \\ (\alpha\tau) \\ (\gamma\tau) \end{bmatrix} = \begin{bmatrix} \mu \\ \alpha \\ \gamma \\ \tau \end{bmatrix} + \begin{bmatrix} 0 \\ (\alpha\gamma) \\ (\alpha\tau) \\ (\gamma\tau) \end{bmatrix} = \begin{bmatrix} \mu \\ \alpha + (\gamma\tau) \\ \gamma + (\alpha\tau) \\ \tau + (\alpha\gamma) \end{bmatrix}$$

Dessa formar, podemos perceber que cada um dos fatores principais está associado a uma das interaçoes de segunda ordem. Perceba que cada uma das linhas da matriz de aliase representa um dos fatores de $\boldsymbol{\beta}_1$ e cada uma das colunas respresenta um dos fatores de $\boldsymbol{\beta}_2$. Mesmo sendo um exemplo simples, esse método pode ser aplicado a projetos mais complexos.

Exemplo 1: Fatorial 2^{4-1}

Considere um experimento fatorial 2^4. Nesse caso, o experimento completo envolve 16 combinações. Suponha que queiramos reduzir o número de combinações de teste para 8. Então, queremos realizar metade de um experimento 2^4, ou seja, $\frac{1}{2}2^4 = 2^{-1}2^4 = 2^{4-1} = 8$ combinações.

Portanto, o experimento básico é um experimento 2^3, com uma matriz de planejamento ilustrada abaixo.

Table 2: Matriz de planejamento para experimento 2^3

Teste	X_1	X_2	X_3
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

Entretanto, uma quarta variável, x_4, precisa ser introduzida neste planejamento. Suponha que x_4 seja introduzida no experimento de forma que a coluna de sinais correspondente ao efeito 123 seja utilizada para definir os níveis de x_4 para as 8 combinações de teste, conforme ilustrado na matriz de planejamento abaixo.

Table 3: Matriz de planejamento para experimento fracionário

Teste	1	2	3	$4 = 123$
1	-	-	-	-
2	+	-	-	+
3	-	+	-	+
4	+	+	-	-
5	-	-	+	+
6	+	-	+	-
7	-	+	+	-
8	+	+	+	+

E a matriz de cáculo dos efeitos para um planejamento como este é dado por:

Table 4: Matriz de cálculo, experimento fracionário 2^{4-1}.

Teste	I	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234
1	+	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+
2	+	+	-	-	+	-	-	+	+	-	-	+	-	-	+	+
3	+	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
4	+	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
5	+	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+
6	+	+	-	+	-	-	+	-	-	+	-	-	+	-	+	+
7	+	-	+	+	-	-	-	+	+	-	-	-	+	+	-	+
8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Examinando a matriz de cálculo, podemos perceber que muitas das colunas são idênticas. Das 16 colunas, apenas 8 são únicas, ou seja, a combinação linear para o cálculo do efeito da variável 1 é o mesmo que para o cálculo do efeito da interação entre 234. Da mesma forma temos os seguintes confundimentos:

1 e 234	12 e 34
2 e 134	13 e 24
3 e 124	23 e 14
4 e 123	Média I e 1234

Assim, as 8 colunas diferentes entre si, quando multiplicada pela coluna de dados e dividindo por 4 (N/2), são usadas para se obter as combinações lineares l_0, l_1, l_2, ..., l_{123} de efeitos confundidos. Assim:

l_0	estima	Média + (½) 1234
l_1	estima	1 + 234
l_2	estima	2 + 134
l_3	estima	3 + 124
l_{12}	estima	12 + 34
l_{13}	estima	13 + 24
l_{23}	estima	23 + 14
l_{123}	estima	4 / 123

Se assumimos que as interações de 3 fatores e de 4 fatores podem ser desprezadas, o experimento produz as seguintes combinações lineares.

l_0	estima	Média
l_1	estima	1
l_2	estima	2
l_3	estima	3
l_{12}	estima	12 + 34
l_{13}	estima	13 + 24
l_{23}	estima	23 + 14
l_{123}	estima	4

Associações, como as definidas anteriormente, definem a resolução de um fatorial fracionado. De modo geral, a resolução de um fatorial fracionado é igual ao menor número de letras de sua relação definidora. É importante lembrar que quanto maior for a resolução, maior será a utilidade prática do fatorial fracionado. A relação entre as possíveis resoluções e o número de fatores pode ser vista na tabela a seguir:

Runs	Factors								
	2	3	4	5	6	7	8	9	10
4	Full	III							
8		Full	IV	III	III	III			
16			Full	V	IV	IV	IV	III	III
32				Full	VI	IV	IV	IV	IV
64					Full	VII	V	IV	IV
128						Full	VIII	VI	V

Resolução III Não há associação entre efeitos principais . Os efeitos principais estão associados as interações de primeira ordem (interações de dois fatores) . As interações de primeira ordem estão associadas entre si. Pouco útil na prática.

Resolução IV Não há associações entre os efeitos principais e interações de primeira orden. As interações de primeira ordem estão associadas entre si.

Resolucção V Não há associação entre efeitos principais. Não há associações entre efeitos principais e interações de preimeira ordem.

Exemplo 2: Fatorial 2^{6-2}

Exemplo 14-9, Montgomery (EAPE): encolhimento excessivo em peças de moldagem por injeção. Analisando 6 fatores em um experimento fatorial fracionado 2^{6-2}, foram realizadas 16 corridas para estes seis fatores. O planejamento foi construído escrevendo um planejamento básico 2^4 nos fatores A, B, C, e D, e então estabelecendo $E = ABC$, e $F = BCD$, pois as relações de definição escolhidas foram $I = ABCE$, e $I = BCDF$. O objetivo do planejamento era identificar os fatores que causavam o encolhimento excessivo, portanto as combinações de fatores que gerem **menores** valores da variável resposta (encolhimento) deverão ser utilizadas no processo de fabricação.

```
library(tidyverse)
library(ggrepel)

tab_dados <- read_csv2("dados/exemplo_fr6-2.csv")

tab_dados
```

```
## # A tibble: 16 x 7
##        A     B     C     D     E     F     y
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1    -1    -1    -1    -1    -1    -1     6
##  2     1    -1    -1    -1     1    -1    10
##  3    -1     1    -1    -1     1     1    32
##  4     1     1    -1    -1    -1     1    60
##  5    -1    -1     1    -1     1     1     4
##  6     1    -1     1    -1    -1     1    15
##  7    -1     1     1    -1    -1    -1    26
##  8     1     1     1    -1     1    -1    60
##  9    -1    -1    -1     1    -1     1     8
## 10     1    -1    -1     1     1     1    12
## 11    -1     1    -1     1     1    -1    34
## 12     1     1    -1     1    -1    -1    60
## 13    -1    -1     1     1     1    -1    16
## 14     1    -1     1     1    -1    -1     5
## 15    -1     1     1     1    -1     1    37
## 16     1     1     1     1     1     1    52
```

Para montarmos a matriz efeitos codificada, podemos utilizar a função `moden.matrix` indicando a formula de interação completa entre todos os fatores.

```
tab_matriz_completa <- model.matrix(~A * B * C * D * E * F, data = tab_dados)

head(tab_matriz_completa)
```

```
##   (Intercept)  A  B  C  D  E  F A:B A:C B:C A:D B:D C:D A:E B:E C:E D:E A:F B:F
## 1           1 -1 -1 -1 -1 -1 -1   1   1   1   1   1   1   1   1   1   1   1   1
```

```
## 2           1  1 -1 -1 -1  1 -1  -1  -1   1  -1   1   1   1  -1  -1  -1  -1   1
## 3           1 -1  1 -1 -1  1  1  -1   1  -1   1  -1   1  -1   1  -1  -1  -1   1
## 4           1  1  1 -1 -1 -1  1   1  -1  -1  -1  -1   1  -1  -1   1   1   1   1
## 5           1 -1 -1  1 -1  1  1   1  -1  -1   1   1  -1  -1  -1   1  -1  -1  -1
## 6           1  1 -1  1 -1 -1  1  -1   1  -1  -1   1  -1  -1   1  -1   1   1  -1
##   C:F D:F E:F A:B:C A:B:D A:C:D B:C:D A:B:E A:C:E B:C:E A:D:E B:D:E C:D:E A:B:F
## 1   1   1   1    -1    -1    -1    -1    -1    -1    -1    -1    -1    -1    -1
## 2   1   1  -1     1     1     1    -1    -1    -1     1    -1     1     1     1
## 3  -1  -1   1     1     1    -1     1    -1     1    -1     1    -1     1    -1
## 4  -1  -1  -1    -1    -1     1     1    -1     1     1     1     1    -1     1
## 5   1  -1   1     1    -1     1     1     1    -1    -1     1     1    -1     1
## 6   1  -1  -1    -1     1    -1     1     1    -1     1     1    -1     1    -1
##   A:C:F B:C:F A:D:F B:D:F C:D:F A:E:F B:E:F C:E:F D:E:F A:B:C:D A:B:C:E A:B:D:E
## 1    -1    -1    -1    -1    -1    -1    -1    -1    -1       1       1       1
## 2     1    -1     1    -1    -1    -1     1     1     1      -1       1       1
## 3     1    -1     1    -1     1    -1     1    -1    -1      -1       1       1
## 4    -1    -1    -1    -1     1    -1    -1     1     1       1       1       1
## 5    -1    -1     1     1    -1    -1    -1     1    -1      -1       1      -1
## 6     1    -1    -1     1    -1    -1     1    -1     1       1       1      -1
##   A:C:D:E B:C:D:E A:B:C:F A:B:D:F A:C:D:F B:C:D:F A:B:E:F A:C:E:F B:C:E:F
## 1       1       1       1       1       1       1       1       1       1
## 2       1      -1      -1      -1      -1       1       1       1      -1
## 3      -1       1       1       1      -1       1      -1       1      -1
## 4      -1      -1      -1      -1       1       1      -1       1       1
## 5       1       1       1      -1       1       1       1      -1      -1
## 6       1      -1      -1       1      -1       1       1      -1       1
##   A:D:E:F B:D:E:F C:D:E:F A:B:C:D:E A:B:C:D:F A:B:C:E:F A:B:D:E:F A:C:D:E:F
## 1       1       1       1        -1        -1        -1        -1        -1
## 2       1      -1      -1        -1         1        -1        -1        -1
## 3       1      -1       1        -1        -1         1         1        -1
## 4       1       1      -1        -1         1         1         1        -1
## 5       1       1      -1        -1        -1         1        -1         1
## 6       1      -1       1        -1         1         1        -1         1
##   B:C:D:E:F A:B:C:D:E:F
## 1        -1           1
## 2         1           1
## 3         1          -1
## 4        -1          -1
## 5         1          -1
## 6        -1          -1
```

Como foi demonstrado a cima, a matriz de efeitos parametrizada com a geradora $E = ABC$, e $F = BCD$ contém diversos fatores que estão confundidos, por isso devemos selecionar dessa tabela apenas os efeitos que sejam únicos. O bloco de código abaixo mostra os efeitos que se repetem em toda a matriz. Como exemplo, mostramos apenas os 3 primeiros resultados.

```r
vec_efeitos <- attr(tab_matriz_completa, "dimnames")[[2]]

lst_out <- vector("list", length = length(vec_efeitos))

for (i in seq_along(vec_efeitos)) {

  names(lst_out)[i] <- vec_efeitos[i]
```

```
  i_efeito <- tab_matriz_completa %>%
    as_tibble() %>%
    pull(!!i)

  lst_out[i][[1]] <- tab_matriz_completa %>%
    as_tibble() %>%
    select_if(~ all(. == i_efeito))
}

head(lst_out, n = 3)
```

```
## $`(Intercept)`
## # A tibble: 16 x 4
##    `(Intercept)` `A:B:C:E` `B:C:D:F` `A:D:E:F`
##            <dbl>     <dbl>     <dbl>     <dbl>
##  1             1         1         1         1
##  2             1         1         1         1
##  3             1         1         1         1
##  4             1         1         1         1
##  5             1         1         1         1
##  6             1         1         1         1
##  7             1         1         1         1
##  8             1         1         1         1
##  9             1         1         1         1
## 10             1         1         1         1
## 11             1         1         1         1
## 12             1         1         1         1
## 13             1         1         1         1
## 14             1         1         1         1
## 15             1         1         1         1
## 16             1         1         1         1
##
## $A
## # A tibble: 16 x 4
##        A `B:C:E` `D:E:F` `A:B:C:D:F`
##    <dbl>   <dbl>   <dbl>       <dbl>
##  1    -1      -1      -1          -1
##  2     1       1       1           1
##  3    -1      -1      -1          -1
##  4     1       1       1           1
##  5    -1      -1      -1          -1
##  6     1       1       1           1
##  7    -1      -1      -1          -1
##  8     1       1       1           1
##  9    -1      -1      -1          -1
## 10     1       1       1           1
## 11    -1      -1      -1          -1
## 12     1       1       1           1
## 13    -1      -1      -1          -1
## 14     1       1       1           1
## 15    -1      -1      -1          -1
## 16     1       1       1           1
##
```

```
## $B
## # A tibble: 16 x 4
##        B `A:C:E` `C:D:F` `A:B:D:E:F`
##    <dbl>   <dbl>   <dbl>       <dbl>
##  1    -1      -1      -1          -1
##  2    -1      -1      -1          -1
##  3     1       1       1           1
##  4     1       1       1           1
##  5    -1      -1      -1          -1
##  6    -1      -1      -1          -1
##  7     1       1       1           1
##  8     1       1       1           1
##  9    -1      -1      -1          -1
## 10    -1      -1      -1          -1
## 11     1       1       1           1
## 12     1       1       1           1
## 13    -1      -1      -1          -1
## 14    -1      -1      -1          -1
## 15     1       1       1           1
## 16     1       1       1           1
```

Considerando apenas os efeitos unicos, temos a matriz a seguir:

```
tab_matriz_frac <- tab_matriz_completa %>%
  unique(MARGIN = 2) %>%
  as_tibble()

tab_matriz_frac
```

```
## # A tibble: 16 x 16
##    `(Intercept)`     A     B     C     D     E     F `A:B` `A:C` `B:C` `A:D`
##            <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1             1    -1    -1    -1    -1    -1    -1     1     1     1     1
##  2             1     1    -1    -1    -1     1    -1    -1    -1     1    -1
##  3             1    -1     1    -1    -1     1     1    -1     1    -1     1
##  4             1     1     1    -1    -1    -1     1     1    -1    -1    -1
##  5             1    -1    -1     1    -1     1     1     1    -1    -1     1
##  6             1     1    -1     1    -1    -1     1    -1     1    -1    -1
##  7             1    -1     1     1    -1    -1    -1    -1    -1     1     1
##  8             1     1     1     1    -1     1    -1     1     1     1    -1
##  9             1    -1    -1    -1     1    -1     1     1     1     1    -1
## 10             1     1    -1    -1     1     1     1    -1    -1     1     1
## 11             1    -1     1    -1     1     1    -1    -1     1    -1    -1
## 12             1     1     1    -1     1    -1    -1     1    -1    -1     1
## 13             1    -1    -1     1     1     1    -1     1    -1    -1    -1
## 14             1     1    -1     1     1    -1    -1    -1     1    -1     1
## 15             1    -1     1     1     1    -1     1    -1    -1     1    -1
## 16             1     1     1     1     1     1     1     1     1     1     1
## # ... with 5 more variables: `B:D` <dbl>, `C:D` <dbl>, `D:E` <dbl>,
## #   `A:B:D` <dbl>, `A:C:D` <dbl>
```

A estimativa dos efeitos é feita pela multiplicação da transpoda da matriz de sinais pelo vetor de resposta seguido da divisão pelo numero de valores que compôem a o contraste.

```r
mat_contrastes <- t(tab_matriz_frac[, -1]) %*% tab_dados$y

mat_contrastes
```

```
##       [,1]
## A      111
## B      285
## C       -7
## D       11
## E        3
## F        3
## A:B     95
## A:C    -13
## B:C    -15
## A:D    -43
## B:D     -1
## C:D     -1
## D:E      5
## A:B:D    1
## A:C:D  -39
```

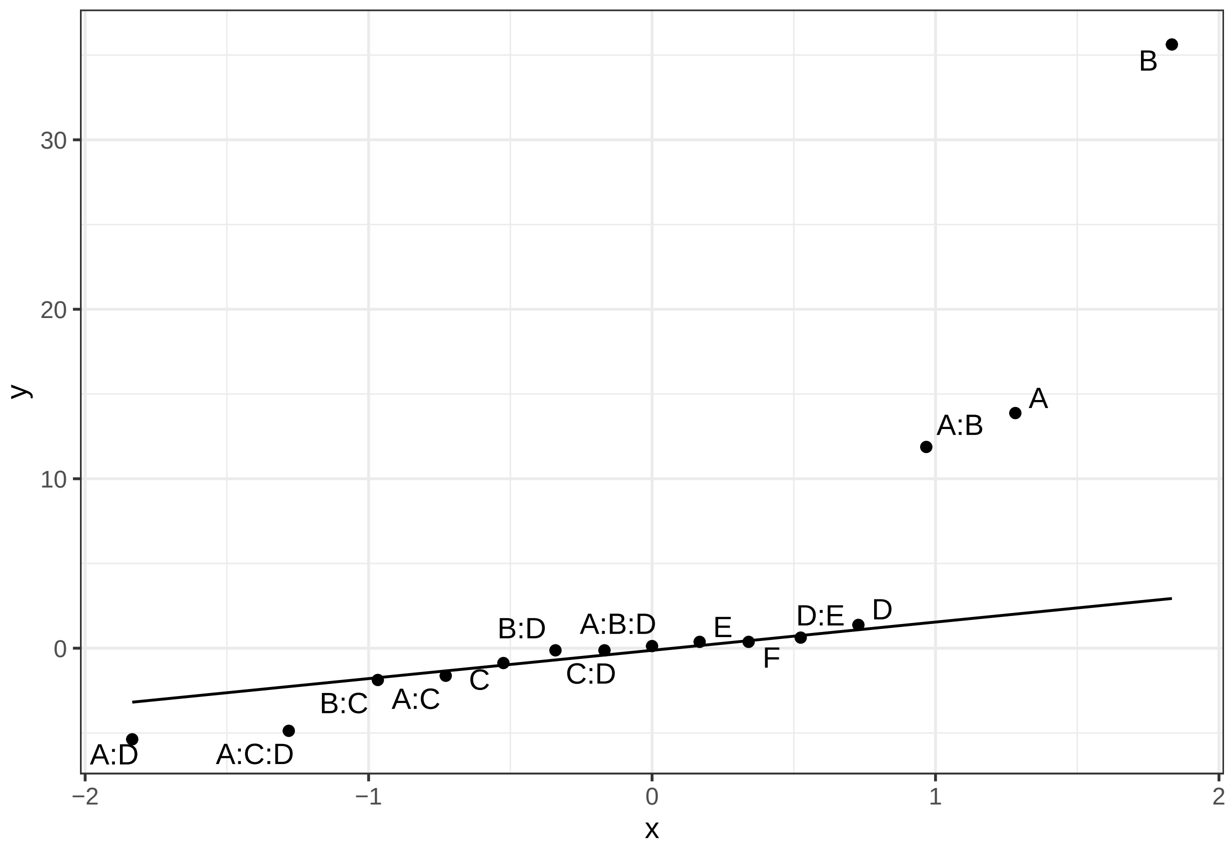
```r
k <- 4
r <- 1
mat_efeitos <- {mat_contrastes/(r * 2^(k - 1))} %>%
  as_tibble(rownames = "efeito")

mat_efeitos
```

```
## # A tibble: 15 x 2
##    efeito     V1
##    <chr>   <dbl>
##  1 A      13.9
##  2 B      35.6
##  3 C      -0.875
##  4 D       1.38
##  5 E       0.375
##  6 F       0.375
##  7 A:B    11.9
##  8 A:C    -1.62
##  9 B:C    -1.88
## 10 A:D    -5.38
## 11 B:D    -0.125
## 12 C:D    -0.125
## 13 D:E     0.625
## 14 A:B:D   0.125
## 15 A:C:D  -4.88
```

Pelo gráfico QQPlot podemos ver quais efeitos podem ser significativos.

```r
mat_efeitos %>%
  ggplot(aes(sample = V1)) +
  geom_qq() +
```

```r
  geom_text_repel(label = arrange(mat_efeitos, V1)$efeito, stat = "qq") +
  stat_qq_line() +
  theme_bw()
```



Ajustando um modelo com apenas os efeitos

```r
lm(y ~ .^2, data = tab_dados) %>%
  summary()
```

```
##
## Call:
## lm(formula = y ~ .^2, data = tab_dados)
##
## Residuals:
##      1      2      3      4      5      6      7      8      9     10     11
##  2.375 -2.375  2.500 -2.500 -2.500  2.500 -2.375  2.375 -2.375  2.375 -2.500
##     12     13     14     15     16
##  2.500  2.500 -2.500  2.375 -2.375
##
## Coefficients: (8 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.3125     1.7241  15.841  0.00396 **
## A             6.9375     1.7241   4.024  0.05657 .
## B            17.8125     1.7241  10.331  0.00924 **
## C            -0.4375     1.7241  -0.254  0.82339
## D             0.6875     1.7241   0.399  0.72862
```

```
## E             0.1875     1.7241   0.109  0.92333
## F             0.1875     1.7241   0.109  0.92333
## A:B           5.9375     1.7241   3.444  0.07496 .
## A:C          -0.8125     1.7241  -0.471  0.68387
## A:D          -2.6875     1.7241  -1.559  0.25939
## A:E          -0.9375     1.7241  -0.544  0.64112
## A:F           0.3125     1.7241   0.181  0.87288
## B:C               NA         NA      NA       NA
## B:D          -0.0625     1.7241  -0.036  0.97438
## B:E               NA         NA      NA       NA
## B:F          -0.0625     1.7241  -0.036  0.97438
## C:D               NA         NA      NA       NA
## C:E               NA         NA      NA       NA
## C:F               NA         NA      NA       NA
## D:E               NA         NA      NA       NA
## D:F               NA         NA      NA       NA
## E:F               NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.897 on 2 degrees of freedom
## Multiple R-squared:  0.9857, Adjusted R-squared:  0.8929
## F-statistic: 10.62 on 13 and 2 DF,  p-value: 0.08928
```

Como vimos no gráfico anterior, os efeitos A, B, e as interações AB, AD, e ACD parecem importantes, então podemos ajustar o modelo

```r
m0 <- lm(y ~ A + B + C + D + A:B + A:D + A:C:D, data = tab_dados)

anova(m0)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq   F value    Pr(>F)
## A          1  770.1   770.1  224.0182 3.919e-07 ***
## B          1 5076.6  5076.6 1476.8182 2.309e-10 ***
## C          1    3.1     3.1    0.8909 0.3728597
## D          1    7.6     7.6    2.2000 0.1762999
## A:B        1  564.1   564.1  164.0909 1.301e-06 ***
## A:D        1  115.6   115.6   33.6182 0.0004060 ***
## A:C:D      1   95.1    95.1   27.6545 0.0007657 ***
## Residuals  8   27.5     3.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como as interações AD e ACD requerem a permanencia dos efeitos principais (não significativos) C e D, podemos ajustar um modelo mais simples, sem estes termos.

```r
m1 <- lm(y ~ A + B + A:B, data = tab_dados)

anova(m1)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## A          1  770.1   770.1  37.149 5.377e-05 ***
## B          1 5076.6  5076.6 244.899 2.392e-09 ***
## A:B        1  564.1   564.1  27.211  0.000216 ***
## Residuals 12  248.8    20.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

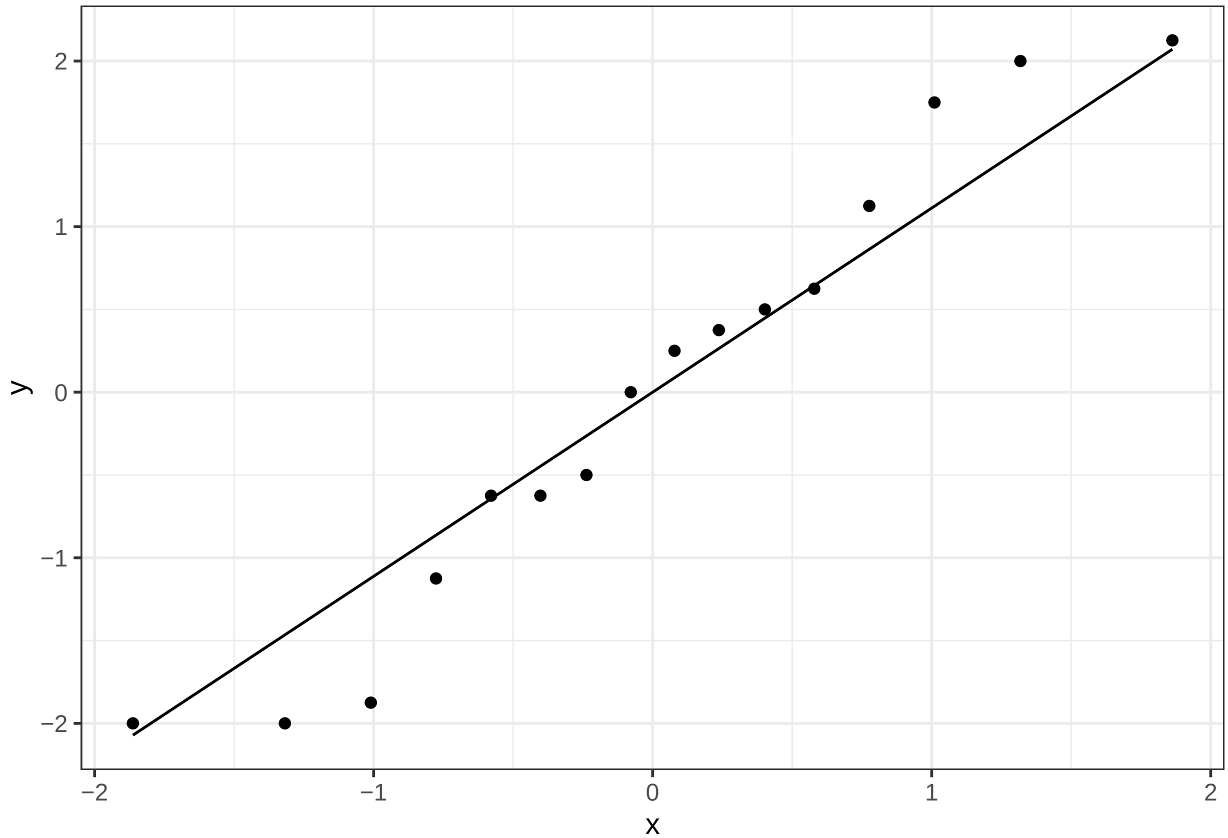
E verificar se existe diferença com o nosso modelo inicial através de um TRV. Se a diferença for significativa, significa que não podemos ignorar os efeitos de AD e ACD (lembre que C e D precisam ser mantidos, mesmo que não significativos, pelo princípio da marginalidade)

```
anova(m1, m0)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ A + B + A:B
## Model 2: y ~ A + B + C + D + A:B + A:D + A:C:D
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     12 248.75
## 2      8  27.50  4    221.25 16.091 0.0006808 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como o resultado é que existe diferença, devemos ficar com o modelo com mais parâmetros (m0)

```
residuals(m0) %>%
  as_tibble() %>%
  ggplot(aes(sample = value)) +
  geom_qq() +
  stat_qq_line() +
  theme_bw()
```


y
x
2
1
0
−1
−2
−2
−1
0
1
2