Planejamento para o ajuste de curvas flexíveis

Luzia Trinca (luzia.trinca@unesp.br)
Unesp, Botucatu, Brazil

VIII Encontro dos Alunos
PG em Estatística e Experimentação Agronômica/ESALQ
Novembro 2018

Introdução

Dados experimentais com fatores contínuos usualmente apresentam padrões de curvaturas que nem sempre são captadas por polinômios 1^a ou 2^a ordem.

Introdução

Dados experimentais com fatores contínuos usualmente apresentam padrões de curvaturas que nem sempre são captadas por polinômios 1^a ou 2^a ordem.

Aumento do grau do polinômio pode levar a curvas ou superfícies de difícil interpretação.

Introdução

Dados experimentais com fatores contínuos usualmente apresentam padrões de curvaturas que nem sempre são captadas por polinômios 1^a ou 2^a ordem.

Aumento do grau do polinômio pode levar a curvas ou superfícies de difícil interpretação.

Inspirados nas transformações Box-Tidwell, Royston e Altman (1994) propuseram os polinômios fracionários (FP) para modelar relações entre variáveis resposta e regressoras contínuas.

Definição

Para uma única regressora $x > 0$, o FP de grau m é escrito como:

$$\eta(x, \boldsymbol{\theta}) = \beta_0 + \varphi(x, \boldsymbol{\alpha}) = \beta_r \sum_{r=0}^{m} H_r(x),$$

$\eta(x, \boldsymbol{\theta})$: função preditora, $\boldsymbol{\theta}$: vetor de todos os parâmetros e

$$H_r(x) = \begin{cases} 1 & r = 0 \\ x^{(\alpha_r)} & \alpha_r \neq \alpha_{r-1}; r = 1, \ldots, m \\ H_{r-1} log(x) & \alpha_r = \alpha_{r-1}; r = 2, \ldots, m \end{cases}$$

com

$$x^{(\alpha_r)} = \begin{cases} x^{\alpha_r} & \alpha_r \neq 0 \\ log(x) & \alpha_r = 0, \end{cases}$$

para $\alpha_1 < \alpha_2 < \ldots < \alpha_m$.

FP's de baixa ordem

- $m = 1 \rightarrow$ FP1:

$$\eta(x, \boldsymbol{\theta}) = \beta_0 + \beta_1 x^{(\alpha)}$$

FP's de baixa ordem

- $m = 1 \rightarrow$ FP1:

$$\eta(x, \boldsymbol{\theta}) = \beta_0 + \beta_1 x^{(\alpha)}$$

- $m = 2 \rightarrow$ FP2:

$$\eta(x, \boldsymbol{\theta}) = \begin{cases} \beta_0 + \beta_1 x^{(\alpha_1)} + \beta_{11} x^{(\alpha_2)} & \alpha_1 \neq \alpha_2 \\ \beta_0 + \beta_1 x^{(\alpha)} + \beta_{11} x^{(\alpha)} \log(x) & \alpha_1 = \alpha_2. \end{cases}$$

FP's de baixa ordem

- $m = 1 \rightarrow$ FP1:

$$\eta(x, \boldsymbol{\theta}) = \beta_0 + \beta_1 x^{(\alpha)}$$

- $m = 2 \rightarrow$ FP2:

$$\eta(x, \boldsymbol{\theta}) = \begin{cases} \beta_0 + \beta_1 x^{(\alpha_1)} + \beta_{11} x^{(\alpha_2)} & \alpha_1 \neq \alpha_2 \\ \beta_0 + \beta_1 x^{(\alpha)} + \beta_{11} x^{(\alpha)} \log(x) & \alpha_1 = \alpha_2. \end{cases}$$

$\Rightarrow$ requer estimação de um parâmetro a mais, para cada grau associado a cada regressora.

Royston e co-autores sugerem

$$\alpha \in S = \{-3, -2, -0.5, 0, 0.5, 1, 2, 3\}$$

engloba valores de α_r que geram modelos bastante flexíveis para cobrir diversos tipos de relações que aparecem nas aplicações.

Royston e co-autores sugerem

$$\alpha \in S = \{-3, -2, -0.5, 0, 0.5, 1, 2, 3\}$$

engloba valores de α_r que geram modelos bastante flexíveis para cobrir diversos tipos de relações que aparecem nas aplicações.

Eles apresentam muitas aplicações em GLM's e modelo de Cox e defendem vantagens a outras alternativas como categorização de variáveis quantitativas.

Royston e co-autores sugerem

$$\alpha \in S = \{-3, -2, -0.5, 0, 0.5, 1, 2, 3\}$$

engloba valores de α_r que geram modelos bastante flexíveis para cobrir diversos tipos de relações que aparecem nas aplicações.

Eles apresentam muitas aplicações em GLM's e modelo de Cox e defendem vantagens a outras alternativas como categorização de variáveis quantitativas.

Estimação por MV para $\alpha \in S$.

Royston e colaboradores

Aplicações para dados de estudos observacionais: n razoavelmente grande.

Royston e colaboradores

Aplicações para dados de estudos observacionais: n razoavelmente grande.

Métodos para seleção de variáveis e inclusão de termos de interação entre regressoras.

Royston e colaboradores

Aplicações para dados de estudos observacionais: n razoavelmente grande.

Métodos para seleção de variáveis e inclusão de termos de interação entre regressoras.

Implementações computacionais: Stata, R (pacote `mfp`) e SAS (macros).

Os FP's para dados experimentais

Potencial para estudos de superfície de resposta.

Como, em geral, n não é muito grande, FP1 e FP2, com FP2 re-definido por

$$\eta(x, \boldsymbol{\theta}) = \beta_0 + \beta_1 x^{(\alpha)} + \beta_{11} \left\{ x^{(\alpha)} \right\}^2,$$

Os FP's para dados experimentais

Potencial para estudos de superfície de resposta.

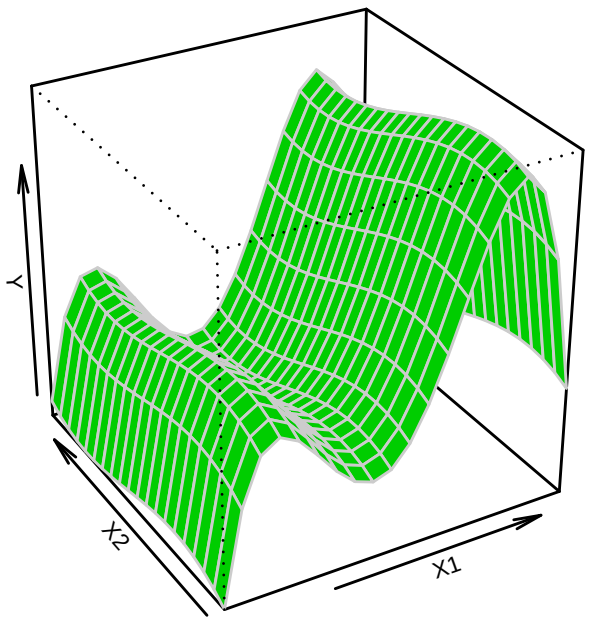
Como, em geral, n não é muito grande, FP1 e FP2, com FP2 re-definido por

$$\eta(x, \boldsymbol{\theta}) = \beta_0 + \beta_1 x^{(\alpha)} + \beta_{11} \left\{ x^{(\alpha)} \right\}^2,$$

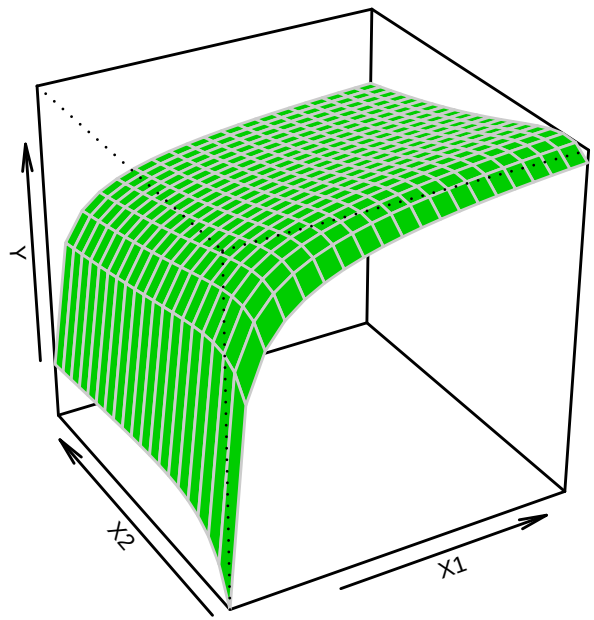
tem a flexibilidade de englobar curvas com assíntotas e curvas com ponto de ótimo, simétricas e assimétricas.

Os FP's para dados experimentais

Polinômio usual (4^a ordem)

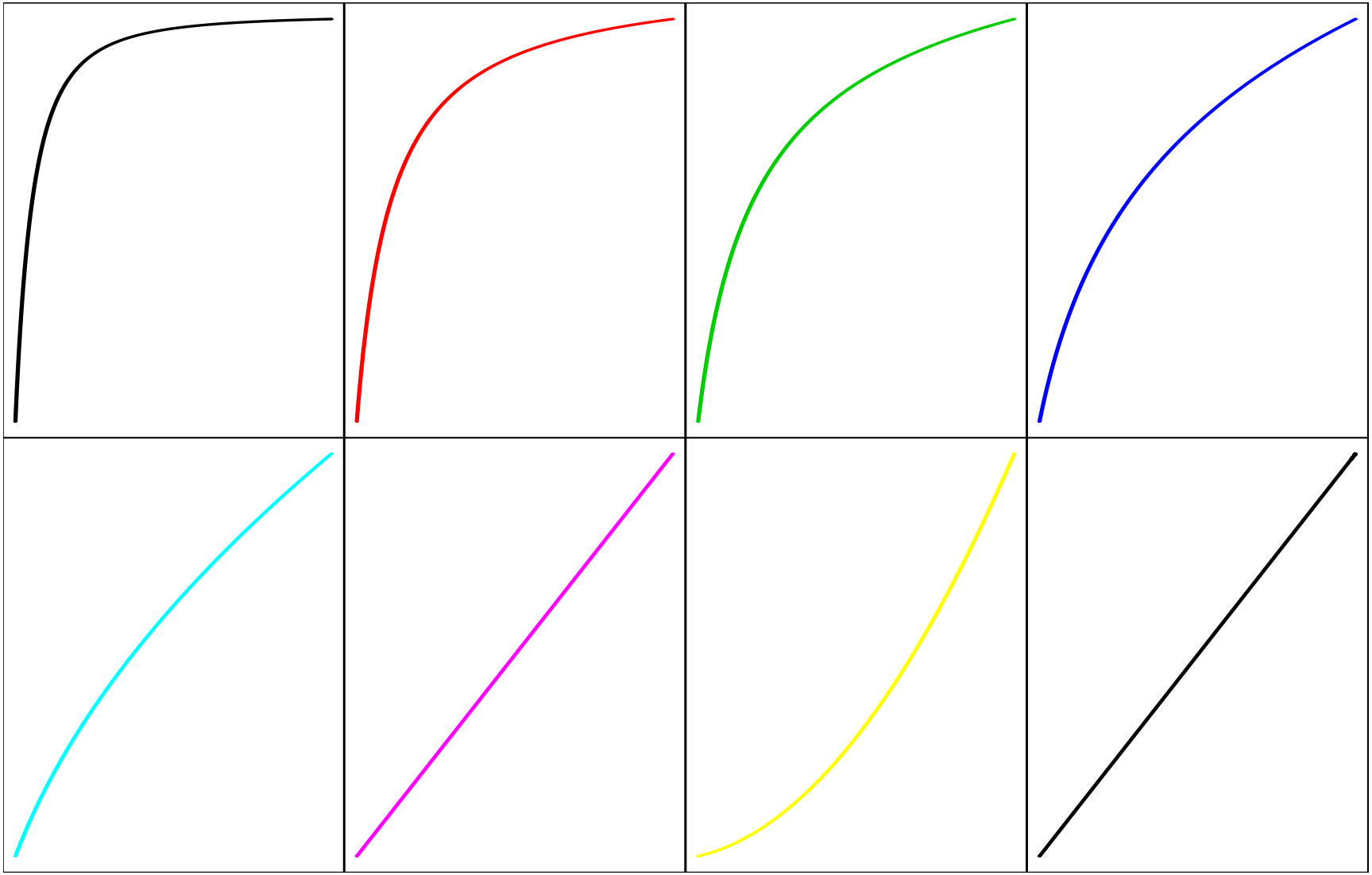


FP (2^a ordem)



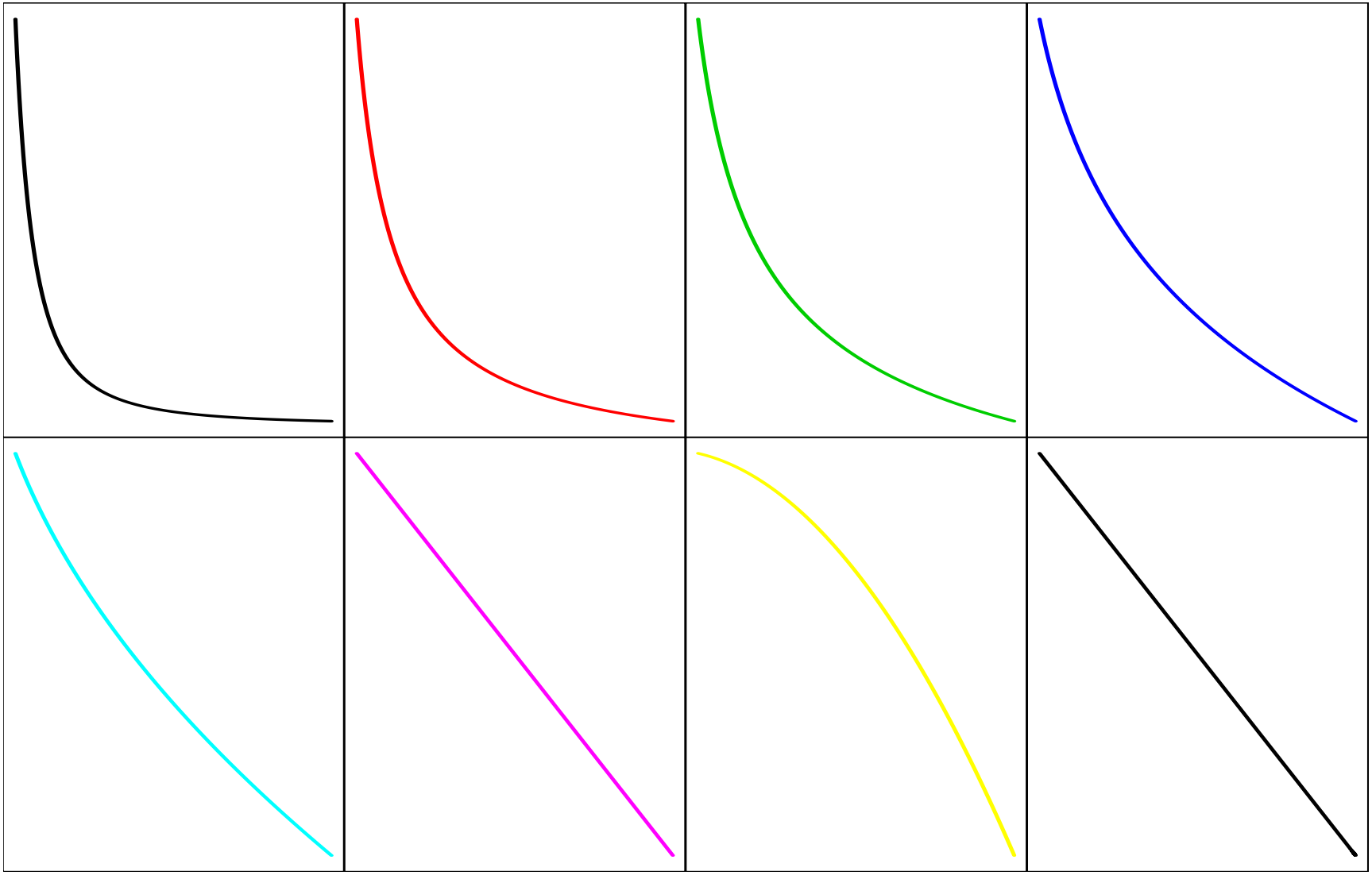
Produção de uma hortaliça em função da densidade de sementes e espaçamento no plantio.

Algumas curvas possíveis FP1



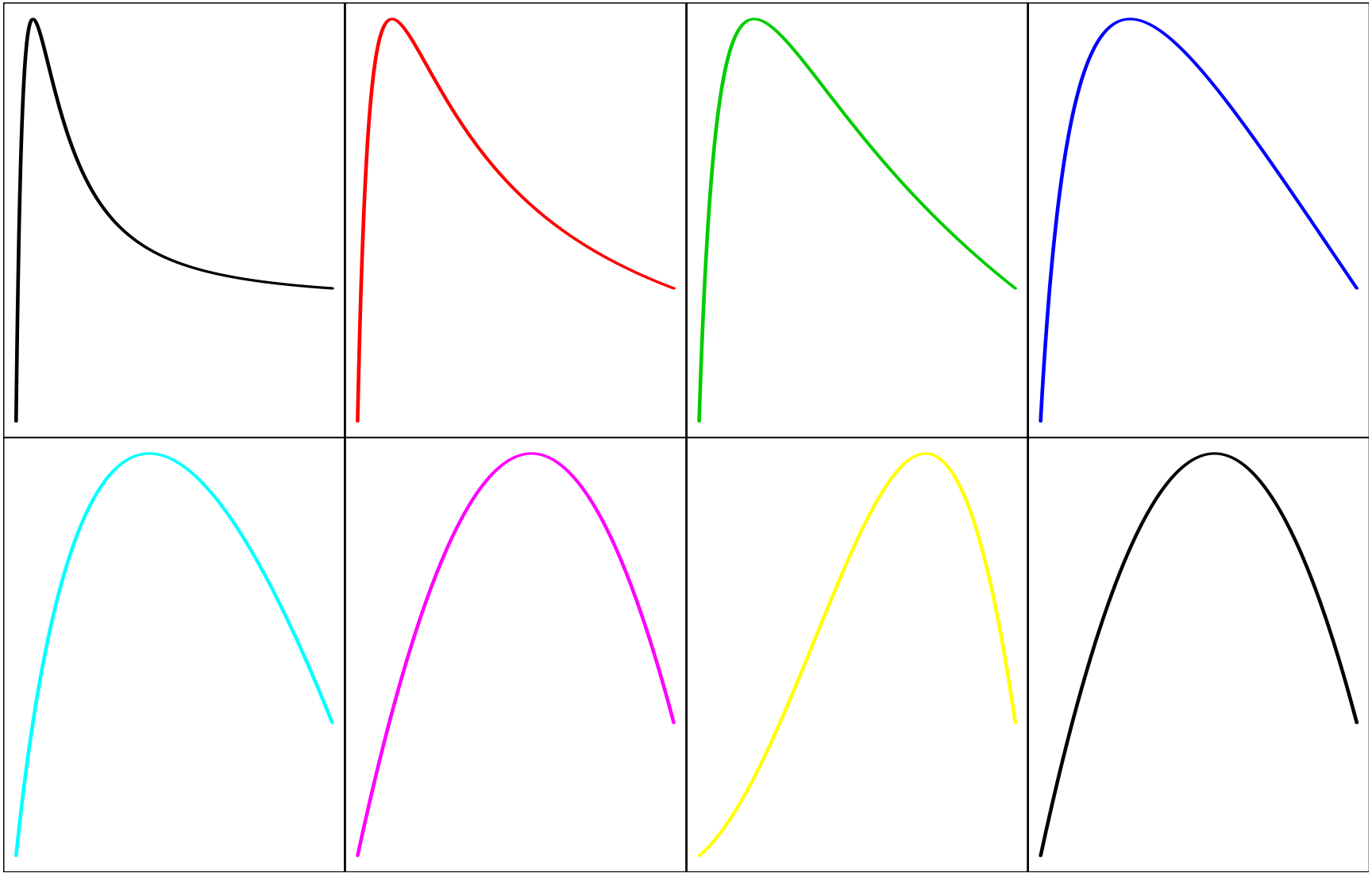
$\alpha \in \{-2, -1, -.5, 0, .5, 1, 2\}$

Algumas curvas possíveis FP1



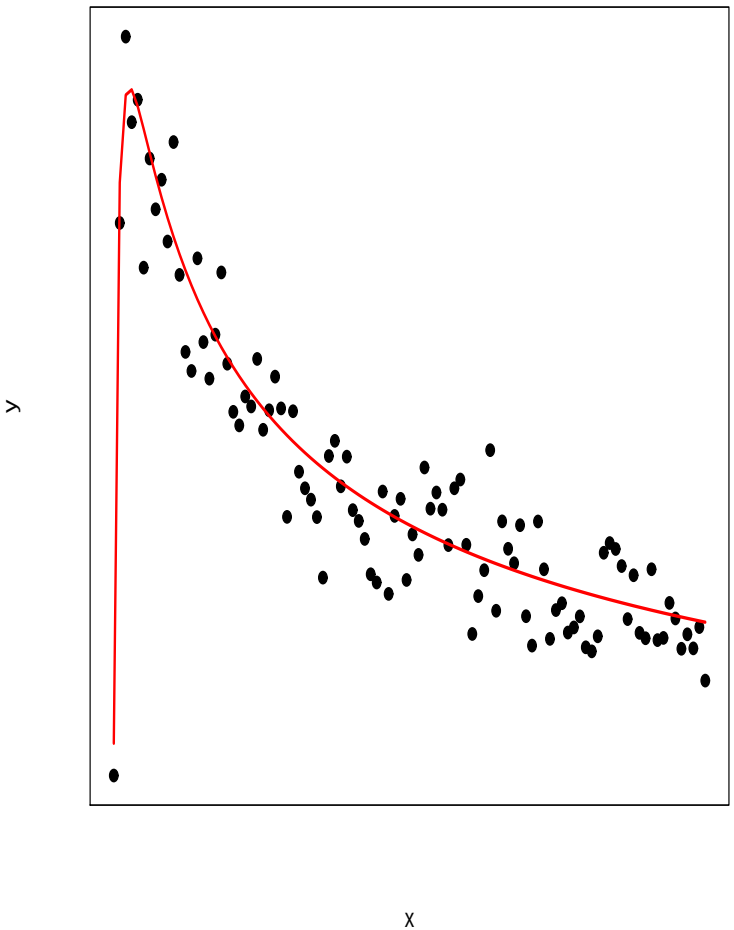
$\alpha \in \{-2, -1, -.5, 0, .5, 1, 2\}$

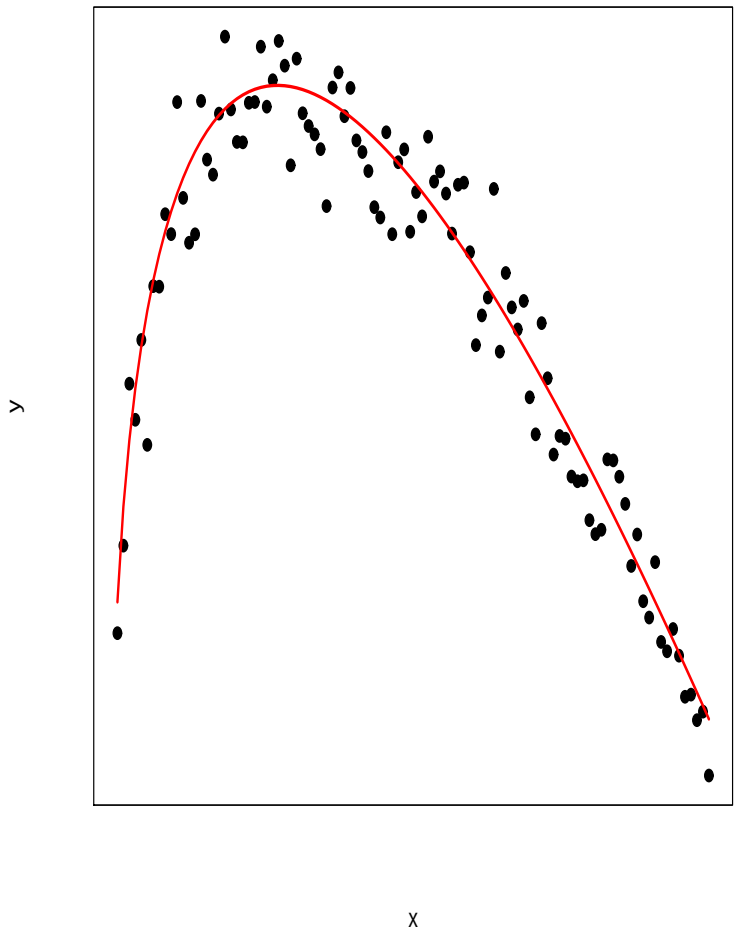
Algumas curvas possíveis FP2



$\alpha \in \{-2, -1, -.5, 0, .5, 1, 2\}$

Algumas curvas possíveis





Delineamento para FP's

Delineamentos experimentais clássicos (poucos níveis igualmente espaçados) são ineficientes (ou até singulares) para estimar os parâmetros dos FP's.

Delineamento para FP's

Delineamentos experimentais clássicos (poucos níveis igualmente espaçados) são ineficientes (ou até singulares) para estimar os parâmetros dos FP's.

Delineamento ótimo: encontra as combinações dos fatores (e repetições) que otimiza alguma propriedade para a estimação dos parâmetros.

Delineamento para FP's

Delineamentos experimentais clássicos (poucos níveis igualmente espaçados) são ineficientes (ou até singulares) para estimar os parâmetros dos FP's.

Delineamento ótimo: encontra as combinações dos fatores (e repetições) que otimiza alguma propriedade para a estimação dos parâmetros.

Como o modelo é não linear, a busca por um delineamento ótimo ou eficiente depende dos valores reais dos parâmetros do FP.

Matriz de informação - FP1

Informação para $\boldsymbol{\theta} = (\beta_0, \beta_1, \alpha)'$, por observação:

$$\mathbf{M}_i(\boldsymbol{\theta}, x_i) = \begin{pmatrix} 1 & x_i^{(\alpha)} & \beta_1 x_i^{(\alpha)} \log(x_i) \\ . & \left\{x_i^{(\alpha)}\right\}^2 & \beta_1 \left\{x_i^{(\alpha)}\right\}^2 \log(x_i) \\ . & . & \beta_1^2 \left\{x_i^{(\alpha)} \log(x_i)\right\}^2 \end{pmatrix}.$$

Matriz de informação - FP1

Informação para $\boldsymbol{\theta} = (\beta_0, \beta_1, \alpha)'$, por observação:

$$\mathbf{M}_i(\boldsymbol{\theta}, x_i) = \begin{pmatrix} 1 & x_i^{(\alpha)} & \beta_1 x_i^{(\alpha)} \log(x_i) \\ . & \left\{x_i^{(\alpha)}\right\}^2 & \beta_1 \left\{x_i^{(\alpha)}\right\}^2 \log(x_i) \\ . & . & \beta_1^2 \left\{x_i^{(\alpha)} \log(x_i)\right\}^2 \end{pmatrix}.$$

Informação para n observações

$$\mathbf{M}(\boldsymbol{\theta}, \mathbf{x}) = \sum_{i=1}^{n} \mathbf{M}_i(\boldsymbol{\theta}, x_i).$$

Matriz de informação - FP1

Informação para $\boldsymbol{\theta} = (\beta_0, \beta_1, \alpha)'$, por observação:

$$\mathbf{M}_i(\boldsymbol{\theta}, x_i) = \begin{pmatrix} 1 & x_i^{(\alpha)} & \beta_1 x_i^{(\alpha)} \log(x_i) \\ . & \left\{x_i^{(\alpha)}\right\}^2 & \beta_1 \left\{x_i^{(\alpha)}\right\}^2 \log(x_i) \\ . & . & \beta_1^2 \left\{x_i^{(\alpha)} \log(x_i)\right\}^2 \end{pmatrix}.$$

Informação para n observações

$$\mathbf{M}(\boldsymbol{\theta}, \mathbf{x}) = \sum_{i=1}^{n} \mathbf{M}_i(\boldsymbol{\theta}, x_i).$$

Para FP2, $\mathbf{M}$ tem dimensão 5×5.

Matriz de informação

Para dois fatores, $x_1,\ x_2 > 0$, modelo incluindo interação,

$$\eta(\mathbf{x}, \boldsymbol{\theta}) = \beta_0 + \beta_1 x_1^{(\alpha_1)} + \beta_{11} \left\{ x_1^{(\alpha_1)} \right\}^2 + \beta_2 x_2^{(\alpha_2)} + \beta_{22} \left\{ x_2^{(\alpha_2)} \right\}^2 + \beta_{12} x_1^{(\alpha_1)} x_2^{(\alpha_2)},$$

a matriz $\mathbf{M}$ é de dimensão 8×8.

Matriz de informação

Para dois fatores, $x_1,\ x_2 > 0$, modelo incluindo interação,

$$\eta(\mathbf{x}, \boldsymbol{\theta}) = \beta_0 + \beta_1 x_1^{(\alpha_1)} + \beta_{11} \left\{ x_1^{(\alpha_1)} \right\}^2 + \beta_2 x_2^{(\alpha_2)} + \beta_{22} \left\{ x_2^{(\alpha_2)} \right\}^2 + \beta_{12} x_1^{(\alpha_1)} x_2^{(\alpha_2)},$$

a matriz $\mathbf{M}$ é de dimensão 8×8.

Um critério de otimização de delineamento é o determinante de $\mathbf{M} \Rightarrow$ critério D.

Critério D

A função critério D ($det(\mathbf{M})$) depende

Critério D

A função critério D $(det(\mathbf{M}))$ depende

- de α para FP1;

Critério D

A função critério D $(det(\mathbf{M}))$ depende

- de α para FP1;
- de todos os parâmetros, exceto β_0, para FP2.

Informação a priori

- valores pontuais $\Rightarrow$ delineamentos localmente ótimos.

Informação a priori

- valores pontuais $\Rightarrow$ delineamentos localmente ótimos.
- discreta para α's e Normais para $\boldsymbol{\beta} \Rightarrow$ delineamentos ótimos pseudo-Bayesianos.

No segundo caso, a função critério é uma soma de integrais múltiplas.

Informação a priori

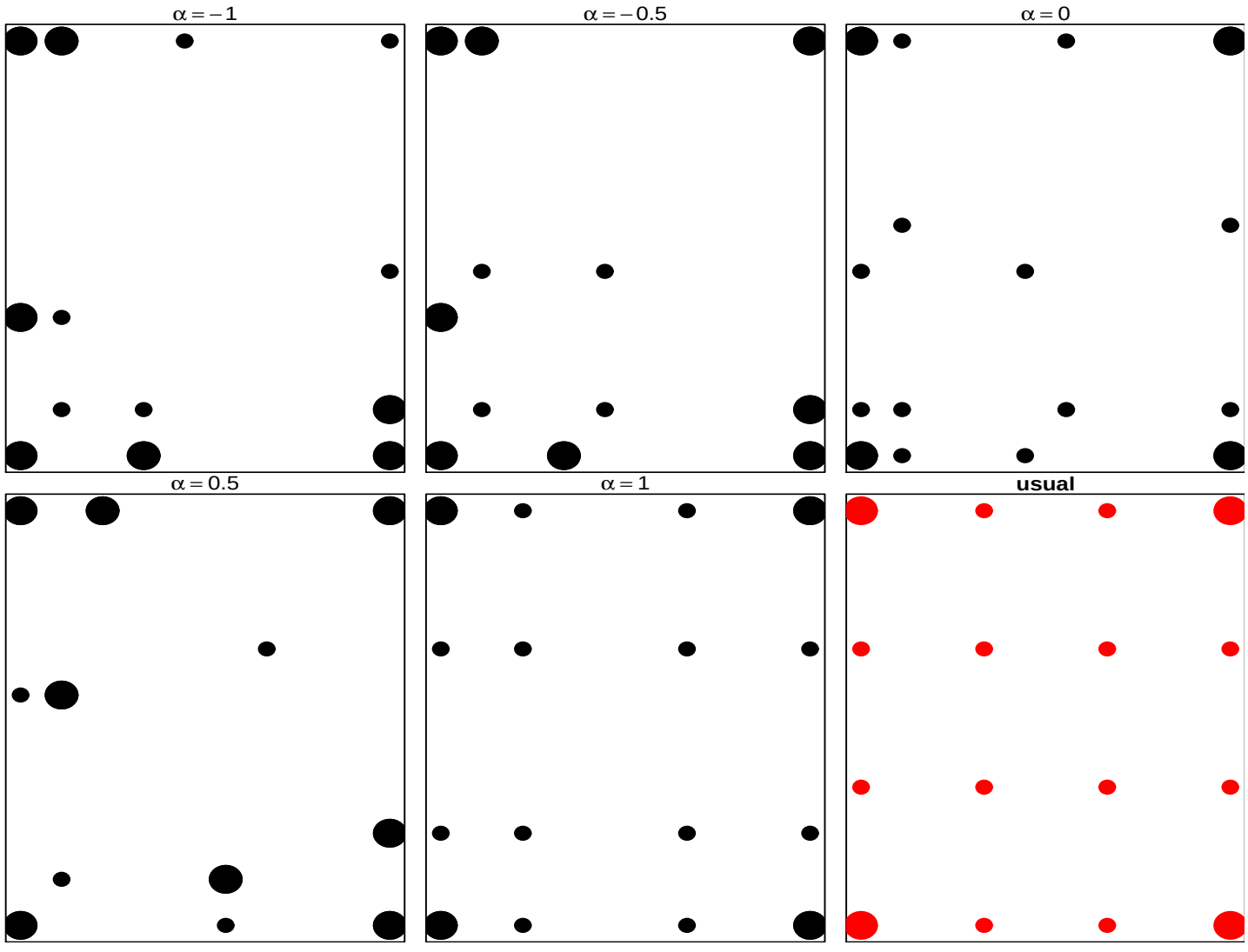
- valores pontuais $\Rightarrow$ delineamentos localmente ótimos.
- discreta para α's e Normais para $\boldsymbol{\beta} \Rightarrow$ delineamentos ótimos pseudo-Bayesianos.

No segundo caso, a função critério é uma soma de integrais múltiplas.

Aqui, a otimização foi num espaço discreto para os fatores e as integrais aproximadas por quadraturas.

Alguns resultados para dois fatores

Delineamentos localmente D-ótimos



$(n = 20,\ \beta_1 = \beta_2 = \beta_{12} = 1.0;\ \beta_{11} = \beta_{22} = -2.5)$

D-eficiências, localmente ótimos

	Delineamento ($\alpha_1 = \alpha_2$)				
$(\alpha_1,\ \alpha_2)$	-1.0	-0.5	0.0	$+0.5$	$+1.0$
$(-1.0,\ -1.0)$	100.0	98.5	90.9	80.2	50.9
$(-0.5,\ -0.5)$	99.4	100.0	97.3	87.1	64.9
$(\ 0.0,\ \ 0.0)$	92.8	96.7	100.0	94.4	81.2
$(+0.5,\ +0.5)$	81.9	89.6	97.4	100.0	97.0
$(+1.0,\ +1.0)$	62.9	72.4	82.4	93.6	100.0

D-eficiências, localmente ótimos

	Delineamento ($\alpha_1 = \alpha_2$)				
$(\alpha_1,\ \alpha_2)$	-1.0	-0.5	0.0	$+0.5$	$+1.0$
$(-1.0,\ -1.0)$	100.0	98.5	90.9	80.2	50.9
$(-0.5,\ -0.5)$	99.4	100.0	97.3	87.1	64.9
$(\ 0.0,\ \ 0.0)$	92.8	96.7	100.0	94.4	81.2
$(+0.5,\ +0.5)$	81.9	89.6	97.4	100.0	97.0
$(+1.0,\ +1.0)$	62.9	72.4	82.4	93.6	100.0
$(-1.0,\ -0.5)$	99.8	99.2	94.1	83.3	57.5
$(-1.0,\ \ 0.0)$	96.7	97.2	95.4	86.3	64.2
$(-1.0,\ +0.5)$	91.3	92.8	94.7	88.7	70.3
$(-1.0,\ +1.0)$	80.5	82.9	87.5	84.9	71.4
$(-0.5,\ \ 0.0)$	96.4	98.4	99.0	90.8	72.8
$(-0.5,\ +0.5)$	90.5	93.9	97.8	92.9	79.5
$(-0.5,\ +1.0)$	79.7	84.1	90.3	89.2	80.7
$(\ 0.0,\ +0.5)$	87.2	92.9	98.8	96.9	88.7
$(\ 0.0,\ +1.0)$	76.6	83.3	91.1	93.3	90.1
$(+0.5,\ +1.0)$	72.0	80.7	89.9	96.9	98.8
perda média	12.8	9.2	6.2	9.4	22.1
perda máxima	37.1	27.6	17.6	19.8	49.1

Delineamentos pseudo-Bayesiano ótimos

Distribuições de probabilidades a priori

	α				
Tipo	-1.0	-0.5	0.0	0.5	1
U_i	.20	.20	.20	.20	.20

Delineamentos pseudo-Bayesiano ótimos

Distribuições de probabilidades a priori

	α				
Tipo	-1.0	-0.5	0.0	0.5	1
U_i	.20	.20	.20	.20	.20
S_i	.10	.20	.40	.20	.10

Delineamentos pseudo-Bayesiano ótimos

Distribuições de probabilidades a priori

	α				
Tipo	-1.0	-0.5	0.0	0.5	1
U_i	.20	.20	.20	.20	.20
S_i	.10	.20	.40	.20	.10
R_i	.45	.30	.15	.07	.03

Delineamentos pseudo-Bayesiano ótimos

Distribuições de probabilidades a priori

	α				
Tipo	-1.0	-0.5	0.0	0.5	1
U_i	.20	.20	.20	.20	.20
S_i	.10	.20	.40	.20	.10
R_i	.45	.30	.15	.07	.03
L_i	.03	.07	.15	.30	.45

Delineamentos pseudo-Bayesiano ótimos

Distribuições de probabilidades a priori

	α				
Tipo	-1.0	-0.5	0.0	0.5	1
U_i	.20	.20	.20	.20	.20
S_i	.10	.20	.40	.20	.10
R_i	.45	.30	.15	.07	.03
L_i	.03	.07	.15	.30	.45

- Distribuições Normais para β.

Delineamentos pseudo-Bayesiano ótimos

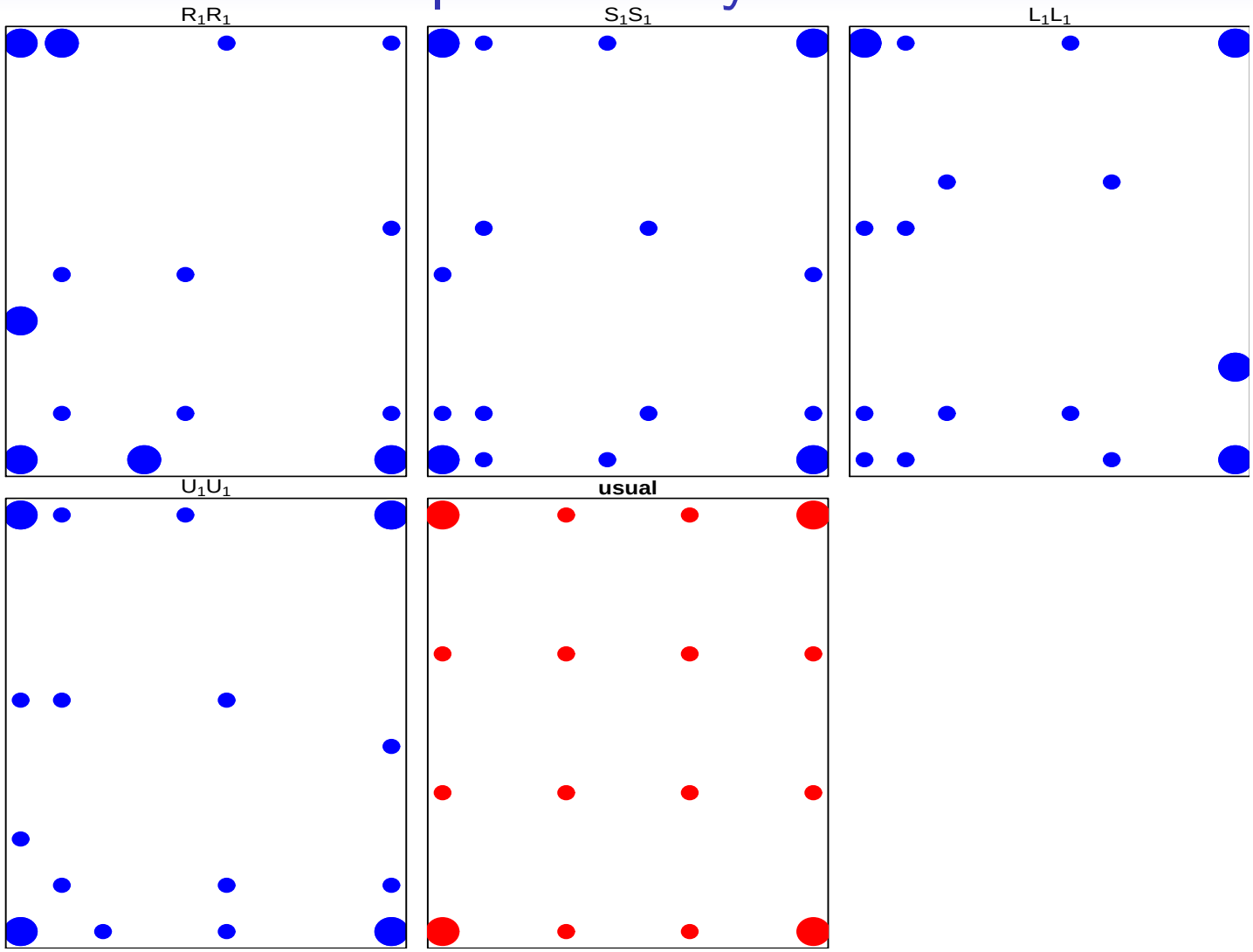
Distribuições de probabilidades a priori

	α				
Tipo	-1.0	-0.5	0.0	0.5	1
U_i	.20	.20	.20	.20	.20
S_i	.10	.20	.40	.20	.10
R_i	.45	.30	.15	.07	.03
L_i	.03	.07	.15	.30	.45

- Distribuições Normais para $\boldsymbol{\beta}$.
- Valores pontuais para $\boldsymbol{\beta}$.

Os delineamentos foram equivalentes para essas duas versões.

Delineamentos pseudo-Bayesiano D- ótimos

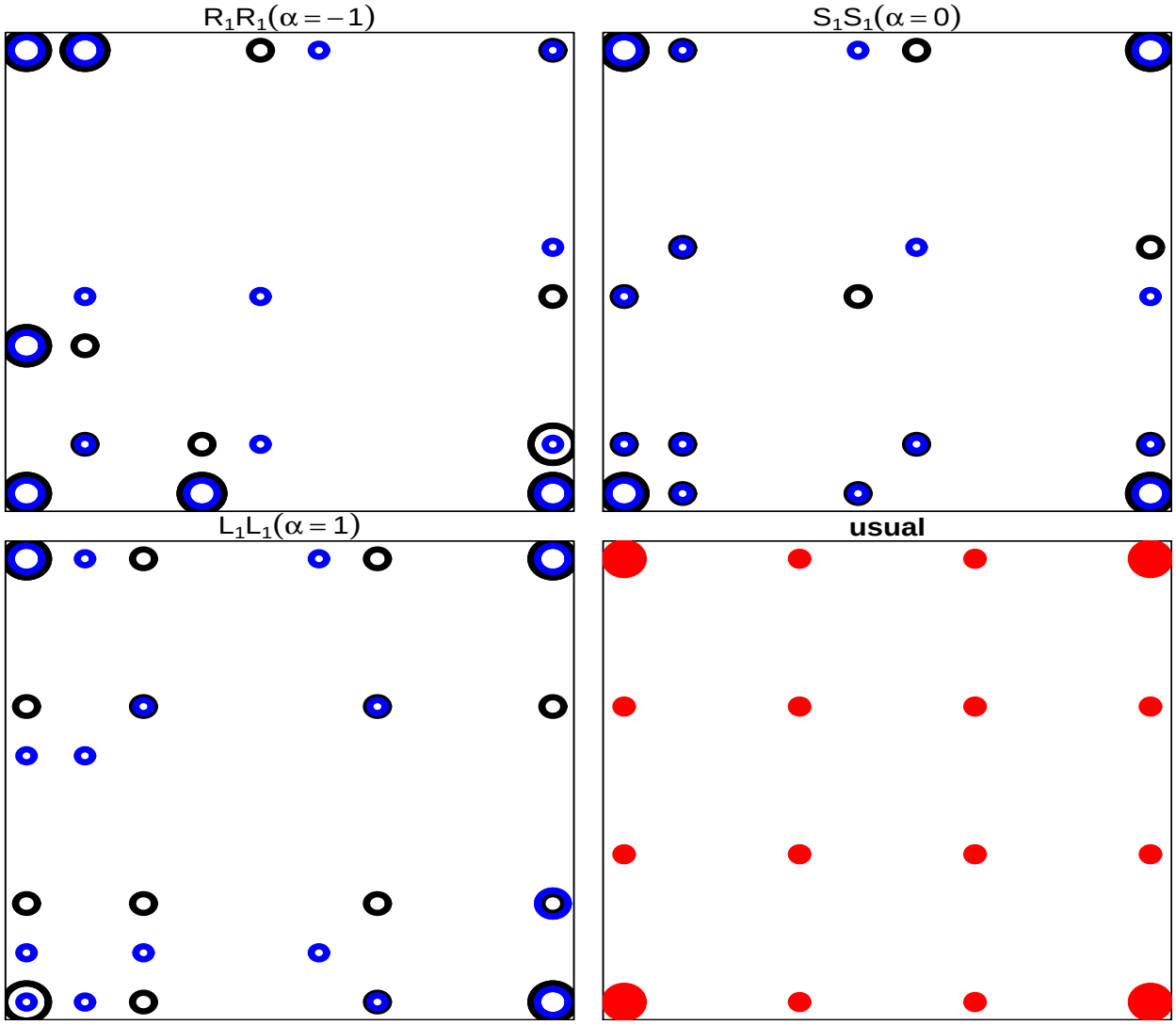


$(n = 20\ \beta_1,\ \beta_2,\ \beta_{12} \sim N(1.0,\ 0.2);\ \beta_{11},\ \beta_{22} \sim N(-2.5,\ 0.5))$

D-eficiências, pseudo-Bayesianos

	Delineamentos Pseudo-Bayesianos							
Priori	U_1U_2 (1)	U_1S_2 (2)	U_1R_2 (3)	U_1L_2 (4)	S_1S_2 (5)	R_1R_2 (6)	L_1L_2 (7)	L_1R_2 (8)
U_1U_2	100.0	99.7	98.6	97.8	99.5	97.1	95.8	96.8
U_1S_2	99.8	100.0	98.7	97.2	99.7	97.2	95.6	97.1
U_1R_2	98.4	98.7	100.0	93.7	98.3	98.5	90.8	99.1
U_1L_2	99.2	98.0	94.6	100.0	98.0	93.1	98.7	91.9
S_1S_2	99.7	99.9	98.6	96.9	100.0	97.4	95.1	96.4
R_1R_2	97.0	97.1	98.5	92.5	97.1	100.0	87.1	96.2
L_1L_2	98.3	97.1	93.6	99.3	96.4	89.1	100.0	92.9
L_1R_2	97.4	97.8	99.1	92.5	96.7	94.5	92.6	100.0
média	1.3	1.5	2.3	3.8	1.8	4.2	5.6	3.7
máx	3.0	2.9	6.4	7.5	3.6	10.9	13.0	8.1

Localmente versus pseudo-Bayesiano ótimos



$(n = 20\ \beta_1,\ \beta_2,\ \beta_{12} \sim N(1.0,\ 0.2);\ \beta_{11},\ \beta_{22} \sim N(-2.5,\ 0.5))$

Localmente versus pseudo-Bayesianos

Priori	Localmente ótimos ($\alpha_1 = \alpha_2 = \alpha$)				
	$\alpha = -1$	$\alpha = -.5$	$\alpha = 0$	$\alpha = .5$	$\alpha = 1$
R_1R_2	99.0	99.7	97.1	88.0	64.7
S_1S_2	92.6	96.6	100.0	95.6	81.5
L_1L_2	79.9	87.6	95.7	99.6	96.4
U_1U_2	91.8	95.9	99.3	95.8	81.2
U_1S_2	92.2	96.2	99.6	95.6	81.3
U_1R_2	95.4	97.6	98.3	91.7	72.4
U_1L_2	85.8	91.6	97.6	97.6	88.5
L_1R_2	89.3	92.8	96.7	93.1	78.9
perda média	9.3	5.3	2.0	5.4	19.4
perda máx	20.1	12.4	4.3	12.0	35.3

Comentários finais

- O delineamento "clássico" para fatores contínuos usa 3 níveis igualmente espaçados. Esse delineamento não permite estimação das potências.

Comentários finais

- O delineamento "clássico" para fatores contínuos usa 3 níveis igualmente espaçados. Esse delineamento não permite estimação das potências.Dependendo da transformação necessária, o delineamento com 4 níveis também é ineficiente.

- Para os casos estudados, os delineamentos obtidos se mostraram robustos em relação aos valores de $\boldsymbol{\beta}$.

- Os delineamentos pseudo-Bayesianos (distribuição a priori para α) se mostram mais robustos em relação aos ótimos locais.

- Para os valores de α considerados, no caso de total desconhecimento do tipo de transformação que será necessária, o uso de distribuição simétrica em zero apresenta as menores perdas de eficiência.

Referências

Box, G. E. P. and Tidwell, P. W. (1962). Transformation of the independent variables. *Technometrics*, **4**, 531-550.

Royston, P. and Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). *Applied Statistics*, **43**, 429-467.

Royston, P. and Altman, D. G. (1997). Approximating statistical functions by using fractional polynomial regression. *The Statistician*, **46**, 411-422.

Royston, P. and Sauerbrei, D. G. (2008). Multivariable Model-Building: A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. John Wiley & Sons, Ltd.