

# Delineamentos ótimos para experimentos com cana-de-açúcar

Alessandra de Lima Goes  
Orientadora: Renata Alcarde Sermarini

ESALQ

21/11/2018

## 1 Introdução

- Motivação
- Principais Delineamentos Utilizados no Melhoramento Genético
- Delineamentos Ótimos
- Medidas de Comparação dos Delineamentos ótimos

## 2 Objetivo

## 3 Estudo de Simulação

- Aspectos Gerais
- Resultados

## 4 Conclusão

## Etapas do Melhoramento Genético:

- Cruzamento de variedades existentes que é o ponto de partida para novos cultivares
- Produção de mudas
- Seleção de clones
- Avaliação de clones
- Comercialização da nova variedade

Experimentos com plantas de gerações iniciais



Maximização do Ganho Genético



Selecionar as melhores Test Lines



**Delineamentos Eficientes**

# Principais Delineamentos Utilizados no Melhoramento Genético

- Segundo KEMPTON (1984) um delineamento frequentemente empregado em ensaios para seleção de novas variedades em programas de melhoramento genético é o **grid-plot**.

## Grid-Plot

- Consiste em alocar parcelas contendo uma variedade padrão (*checks*) que são variedades já consolidadas no mercado entre as *test lines* não repetidas.

### **Vantagem:**

- Não é essencial aleatorizar as *test lines* podendo até mesmo utilizar um arranjo sistemático.

# Exemplos Grid-Plot

- Sistemático

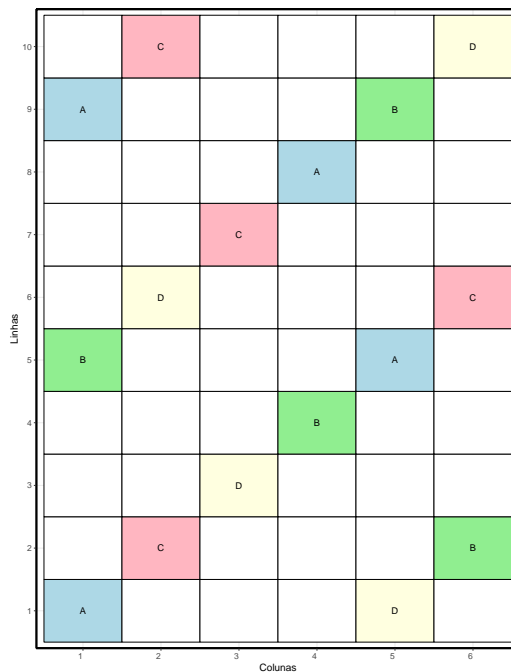


Figure 1: Grid-Plot Sistemático - 4 Checks

# Principais Delineamentos Utilizados no Melhoramento Genético

- FEDERER (1975), por sua vez, introduziu uma nova classe de delineamentos, chamados de delineamento aumentados, buscando melhores delineamentos experimentais nessas etapas iniciais onde não se tem material suficiente para a replicação.

## Delineamento Aumentado

- Usualmente neste tipo de delineamento as variedades padrões (*checks*) são aleatorizadas em blocos completos, ou até mesmo incompletos, e cada bloco é aumentado com as variedades em teste.
- A desvantagem desse delineamento é a necessidade de aleatorizar as *checks* dentro de cada bloco o que produz um arranjo irregular no campo e pode dificultar a avaliação visual.

# Exemplo Delineamento Aumentado

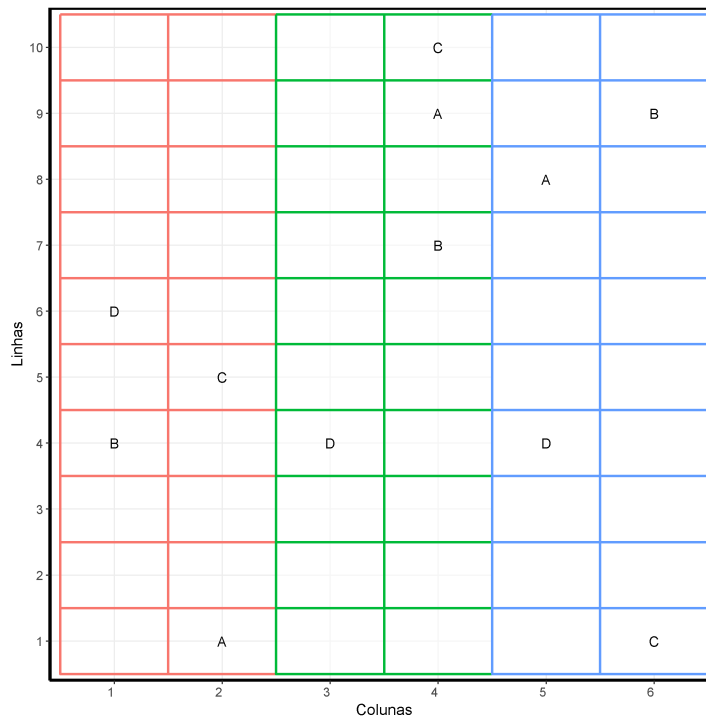


Figure 2: Delineamento Aumentado



# Principais Delineamentos Utilizados no Melhoramento Genético

- CULLIS et al. (2006) propuseram delineamentos onde no lugar do uso de *checks* as parcelas recebam uma certa proporção  $p$  das variedades de interesse, que denotaram como parcialmente replicados, **p-reps**.

## $p$ -reps

- Consiste em replicar  $p\%$  das *test lines*

### **Vantagem:**

- Todas as parcelas são usadas por *test lines*.

# Exemplo p-reps

Linhas	1	2	3	4	5	6
10	1	8	4		5	1
9		13		16	9	3
8	6		3	4		9
7			10			
6		7	12			
5	14				10	
4		15		2		8
3		7	12		13	15
2	14		16		6	
1	5	11	2			11

Figure 3: P-rep - 60 variedades de interesse e 14 replicadas (23,4%)

Um modo de se obter delineamentos eficientes para seleção de novos genótipos em estudo de melhoramento genético é a utilização de delineamentos ótimos, os quais dependem do critério de otimização selecionado.

Muitos desses critérios envolvem funções da matriz de informação que é dada por:

$$\mathbf{C} = \mathbf{W}'(\mathbf{ZDZ}' + \mathbf{R})^{-1}\mathbf{W} + \mathbf{G}^{-1} - \\
- \mathbf{W}'(\mathbf{ZDZ}' + \mathbf{R})^{-1}\mathbf{X}(\mathbf{X}'(\mathbf{ZDZ}' + \mathbf{R})^{-1}\mathbf{X})^{-1}\mathbf{X}' \times \\
\times (\mathbf{ZDZ}' + \mathbf{R})^{-1}\mathbf{W}$$

onde ,

**W** é a matriz de delineamento para os efeitos de tratamentos;

**Z** é a matriz de delineamento para os efeitos aleatórios, com matriz de variâncias **D**;

**X** é a matriz de delineamentos de efeitos fixos

**R** é a matriz de correlação residual e

**G** é a matriz de variâncias relacionada aos efeitos de tratamentos, se os tratamentos são de efeito fixo  $\mathbf{G}^{-1} = \mathbf{0}$

## Critério A

Que busca minimizar o traço da inversa da matriz de informação, isto é,

$$APEV \Rightarrow \text{minimiza } tr(\mathbf{C}^{-1})$$

## Critério C

Que minimiza a média das variâncias dos pares de diferenças dos efeitos de tratamentos (AVPD), ou *test lines*, isto é,

$$AVPD \Rightarrow \text{minimiza } \frac{2}{v-1} \left( tr(\mathbf{C}^{-1}) - \frac{1}{v} \mathbf{1}' \mathbf{C}^{-1} \mathbf{1} \right)$$

## Critério D

Que minimiza o determinante da inversa da matriz de informação, isto é,

$$PEGV \Rightarrow \text{minimiza } |\mathbf{C}^{-1}|$$

- Tais critérios dependem do modelo do delineamento, que de acordo com Butler, Cullis e Taylor (2014) deve ser o mais similar possível ao modelo de análise.

Os delineamento ótimos obtidos podem ser comparados com relação aos próprios critérios de obtenção. Entretanto, fazendo uso de estudos de simulação, critérios que representem a qualidade da seleção podem ser utilizados.

## CULLIS et al. (2006)

Fizeram o uso do Ganho Genético Percebido (RGG) que refere-se ao ganho que a cana-de-açúcar obteve após o melhoramento, selecionando uma porcentagem das *test lines*. O RGG é dado pela média dos  $s\%$  maiores valores preditos para as *test lines*, dividida pela média dos  $s\%$  maiores verdadeiros efeitos de *test lines*.

## WILLIAMS E PIEPHO (2013)

Adotaram o fator de eficiência. A eficiência é medida pela média harmônica dos autovalores não nulos da matriz de informação de tratamento ponderada pelo número de repetições.

Fouilloux and Lalole, 2001; Smith, Butler, and Cullis, 2011; Moehring, Williams, and Piepho, 2014

Uma medida alternativa para a comparação é a correlação entre os verdadeiros valores dos efeitos de tratamentos e os EBLUPs.

## Objetivo:

- Avaliar os diferentes delineamentos, entre eles o *grid-plot* e *p-rep*, considerando-se número diferentes de variedades padrão, diferentes porcentagens de replicação das variedades em teste e cenários para os dados.



# Estudo de Simulação - Aspectos Gerais

- Experimentos com 24 linhas e 10 colunas
- 180 *teste lines*
- Usando critério D e C
- Seis Delineamentos Distintos

**Table 1:** Caracterização dos tratamentos para os diferentes experimentos

	Nº de trat. não duplicados	Nº de trat. duplicados	Nº de Checks	Nº de repetição para cada check
<b>Del. 1</b>	180	0	4	15
<b>Del. 2 (p=11%)</b>	160	20	4	10
<b>Del. 3 (p=22%)</b>	140	40	4	5
<b>Del. 4 (p=33%)</b>	120	60	0	0
<b>Del. 5 (p=22%)</b>	140	40	2	10
<b>Del. 6 (p=22%)</b>	140	40	1	20

## Modelo proposto para obtenção do delineamento

- Assumimos tratamento de efeito fixo no modelo, por apresentar melhores resultados para o RGG, de acordo Sermarini et al.

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_o\mathbf{u}_o + \mathbf{e}$$

onde  $\mathbf{y}$  é o vetor de observações ( $n \times 1$ ),  $\boldsymbol{\tau}$  é um vetor ( $t \times 1$ ) de efeitos fixos dos tratamentos associado à matriz  $\mathbf{X}$  ( $n \times t$ ) de posto coluna completo,  $\mathbf{u}_o$  vetor ( $b \times 1$ ) de efeitos aleatórios não genéticos, e de linhas e colunas associado à matriz de delineamento  $\mathbf{Z}_o$  ( $n \times b$ ), e  $\mathbf{e}$  vetor ( $n \times 1$ ) de resíduos aleatórios.

Além disso

$$E \begin{bmatrix} \mathbf{u}_o \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

e matriz de variâncias:

$$\text{Var} \begin{bmatrix} \mathbf{u}_o \\ \mathbf{e} \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{G}_o(\boldsymbol{\gamma}) & \mathbf{0} \\ \mathbf{0} & \mathbf{R}(\boldsymbol{\phi}) \end{bmatrix}$$

Foram adotados  $\sigma^2 = 1$ ,  $\gamma_c = \gamma_r = 0.1$ ,  $\phi_c = 0.6$  e  $\phi_r = 0.4$  e  $\mathbf{R} = \text{AR1}(\phi_r) \otimes \text{AR1}(\phi_c)$ .

## Encontrando Delineamentos Ótimos

- D-ótimo foi obtido usando o PROC OPTEX do software SAS.
- C-ótimo foi obtido usando a biblioteca od (BUTLER,2014)

## Para os delineamentos usando D-ótimo e C-ótimo

- *test lines* foram consideradas de efeitos fixos
- 1000 interações aleatórias

# Modelos Proposto Simulação dos dados

Os dados foram simulados a partir do seguinte modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_o\mathbf{u}_o + \boldsymbol{\epsilon},$$

em que  $\boldsymbol{\beta}$  é o vetor de efeitos fixos, com matriz de delineamento  $\mathbf{X}$ ,  $\mathbf{u}_g$  é o vetor de efeitos genéticos (test lines) aleatório, com matriz de delineamento  $\mathbf{Z}_g$ ,  $\mathbf{u}_o$  é o vetor de efeitos aleatórios de linhas e colunas, com correspondente matriz de delineamento  $\mathbf{Z}_o$  e  $\boldsymbol{\epsilon}$  é o vetor de erros. Adicionalmente,

$$E = \begin{bmatrix} \mathbf{u}_g \\ \mathbf{u}_o \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

## Modelos Proposto Simulação dos dados

e matriz de variancia:

$$\text{Var} \begin{bmatrix} \mathbf{u}_g \\ \mathbf{u}_o \\ \epsilon \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{G}_g(\gamma_g) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_o(\gamma_o) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}(\phi) \end{bmatrix}$$

em que  $\mathbf{G}_o(\gamma_o)$  e  $\mathbf{R}(\phi)$  são descritos conforme o modelo de obtenção de delineamentos e  $\mathbf{G}_g(\gamma_g) = \gamma_g \mathbf{I}$ , em que  $\mathbf{I}$  é a matriz identidade.

Foram utilizados  $\gamma_c = \gamma_r = 0.1$  e  $\gamma_g = 0.8$  e também  $\gamma_g = 0.3$ . Os demais parâmetros serão conforme o modelo para o delineamento.

Foram geradas 1000 amostras de uma distribuição normal padrão utilizando o software R e para cada variância genética foi feita uma transformação na variável.

A análise dos dados foi feita a partir do software SAS seguindo o modelo para a simulação dos dados. Desse modo, quando presentes, as variedades *checks* foram consideradas de efeito fixo; as *test lines* foram assumidas de efeito aleatório, bem como os efeitos globais de linha e coluna; ainda, para os resíduos, foram ajustados modelos separáveis auto regressivos de primeira ordem.

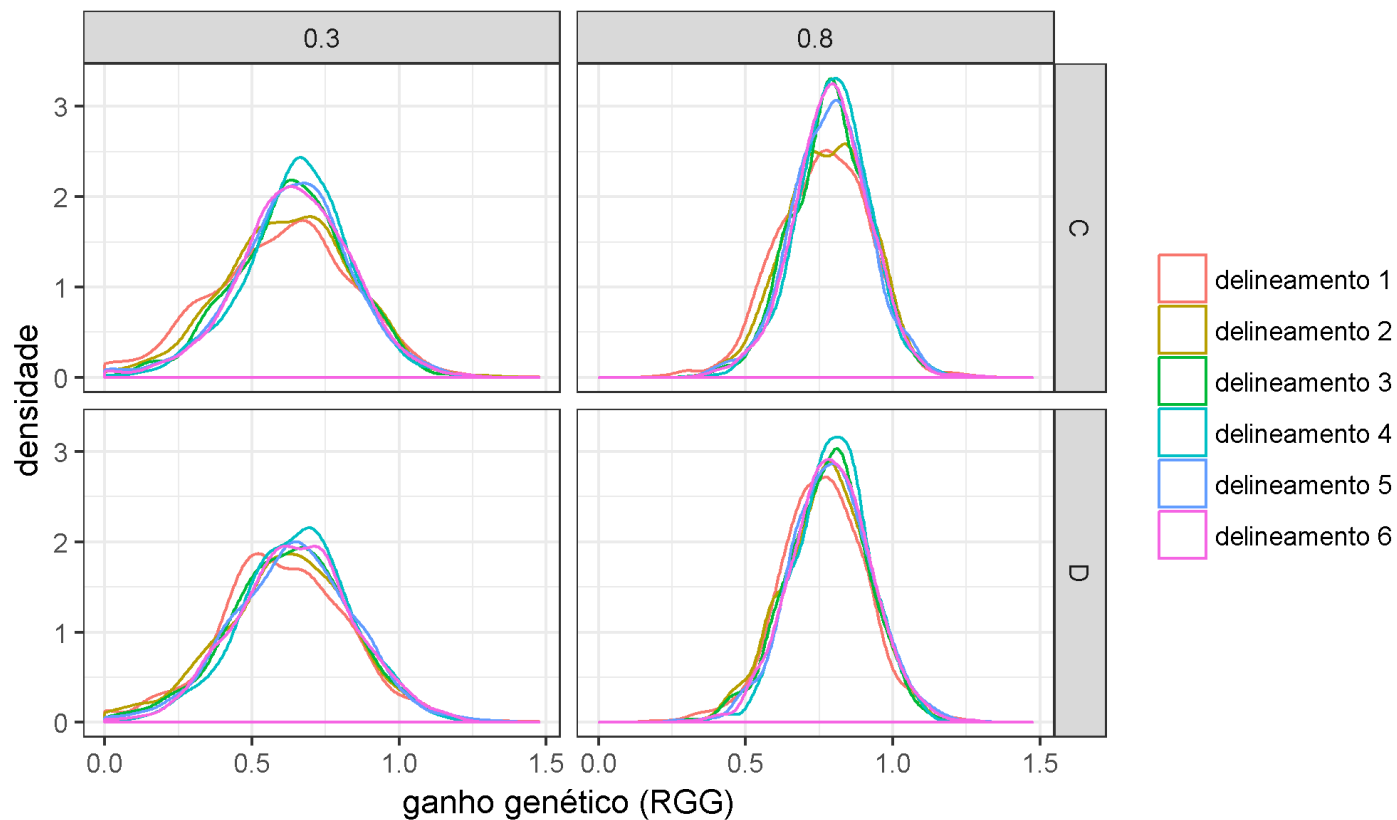
## Medidas de Comparação

- **Realized Genetic Gain** – Que foi calculado pela média dos 10% maiores valores preditos para as *test lines*, dividida pela média dos 10% maiores verdadeiros efeitos de *test lines*.
- **Correlação** - Entre os valores preditos e os verdadeiros valores
- **Assertividade** - Dada pela porcentagem das melhores *test lines* que foram selecionadas pelo modelo.



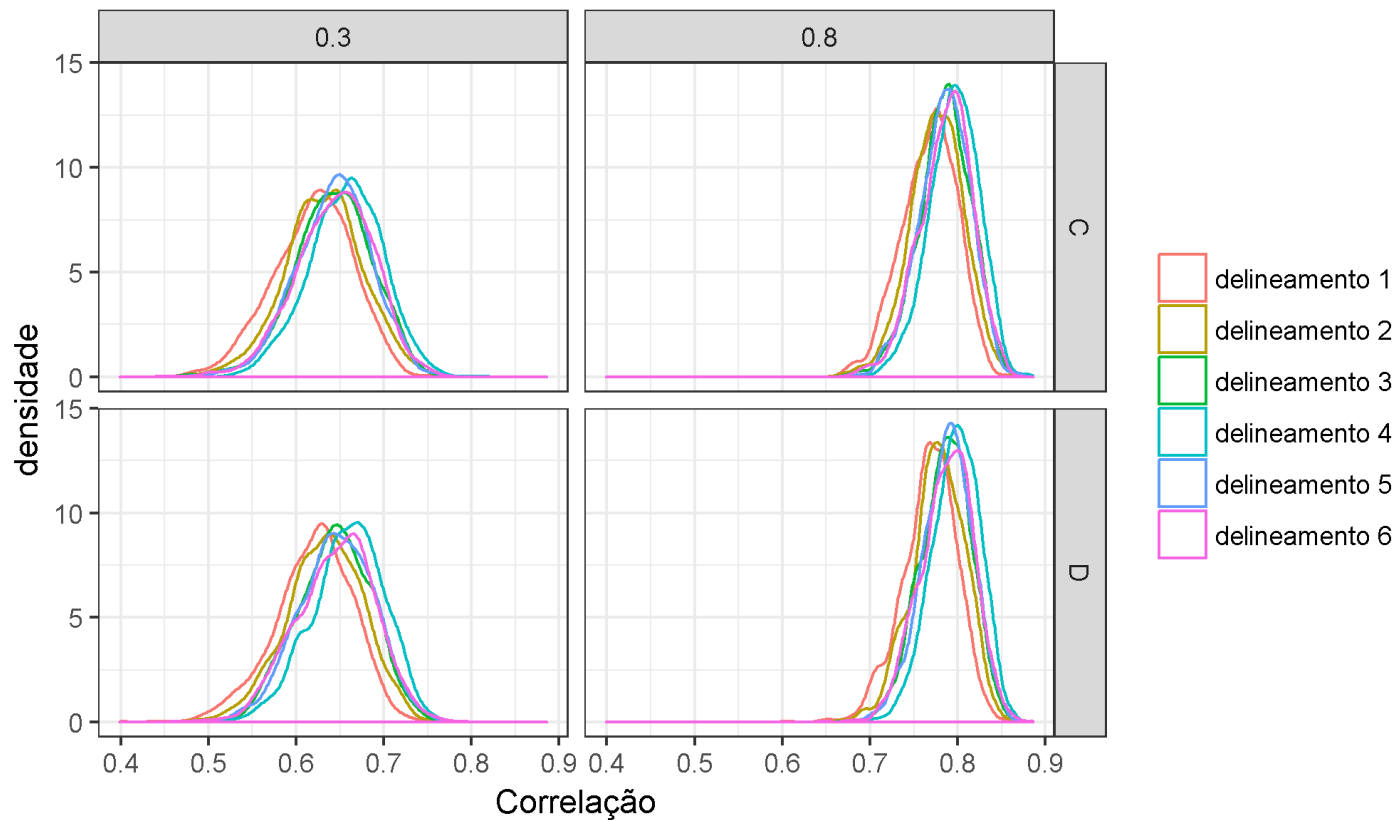
# Resultados

## ● RGG



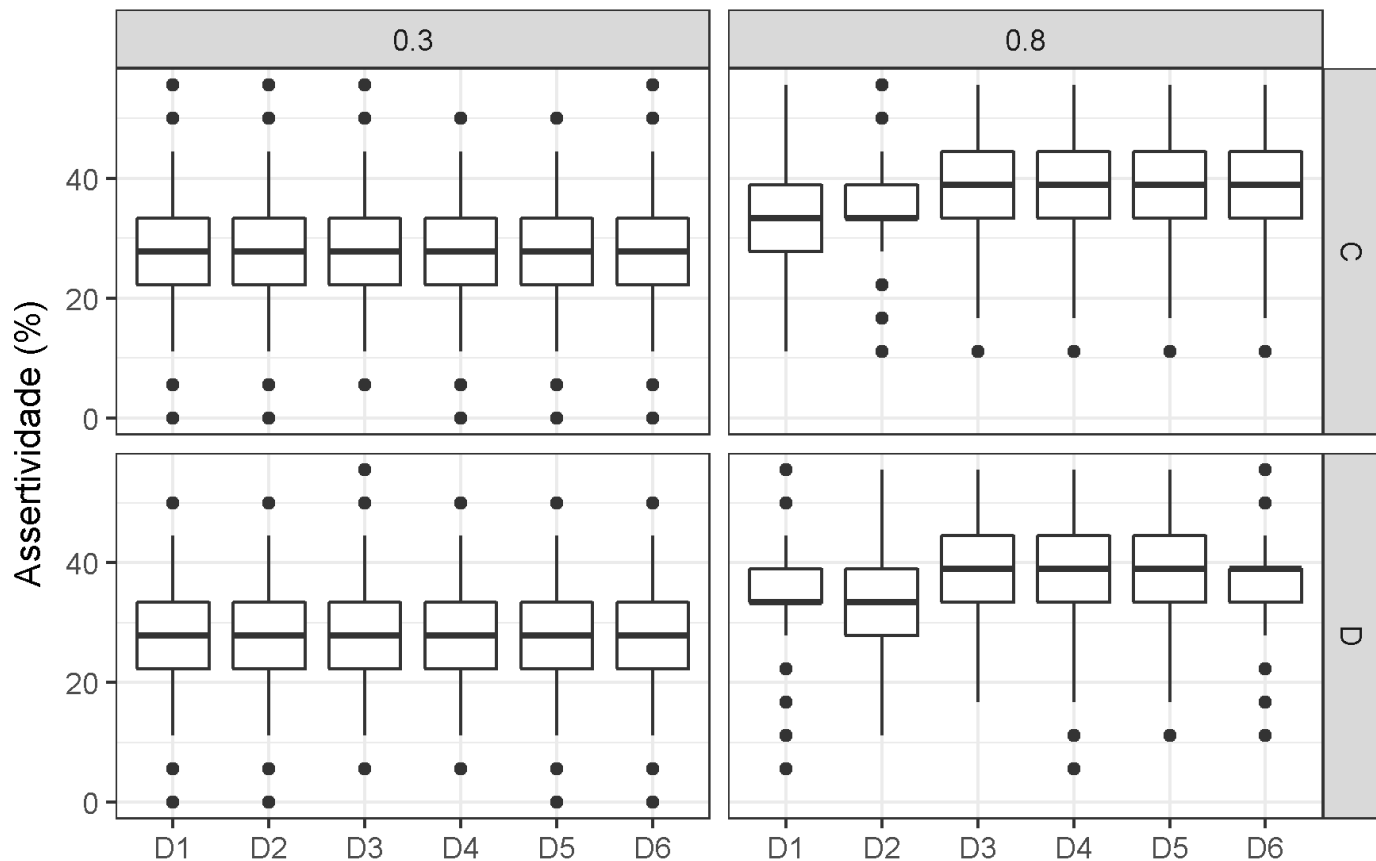
# Resultados

## ● Correlação



# Resultados

## ● Assertividade



- Não há precisamente uma grande diferença entre os critérios de otimização ;
- Quanto maior a variância genética melhor a seleção;
- Os delineamentos p-rep apresentam um melhor desempenho, principalmente na presença de baixa variância genética, para os cenários avaliados.

# Principais Referências

- CLARKE, G.P.; STEFANOVA, K.T. (2011). Optimal Design for Early-Generation Plant-Breeding Trials with Unreplicated or Partially Replicated Test Lines. *Australian & New Zealand Journal of Statistic*. v. 53, n. 4, p. 461-480.
- CULLIS, B.R.; SMITH, A.B., COOMBES, N.E. (2006). On the Design of Early Generation Variety Trials With Correlated Data. *Journal of Agricultural, Biological, and Environmental Statistics*. v. 11, n. 4, p. 381-393.
- KEMPTON, R.A.; TALBOT, M. (1988). The Development of New Crop Varieties. *Journal of Royal Statistical Society*. v. 151, n. 2, p. 327-341.
- WILLIAMS, E. R.; PIEPHO, H.-P. (2013). A comparison of spatial designs for old variety trials”. In: *Australian and New Zealand Journal of Statistics* 55.3, pp. 253-258.
- BUTLER D.; Generate optimal experimental designs. Package ‘OptimalDesign’. Maio, 2016.
- SANTOS A. ; Design and analysis of sugarcane breeding experiments : a case study. 2017

**OBRIGADA!!**