



Sequenciamento sistemático de experimentos fatoriais como alternativa à ordem aleatória

Systematic sequencing of factorial experiments as an alternative to the random order

Pedro Carlos Oprime¹
Vitória Maria Miranda Pureza¹
Samuel Conceição de Oliveira²

Resumo: Este trabalho tem como objetivo principal discutir o uso de métodos sistemáticos para geração de designs de experimentos com boas propriedades estatísticas e custos baixos. O foco da pesquisa é o sequenciamento dos experimentos, de maneira que são analisados os resultados de três diferentes abordagens para construção de designs fatoriais (ortogonais e não ortogonais) com dois níveis, em que o sequenciamento é feito de forma aleatória ou sistemática. Em particular, simulou-se a condução do design gerado por cada abordagem no contexto de um processo real de fabricação de embalagens de vidro, sem a presença de efeitos de tendências lineares e com a presença desses efeitos. Os resultados das análises indicam que em relação à ordem aleatória, sequências sistemáticas podem resultar em menor número de mudanças de níveis dos fatores e maior robustez a efeitos de tendências lineares, compatibilizando, portanto, o custo e a qualidade do design.

Palavras-chave: Projeto fatorial de experimentos; Sequências aleatórias e sistemáticas; Tendências lineares; Custo; Simulação.

Abstract: *The current study aims to discuss the use of systematic methods to generate experimental designs with good statistical properties and low costs. The research focuses on the sequence of experiments and on analysis the results of three different approaches used to build (orthogonal and non-orthogonal) two-level factorial designs, wherein sequencing is randomly or systematically performed. The study simulated the design generated by each approach in the context of an actual glass container manufacturing process, with and without the presence of linear trend effects. The results indicate that, in comparison to the random order, systematic sequences may lead to fewer factor level changes and to increased robustness to linear trend effects. Therefore, they may attach design cost and quality.*

Keywords: *Factorial design of experiments; Random and systematic sequences; Linear trends; Cost; Simulation.*

1 Introdução

O planejamento de experimentos (DoE) é uma das técnicas estatísticas mais utilizadas em projetos de melhoria e de desenvolvimento de produtos e processos, tendo ampla divulgação com o movimento da Qualidade Total. No Brasil, sua divulgação deu-se no final da década de 1980 e início de 1990, quando surgiram os primeiros conceitos de qualidade segundo o modelo japonês. Sobre esse tema, há vários livros (Toledo, 1986; Imai, 1997; Kume, 1993) e artigos que relacionam o uso dessas técnicas à melhoria contínua (Marin-Garcia et al., 2008; Oprime et al., 2010), bem como os efeitos dessas atividades na

produtividade (Bessant & Caffyn, 1997; Savolainen, 1999; Harrison, 2000; Bessant et al., 2001; Delbridge & Barton, 2002; Hyland et al., 2003).

Em linhas gerais, o DoE é definido como uma combinação de experimentos (tratamentos) planejados que permite relacionar o efeito de um conjunto de níveis (valores) de fatores (variáveis) independentes sobre uma ou mais variáveis de resposta dependentes, julgadas de interesse. A partir desses experimentos, é possível aplicar testes estatísticos sobre a significância dos efeitos dos fatores, bem como desenvolver modelos matemáticos empíricos que permitam, para o intervalo

¹ Universidade Federal de São Carlos – UFSCar, Rodovia Washington Luís, Km 235, SP-310, CEP 13565-905, São Carlos, SP, Brasil, e-mail: pedro@dep.ufscar.br; toledo@dep.ufscar.br

² Departamento de Bioprocessos e Biotecnologia, Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista – UNESP, Rodovia Araraquara-Jaú, Km 1, CEP 14801-902, Araraquara, SP, Brasil, e-mail: samueloliveira@fcfar.unesp.br

Recebido em Fev. 14, 2015 - Aceito em Set. 2, 2015

Suporte financeiro: Nenhum.

de experimentação considerado, prever os efeitos de determinadas combinações desses fatores sobre as variáveis de resposta do sistema (Davis, 1956; Box et al., 1978; Montgomery, 1991).

A partir do trabalho pioneiro de Fisher (1926), a literatura especializada vem mostrando constante evolução das técnicas de DoE. Até 1950, as pesquisas concentraram-se em arranjos ortogonais, com pouco avanço em arranjos não ortogonais (Addelman, 1972). A partir de 1965, um maior número de publicações sobre DoE abordaram aspectos relacionados ao fracionamento de experimentos regulares (associado ao conceito clássico de experimentos ortogonais). Estudos com experimentos irregulares foram introduzidos no início da década de 1970, buscando um compromisso entre a qualidade do planejamento e os custos experimentais associados. Com isso, maior foco foi dado a *designs* irregulares fracionados (associado aos experimentos não ortogonais) e em estudos sobre ordem de execução (sequenciamento) dos tratamentos (Atkinson & Bailey, 2001).

De modo geral, identificam-se três problemas básicos de pesquisa em DoEs: (i) a seleção de experimentos para composição de *designs* ótimos, ou seja, que produzam o menor erro de estimativas dos efeitos e parâmetros estatísticos do modelo empírico; (ii) a definição do sequenciamento de experimentos previamente definidos, de forma a minimizar o custo de execução dos experimentos; e (iii) o planejamento de *designs* (seleção e sequenciamento) de experimentos robustos a efeitos de tendências lineares (ou seja, a efeitos do tempo).

No que diz respeito à geração de *designs* ótimos, procura-se minimizar a variância associada aos erros de estimativa dos efeitos dos tratamentos pela maximização do determinante $|X'X|$, em que X é a matriz do *design*. Quanto maior é o determinante, menor é o erro de estimativa dos efeitos e dos parâmetros estatísticos dos modelos de regressão múltipla utilizados na aplicação da técnica de superfície de resposta. Maiores detalhes sobre a construção de experimentos ótimos podem ser encontrados em Dykstra (1971), Galil & Kiefer (1980), Aggarwal et al. (2003), Street & Burgess (2008), Wilmut & Zhou (2011), Alonso et al. (2011) e Suen & Midha (2013).

Os outros dois problemas de pesquisa, por sua vez, consideram que a ordem de execução dos experimentos impacta não só o custo de transição entre os experimentos (Daniel & Wilcoxon, 1966; Draper & Stoneman, 1968; Cheng, 1990; Wang, 1991; Wang & Jan, 1995; Wang & Chen, 1998; Garroi et al., 2009), mas também a robustez do *design*, uma vez que as estimativas dos efeitos principais e das interações ao longo da execução dos experimentos podem ser suscetíveis a variáveis não controladas e produzir vieses nas estimativas (Hilow, 2013). A ordem sistemática (não aleatória) dos experimentos tem, portanto, grande

relevância prática, em função de seu impacto nessas duas medidas. Nesse sentido, a sistematização da ordem de execução dos experimentos confronta-se diretamente com um dos principais paradigmas do DoE: a aleatorização do sequenciamento (Box et al., 1978; Montgomery, 1991; Montgomery et al., 2009).

Alinhado a esse debate, o presente trabalho traz um estudo comparativo de três abordagens de geração de *designs* de experimentos fatoriais de dois níveis: (i) geração de *designs* ortogonais e não ortogonais com a técnica DETMAX (veja Cook & Nachtsheim (1980) para o detalhamento matemático da técnica), seguida do sequenciamento aleatório dos experimentos; (ii) construção sistemática de *designs* ortogonais robustos a efeitos de tendências lineares com o algoritmo de Angelopoulos et al. (2009); e (iii) sequenciamento sistemático dos *designs* ortogonais e não ortogonais gerados pela técnica DETMAX pela resolução de um modelo de programação matemática com um método de otimização exato.

As abordagens foram aplicadas a seis exemplos com 12 a 28 experimentos e 4 e 5 fatores, sendo os *designs* resultantes analisados segundo quatro critérios: D-eficiência, contagem de tempo, correlação dos fatores com o tempo e número de mudanças de fatores. Em seguida, a ausência e a presença de efeitos de tendências lineares foram simuladas tomando-se uma aplicação real de DoE com 16 experimentos e 5 fatores, considerando-se os *designs* das três abordagens. Os erros tipo I e II foram então avaliados para cada *design*.

O restante deste artigo é organizado como se segue. A Seção 2 apresenta uma breve fundamentação teórica sobre experimentos fatoriais de dois níveis na presença de efeitos de tendências lineares e a revisão bibliográfica. Na Seção 3, são descritas as abordagens sistemáticas e aleatória consideradas no estudo, enquanto a Seção 4 discute os resultados com essas abordagens para os seis exemplos. Na Seção 5, é descrito o procedimento de simulação e analisados os resultados das abordagens para o caso real. Finalmente, a Seção 6 apresenta as conclusões do estudo e perspectiva de pesquisa futura.

2 *Designs* de experimentos fatoriais de dois níveis na presença de efeitos de tendências lineares

Uma maneira prática de planejar experimentos é construir experimentos fatoriais. Sua principal vantagem é a de permitir a análise de um grande número de fatores simultaneamente, permitindo a identificação do efeito de cada fator na variável resposta, assim como o efeito de interações entre fatores. Em experimentos fatoriais, o experimentador seleciona um número fixo de níveis para cada um

dos k fatores e executa experimentos com todas as combinações de níveis.

Experimentos fatoriais 2^k (experimentos com k fatores, cada qual com dois níveis) é uma classe de planejamento experimental bastante utilizada na indústria, cujo modelo matemático é dado pela seguinte Equação 1 geral:

$$y = \beta_0 + \sum_{k=1}^p \beta_k x_k + \sum_{k=1}^p \sum_{k'=1}^p \beta_{kk'} x_k x_{k'} + \varepsilon \quad (1)$$

em que o parâmetro β_0 é a média global, β_k são os parâmetros referentes aos efeitos principais, $\beta_{kk'}$ são os parâmetros referentes às interações entre cada dois fatores (2^a ordem) e ε é o erro experimental aleatório. Nos casos nos quais há um grande número de fatores, um fracionamento do planejamento completo torna-se conveniente, à custa, entretanto, da sobreposição dos efeitos principais e de interações de 2^a ordem. Os níveis de confundimento (sobreposição) dos efeitos determinam o grau de resolução do planejamento. Deste modo, quanto maior o confundimento menor será a resolução do delineamento (Box et al., 1978; Montgomery, 1991). Um delineamento fatorial fracionário de dois níveis é denotado como fatoriais 2^{k-p} , em que p indica o fracionamento do experimento.

Há três propriedades centrais na construção de experimentos com boas propriedades estatísticas: a ortogonalidade e o balanceamento da matriz X do *design* de experimentos e a robustez a efeitos de tendências lineares. Em *designs* ortogonais (ou semiortogonais), a matriz $X'X$ é diagonal, ou seja, seus elementos têm valores iguais a zero (ou próximos de zero) fora da diagonal. Assim, a variância da estimativa dos parâmetros do modelo (1) (estimativa esta dada por $\hat{\beta} = (X'X)^{-1} X'y$) é minimizada e a correlação entre os fatores de X é zero (ou próxima zero) (Dykstra, 1971; Mitchell, 1974; Galil & Kiefer, 1980). Uma matriz balanceada, por sua vez, tem o mesmo número de níveis em cada fator (Addelman, 1972; Adekeye & Kunert, 2005).

A máxima eficiência ocorre quando X é balanceada e ortogonal, o que define diferentes medidas de eficiência que fornecem informação sobre a qualidade das estimativas dos parâmetros do modelo. A chamada *D-eficiência*, utilizada em Atkinson (1996), Tack & Vandebroek (2004), Atkinson et al. (2007), Triefenbach (2008) e Alonso et al. (2011), é calculada pela Equação 2 dada por

$$D_{eff} = \frac{|XX'|^{\frac{1}{p}}}{N} \quad (2)$$

em que N é o número de experimentos do *design* e p é o número dos parâmetros do modelo. Note que $0 \leq D_{eff} \leq 1$ e que o ideal é que D_{eff} seja igual a 1.

No que diz respeito à minimização de possíveis efeitos de tendências lineares, abordagens clássicas prescrevem que a sequência de realização dos

experimentos seja produzida de forma aleatória. Ressalta-se que, além da aleatorização, outro procedimento fundamental é o diagnóstico do modelo para verificar o comportamento da distribuição dos resíduos (possíveis efeitos de tendências lineares) e a adequação do modelo aos dados experimentais (Box et al., 1978).

Como alternativa à aleatorização, o sequenciamento dos experimentos pode ser definido de forma sistemática, considerando-se o critério de *contagem de tempo* (TC) (Draper & Stoneman, 1968). Este critério mede explicitamente a correlação entre os fatores da matriz XX e o tempo (ou ordem de execução dos tratamentos). Para um *design* com k fatores, 2 níveis e N experimentos, ele é dado pela Equação 3:

$$TC = \text{Max}\left\{ \sum_{j=1}^N i^* u_{ij} \right\} \quad (3)$$

em que $u_{ij} \in \{-1, +1\}$ denotam o nível (inferior e superior, respectivamente) do fator j no i -ésimo experimento realizado. Quando a contagem de tempo é igual a zero para um determinado fator, não há correlação entre esse fator e ordem temporal de execução dos experimentos; ou seja, o fator não é suscetível a efeitos de tendências lineares. Em outras palavras, não haverá viés na estimativa do parâmetro do modelo de regressão associado ao fator. A correlação ρ é obtida a partir de TC pela Equação 4 (Angelopoulos et al., 2009):

$$\rho = \frac{TC}{N \sqrt{\left(\frac{N^2 - 1}{12} \right)}} \quad (4)$$

Ou seja, quanto maior o valor de TC , maior é a correlação ρ .

Além da qualidade estatística, o custo da experimentação é claramente um critério que influencia a definição do planejamento de experimentos em contextos práticos. No caso geral, o custo é dado pelo *número de mudanças dos níveis dos fatores* (NFC) conforme os experimentos são executados (Draper & Stoneman, 1968). Apesar de, na prática, o custo de transição poder ser diferente para cada fator, admite-se que quanto maior NFC , mais caro é o *design*. Para um *design* com k fatores, 2 níveis e N experimentos, esse critério é formalizado por (Equação 5)

$$NFC = \sum_{i=1}^{N-1} \sum_{j=1}^k |u_{ij} - u_{(i+1)j}| \quad (5)$$

em que, mais uma vez, $u_{ij} \in \{-1, +1\}$ denotam o nível (inferior e superior, respectivamente) do fator j no i -ésimo experimento realizado.

Nos últimos 50 anos, um grande esforço de pesquisa tem sido observado no estudo ou proposição de métodos para produção de *designs* que levem

em conta um ou mais critérios dentre aqueles supracitados. Em particular, o artigo de Daniel & Wilcoxon (1966) foi um dos primeiros trabalhos a discutir o efeito do sequenciamento de experimentos 2-fatoriais no número de mudanças de níveis dos fatores e na máxima contagem de tempo. Note que o uso de enumeração explícita para escolha da melhor sequência segundo algum desses critérios tem aplicação limitada, uma vez que o número de sequências cresce geometricamente com o número de fatores, impossibilitando a obtenção de soluções ótimas para *designs* de maior porte. Soluções para esse entrave são discutidas em Dickinson (1974) e Joiner & Campbell (1976), com propostas de algoritmos que buscam gerar sequências viáveis para *designs* com até 16 experimentos. Soluções de alta qualidade para *designs* maiores foram desenvolvidas efetivamente a partir deste século, com a evolução da tecnologia computacional.

Adelman (1972) apresenta uma revisão da literatura sobre o sequenciamento de experimentos fatoriais e fatoriais fracionados envolvendo questões de custo e robustez aos efeitos de tendências lineares. Apesar de os trabalhos de Draper & Stoneman (1968), Dickinson (1974) e Joiner & Campbell (1976) já considerarem efeitos de tendências lineares por meio da máxima contagem de tempo, somente no final do século XX esses critérios foram efetivamente retomados para a elaboração de *designs* fatoriais, em particular, nos artigos de Cheng & Jacroux (1988), Bailey et al. (1992) e Atkinson & Donev (1996). Coster & Cheng (1988), Jacroux (1994), Githinji & Jacroux (1998) e Tsao & Liu (2008), por sua vez, abordaram o problema tanto com base na contagem de tempo como no número de mudanças de níveis dos fatores.

Procedimentos para a construção de *designs* ótimos, bem como robustos a tendências lineares e de custo mínimo, foram propostos a partir dos anos 1990 por vários autores. Cheng (1985) desenvolveu o esquema generalizado de *foldover* para construir sequências de experimentos robustos aos efeitos de tendências lineares de *designs* fatoriais completos e fatoriais fracionados. Posteriormente, Cheng & Jacroux (1988) generalizaram o esquema, minimizando os vieses de estimativas dos efeitos principais e de interações duplas de experimentos 2^k . Tack & Vandebroek (2004) inovaram as pesquisas na área ao estudarem simultaneamente a robustez e custos de experimentos 2-fatoriais ortogonais e semiortogonais. Angelopoulos et al. (2009), por sua vez, propuseram um procedimento construtivo que produz *designs* 2-fatoriais ortogonais de mínimo custo com máxima D-eficiência e robustos a efeitos de tendências lineares. Mais recentemente, Hilow (2013) estendeu as análises com o estudo de quatro algoritmos para sequenciar *designs* 2^k e visando dois critérios: i) minimizar o número de mudanças de fatores; e ii) minimizar os

efeitos principais e interações de 2ª ordem aos efeitos de tendências lineares.

A análise da abordagem tradicional da aleatorização e da abordagem de sistematização do sequenciamento dos experimentos é formalmente endereçada em Adekeye & Kunert (2005). Os autores partem da premissa de que como na aleatorização os efeitos de tendências lineares são teoricamente diluídos no erro experimental (a presença de variáveis não controladas aumenta o erro aleatório), aparentemente é mais apropriado usar uma ordem sistemática de experimentação. Entretanto, os resultados do estudo a partir da aplicação de métodos de simulação, não mostraram vantagens da ordem sistemática em relação à aleatorizada.

Tal conclusão é, entretanto, questionada por outros autores que afirmam que a aleatorização não é necessariamente a melhor prática. Cheng & Jacroux (1988) apresentam provas matemáticas de que a aleatorização é inadequada para experimentos sujeitos a efeitos de tendências lineares e propuseram a construção de planejamentos robustos aos efeitos do tempo para obter covariância nula entre as variáveis e o tempo. Bertsimas et al. (2015) mostram também os benefícios da abordagem sistemática em experimentos com cobaias, aplicando técnicas de otimização matemática, e Ganju & Lucas (2004) estudaram o sequenciamento sistemático da ordem dos experimentos, indicando inadequação da aleatorização como prática a ser seguida em qualquer situação.

3 Abordagens do estudo

Nesta seção, são discutidas as três abordagens de geração de *designs* 2-fatoriais, cujas soluções foram analisadas em termos de custo e qualidade estatística.

A primeira abordagem (aqui denominada **RAN**) consiste da aplicação da técnica DETMAX (Mitchell, 1974) por meio do *software* comercial *Statistica*. A matriz do *design* é construída visando máxima eficiência, enquanto o sequenciamento de seus experimentos é feito de forma aleatória.

O segundo procedimento (denominado **AEK09**) é o algoritmo proposto por Angelopoulos et al. (2009), projetado para construir *designs* ortogonais e balanceados, livre de efeitos de tendências lineares nos efeitos principais, alta D-eficiência e número mínimo de mudanças de variáveis. Para tal, é inicialmente elencado o conjunto de colunas que descrevem todas as possíveis combinações de níveis que resultam em contagem de tempo nula, considerando-se N experimentos. As colunas são então divididas em $N-1$ conjuntos S_j , disjuntos com j mudanças de fatores, e organizados em ordem crescente de j . Ou seja, as colunas do conjunto S_1 têm uma mudança de fator, as colunas de S_2 têm duas mudanças de fatores, e assim por diante. Para um número de k fatores, constroem-se todas as matrizes ortogonais possíveis,

selecionando-se k colunas a partir dos primeiros k conjuntos não vazios em ordem crescente. Dentre as matrizes geradas, escolhe-se a mais eficiente e, se sua eficiência é igual ao valor máximo conhecido para a aplicação, o procedimento é finalizado, retornando-se para a matriz e sequenciamento dos experimentos. Caso contrário, a pesquisa é expandida para os primeiros $k + 1$ conjuntos. Note que a construção da matriz define implicitamente o sequenciamento, visando o mínimo número de troca de fatores.

Finalmente, a terceira abordagem (aqui denominada **POC**) consiste da aplicação do método de otimização linear inteira *branch and cut* (Cordier et al., 1999) incluso no *software* comercial GAMS/CPLEX ao modelo de programação inteira mista de Pureza et al. (2014). Tal modelo formaliza o problema de sequenciamento de N experimentos com 2 fatores. Especificamente, o objetivo do modelo é encontrar, para uma dada matriz de experimentos, uma sequência de execução com mínimo número de mudanças de níveis de fatores e mínima contagem de tempo. Pesos diferentes são dados a estas medidas, de modo que a minimização do número de mudanças de fator domina a minimização da contagem de tempo. O sequenciamento dos tratamentos é dado por variáveis binárias x_{ij} , em que i e j denotam experimentos distintos da matriz. A variável é igual a 1 se o experimento i precede o experimento j , caso contrário, assume o valor 0.

Por conseguinte, quando o modelo é resolvido, os valores das variáveis definem um caminho a partir do primeiro experimento para o último (ou seja, uma

ordem de execução). O modelo inclui uma restrição que calcula o número de mudanças dos níveis dos fatores, três famílias de restrições que contabilizam a contagem de tempo, duas famílias de restrições que impõem que apenas um experimento pode preceder e suceder cada experimento, e uma família de restrições que eliminam subciclos. No presente estudo, o modelo foi utilizado para sequenciar os *designs* gerados pela abordagem RAN.

4 Análise comparativa das abordagens sistemática e aleatória em seis exemplos

A Tabela 1 apresenta as soluções obtidas com as três abordagens para seis exemplos. Os números constantes na primeira coluna caracterizam os dados de entrada do exemplo, indicando o número de experimentos e o número de fatores desejados no *design*; por exemplo, 12.4 indica *designs* com 12 experimentos e 4 fatores. A segunda coluna apresenta a abordagem utilizada enquanto a terceira coluna fornece os experimentos dos *designs* em ordem de execução. Note que foi utilizada uma notação concisa para apresentação da sequência de experimentos; considerando que o primeiro fator é denotado pela letra “a”, o segundo fator é denotado pela letra “b”, e assim por diante, apenas os fatores no nível superior são apontados em cada experimento da sequência. Experimentos com todos os fatores no nível inferior são denotados pelo símbolo (1).

Tabela 1. Resultados das abordagens.

Exemplo	Abordagem	Experimentos em ordem de execução
12.4	RAN	d b abcd 1 a bd abc acd abd c cd bc
	AEK09	(1) abcd ac ad bcd bd ab bc c d ab acd
	POC	abc ac a bd b d cd c bcd ab abcd ad
16.5	RAN	d cde c ace abcde a bcd acd abe be ade b abd abc e bde
	AEK09	(1) ab abcd abce acde de ce cd bd bcde be ae abde ad ac bc
	POC	c b bde cde ade abe abc acd abcde ace a abd d bcd bce e
20.4	RAN	abcd cd bd bc a acd abd d ac ab (1) bcd abc ad b c abd bcd d ac
	AEK09	(1) (1) bd bcd bcd abc ac acd acd ad a abc abd abd ab b bc c cd d
	POC	ad acd cd bcd bc b ab a a ac c c (1) d d bd abd abcd abcd abc
24.4	RAN	d abcd cd c bd bc ac acd abcd abc d bcd abd a cd bc ac a b abd ab ad b (1)
	AEK09	(1) (1) a abd abd abcd acd acd cd c bc bc bcd bcd bd b ab ab abc ac ac ad d d
	POC	ab ab abc bc c c cd d d ad ad abd bd bcd bcd abcd abcd acd ac ac a (1) b b
28.4	RAN	ad bc ac acd abcd cd d d ab bcd c c abcd bc a d bcd (1) ab ac ad b (1) ab bd acd abd bd
	AEK09	(1) a ac cd c bc bcd bcd bd bd abd abd ad a ab ab abc abc abcd acd acd cd d (1) c bc b
	POC	d d ad ad ab ab b b bc bc abc abcd abcd bcd cd cd acd a a ac ac c c (1) bd bd abd abd cd de ce ac bc bcde ab ad abde (1) bcde be abde abce abcd ce bd bc acde ac (1) ab ae abcd ae ad de
28.5	AEK09	(1), ac,ab,ade,de,ac,cd,cde,bc,bcde,bd,abcde,abce,abcd,abd,abde,ae,ae, be, be,bce,ce,(1),cd,bd,ad,abc,acde
	POC	a ad abd ab b bc bce ce cde de ade abde abcde abcd abc ac c (1) e ae abe abce ace acd d bde bd bcd

As medidas relevantes dos *designs* obtidos (D-eficiência (coluna D_{eff}), contagem de tempo (coluna TC), número de mudanças de nível de cada fator (coluna FC), número total de mudanças de nível dos fatores (coluna NFC), correlação média (coluna $\bar{\rho}$) e correlação máxima (coluna ρ_{max}) são apresentadas na Tabela 2. Essas medidas foram calculadas pelas Equações 2-5, utilizando-se o *software Maple*.

Analisando-se a Tabela 2, verifica-se que a correlação máxima entre as colunas de interesse (efeitos principais e interações de 2ª ordem) e o tempo é menor ao se empregar a ordem aleatória de execução dos experimentos. Nos exemplos tratados, os desvios percentuais dessa medida em relação aos resultados das abordagens sistemáticas é cerca de -28%. Por outro lado, a correlação média das sequências aleatorizadas é, em média, 10% maior que a das sequências de AEK09 e POC. Outro aspecto desfavorável de RAN é que ela resultou em um número substancialmente maior de mudanças dos níveis dos fatores, com efeitos evidentes sobre os custos e o tempo de execução dos planejamentos experimentais. Em relação às abordagens sistemáticas, o número de mudanças de níveis representa um aumento médio de 115%.

No que diz respeito às abordagens sistemáticas, as sequências de AEK09 têm correlações médias e máximas menores que as observadas nas sequências de POC (em média, 27% e 11%, respectivamente), porém apresenta, em média, um número médio de mudanças de níveis de fatores 7% maior. Quanto ao critério D-eficiência, as sequências de POC são ligeiramente superiores às sequências de AEK09, o

que é justificado pelo fato de os *designs* não estarem limitados ao tipo ortogonal. As sequências produzidas por RAN e POC, por sua vez, apresentaram o mesmo valor de D-eficiência, o que é esperado uma vez que o conjunto de experimentos é o mesmo em ambos e o sequenciamento não interfere nessa medida.

5 Análise das sequências sistemáticas e aleatória em um caso real

A análise das correlações máximas para os seis exemplos da seção anterior indica algumas vantagens da aleatorização do sequenciamento sobre sua sistematização. Entretanto, a correlação não é um indicador definitivo, uma vez que não avalia os efeitos dos vieses provocados pelas tendências lineares para cada fator individualmente.

Uma maneira de avaliar esses efeitos é reproduzir a condução dos experimentos via Simulação. De fato, Gibbons & Chakraborti (2011) indicam a Simulação como um método eficiente para determinar erros tipo I (falsos positivos) e os erros tipo II (falsos negativos). O erro padrão da estimativa do erro tipo I (também conhecido como erro α) e do erro tipo II (erro β) são dados, respectivamente, por (Equações 6 e 7)

$$\sigma_{\alpha} = \sqrt{\frac{\alpha(1-\alpha)}{\eta}} \quad (6)$$

$$\sigma_{\beta} = \sqrt{\frac{\beta(1-\beta)}{\eta}} \quad (7)$$

Tabela 2. Medidas dos *designs* obtidos.

Exemplo	Abordagem	D_{eff} (%)	FC	NFC	$\bar{\rho}$	ρ_{max}
12.4	RAN	85,473	6 7 5 9	27	0,311	0,435
	AEK09	79,750	5 6 7 7	25	0,155	0,628
	POC	85,473	2 5 5 5	17	0,193	0,483
16.5	RAN	100	8 9 8 10 7	42	0,208	0,461
	AEK09	100	4 4 7 7 8	30	0,147	0,651
	POC	100	2 10 7 5 6	30	0,130	0,705
20.4	RAN	96,750	10 13 14 9	46	0,147	0,434
	AEK09	93,950	2 4 6 7	19	0,169	0,468
	POC	96,750	4 3 5 3	15	0,277	0,763
24.4	RAN	96,530	10 14 8 13	45	0,195	0,494
	AEK09	96,840	4 4 4 5	17	0,171	0,554
	POC	96,530	5 4 4 2	15	0,209	0,506
28.4	RAN	97,920	15 17 14 12	58	0,129	0,274
	AEK09	96,840	4 5 6 6	21	0,120	0,451
	POC	97,920	7 3 4 4	18	0,182	0,787
28.5	RAN	96,420	12 14 11 18 17	72	0,164	0,424
	AEK09	94,090	5 6 7 7 7	32	0,126	0,557
	POC	96,420	5 7 7 5 6	30	0,218	0,478

em que α é probabilidade de ocorrência de falsos positivos, β é a probabilidade de ocorrência de falsos negativos e η é o número de simulações.

Com este objetivo, foram realizadas simulações da condução dos *designs* com 16 tratamentos e 5 fatores (16.5) gerados por cada abordagem, tanto na condição estacionária como submetidos a tendências lineares. Esse exemplo foi selecionado por ter o mesmo número de experimentos e fatores de um *design* aplicado em um estudo de caso na indústria. Os parâmetros μ_i de cada tratamento i e σ do erro aleatório (erro experimental) considerados nas simulações foram obtidos, portanto, dos dados experimentais coletados no estudo de caso.

5.1 Descrição do procedimento de simulação

As simulações consideram um processo de fabricação de embalagens de vidro utilizadas na indústria alimentícia. Uma descrição sucinta desse processo indica quatro macroetapas: i) a etapa de fusão, cujas propriedades químicas do líquido fundido têm grande influência sobre a qualidade final do produto; ii) a etapa de conformação a quente, cujos elementos-chave são os componentes mecânicos e procedimentos operacionais; iii) a etapa de resfriamento dos produtos, cuja qualidade final depende do ciclo de resfriamento; e, finalmente, iv) a etapa de inspeção final de 100% dos produtos, cuja variável crítica é a instabilidade dos equipamentos de inspeção. Esta etapa é um fator que pode produzir efeitos de tendências lineares devido à perda de acuracidade do sistema de medição ao longo do tempo.

A fim de analisar os *designs* 16.5 das três abordagens, foi elaborada uma rotina em *Maple* 13. Para cada *design*, o estudo seguiu os seguintes passos: i) simulamos η vezes a execução do *design*, e estimamos os parâmetros do modelo da Equação 1, sendo os valores de cada experimento gerados segundo a distribuição normal com μ_i e σ ; ii) determinamos o intervalo de confiança $\hat{\mu} \pm 2\sigma$ dos parâmetros estatísticos do modelo; iii) caso

o intervalo de confiança não contivesse o valor zero, haveria evidência para o tomador de decisão afirmar que o parâmetro era estatisticamente significativo, caso contrário, nada se poderia afirmar. O erro α foi estimado de acordo com a frequência com que o evento ocorreu, podendo esta ser expressa por $\alpha = 1 - P[0 \in (-2\sigma; 2\sigma)]$.

O modelo simulado é aquele representado pela Equação 1, com termos referentes aos efeitos principais e às interações de 2ª ordem consideradas relevantes na ocasião do estudo de caso (ou seja, não foram consideradas todas as interações duplas possíveis do modelo da Equação 1). Foram simuladas $\eta = 1000$ execuções dos *designs* em condição estacionária, ou seja, sem efeitos de tendências lineares, e em condição dinâmica, isto é, com efeitos de tendências lineares oriundos de variáveis não controladas. Como a ocorrência de um ponto fora do intervalo de confiança $(-2\sigma; 2\sigma)$ segue a distribuição binomial, é possível estimar o intervalo de confiança de 95% para os erros α e β ($1,96\sigma_{\alpha,\beta}$), e assim inferir sobre o impacto dos efeitos de tendências lineares nos erros tipos I e II nos testes estatísticos para determinar a significância dos parâmetros do modelo da Equação 1. O procedimento completo de simulação encontra-se esquematizado na Figura 1.

Cinco fatores (denotados por A, B, C, D, E) relacionados ao processo de fabricação foram selecionados: i) parâmetros do processo de fusão; ii) lubrificação dos moldes de fusão; iii) características das matérias-primas utilizadas na fusão; iv) parâmetros do processo de conformação; v) ciclo de vida dos equipamentos da etapa de conformação. A variável de resposta é o rendimento do processo (número de garrafas sem defeitos), dado em porcentagem.

Os parâmetros referentes aos efeitos principais são identificados pela ordem relativa de seus respectivos fatores, ou seja, como β_1 (fator A), β_2 (fator B), ..., β_5 (fator E), enquanto, como já discutido, β_0 representa a média global. Os parâmetros das interações de 2ª ordem ($\beta_{kk'}$) são identificados pela combinação das ordens dos fatores envolvidos; por exemplo,

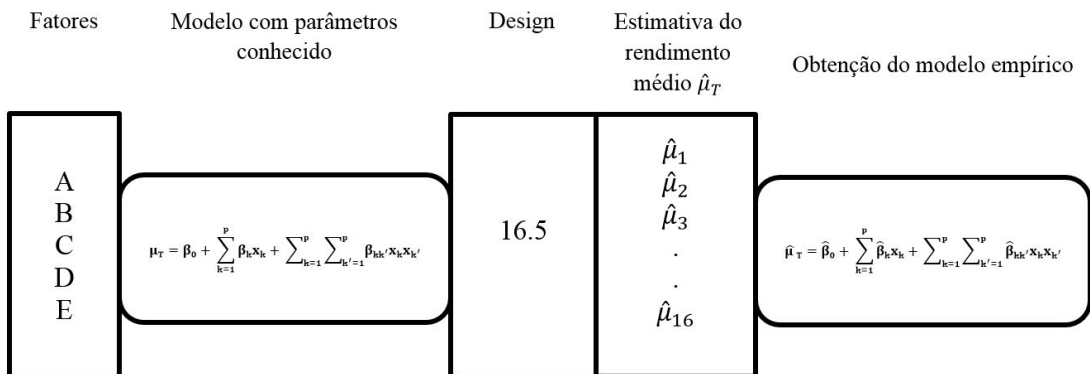


Figura 1. Síntese do procedimento de simulação.

12 representa a interação entre os fatores A e B, sendo β_{12} o respectivo parâmetro.

Os valores dos parâmetros do modelo utilizados na simulação são:

$$\begin{aligned} \beta_0 = 70; \beta_1 = \beta_0 = 70; \beta_1 = -2,79; \beta_2 = -2,27; \beta_3 = -0,17; \\ \beta_4 = 0; \beta_5 = -3,80; \beta_{12} = 1,74; \beta_{13} = 1,0; \beta_{14} = 0; \\ \beta_{15} = 2,41; \beta_{23} = 0,5; \beta_{24} = 0; \beta_{25} = 0; \beta_{34} = 0; \beta_{35} = 0 \\ \text{e } \beta_{45} = 0. \end{aligned}$$

Esses parâmetros foram utilizados para gerar os valores estimados da resposta em cada experimento, sendo $\sigma_e = 0,2145$ o desvio padrão do erro experimental adotado nas simulações. O efeito de tendência linear foi de 1% acumulado ao longo dos experimentos; ou seja, o primeiro experimento executado teve 1% de viés na média populacional ($0,01\mu_i$), o segundo experimento executado teve 2% de viés ($0,02\mu_i$), e assim por diante, até o décimo sexto experimento com um viés de 16% ($0,16\mu_i$). Esse viés é razoável para o caso estudado, uma vez que a instabilidade dos sistemas de medição acarreta um efeito aproximadamente linear.

5.2 Análise das simulações

A Tabela 3 mostra os resultados das simulações para cada tratamento do *design* 16.5 resultante da abordagem RAN (sequenciamento aleatório) sem efeitos de tendência linear (Normal) e com efeito de tendências lineares (TL), assim como o número de ocorrências em que o efeito da coluna é detectado como estatisticamente significativo (coluna Detecção de significância estatística). O fator A foi detectado como estatisticamente significativo nos 1000 testes de significância realizados, tanto para a condição Normal quanto para a condição TL, resultado este esperado, uma vez que $\beta_1 = -2,79$ e $\hat{\beta}_1 = -2,79605$ sem efeitos de tendências lineares e $\hat{\beta}_1 = -2,72375$ com efeitos de tendências lineares. Para o fator D (em que $\beta_4 = 0$),

obtiveram-se 69 testes estatisticamente significativos (erro $\alpha = \frac{69}{1000} = 0,069$) para a condição Normal. Para a situação TL, foram obtidos 427 falsos positivos ($\alpha = 42,7\%$). Esse resultado mostra, portanto, o impacto dos efeitos de tendências lineares nos testes de significância estatística para a sequência aleatorizada.

O viés da estimativa é mostrado na coluna Dif da Tabela 3, indicando que a ocorrência de efeitos de TL aumenta o erro tipo I (α), o qual postula que uma variável é estatisticamente significativa quando ela não é. Isso é percebido, por exemplo, pelos resultados da variável D e da interação AD. Para esta última, o erro de falso positivo é de 0,315; sob a condição Normal, esse erro está abaixo de 0,07.

A Tabela 4 mostra a contagem de tempo e a correlação para cada uma das colunas de efeitos principais e interações de 2ª ordem (aquelas consideradas relevantes no estudo do caso real) do *design* 16.5 produzido por RAN. Tal análise é importante, pois indica a coluna com maior viés quando os experimentos estão sujeitos a efeitos de tendência linear. Assim, observa-se que a maior correlação sob esse tipo de efeito ocorre na coluna B, seguido da coluna D. Esse resultado é significativo para o planejamento de experimentos industriais; desconsiderá-lo pode levar a conclusões equivocadas sobre o processo, com efeitos na qualidade e produtividade.

Análise similar foi feita para a sequência da abordagem AEK09 (Tabelas 5 e 6). Nota-se na Tabela 6 que as colunas A, B, C, D, E e AE apresentam contagem de tempo igual a zero, o que significa correlação com o tempo nula (fatores das colunas ortogonais ao tempo). A robustez das colunas A-E deve-se à própria natureza de AEK04, uma vez que a abordagem foi projetada para construir exclusivamente *designs* com efeitos principais livres de tendências lineares. Essa propriedade não foi observada nas colunas do *design* produzido por RAN. E como identificado

Tabela 3. Estimativa dos efeitos principais e das interações de 2ª ordem para o *design* 16.5 produzido pela abordagem RAN.

Fonte de variação	Estimado		Dif	Detecção de significância estatística	
	Normal	TL		Normal	TL
A	-2,79605	-2,72375	-0,0723	1000	1000
B	-2,27672	-2,48656	0,209839	1000	1000
C	-0,16492	-0,10071	-0,06421	384	142
D	-0,00299	-0,18243	0,179438	69	427
E	-3,79776	-3,97455	0,176787	1000	1000
AB	1,734983	1,719464	0,015519	1000	1000
AC	1,004719	0,988173	0,016546	1000	1000
AD	0,002595	0,153406	-0,15081	63	315
AE	2,411162	2,535522	-0,12436	1000	1000
BC	0,501172	0,407948	0,093224	992	964
BD	0,0045	-0,02836	0,032861	65	59
BE	0,00219	-0,15005	0,152237	62	78
	Média		0,03873		

nas simulações, os *designs* de AEK09 têm menores vieses nas estimativas dos parâmetros do modelo da Equação 1 (observe e compare a coluna Dif das Tabelas 3 e 5); consequentemente, os erros α e β são menores para a abordagem sistemática AEK09 que os *designs* de RAN quando na presença de tendências lineares (por exemplo, observe na Tabela 3, que, quando na presença de efeitos de tendências lineares, o falso positivo, α , da variável D é de 42,7%, contra 6,9% em condições normais).

Os resultados das análises da abordagem POC são mostrados nas Tabelas 7 e 8. Na Tabela 8, observa-se robustez nas colunas A, B, C, D, E, AB, AC e AD. Em relação à sequência de AEK09, o viés médio é menor (de 0,020156 para 0,00499), e quando compramos os *designs* com abordagem sistemática com o aleatório, observamos um número menor de falsos positivos para o *design* POC e AEK09 (observe as Tabelas 3, 5 e 7 na coluna detecção de significância estatística). Esses resultados motivam a

pesquisa de métodos sistemáticos que gerem *designs* de experimentos de menor custo e maior robustez aos efeitos de tendências lineares.

Constata-se, na Figura 2, que, para o exemplo 16.5, o *design* produzido pela abordagem RAN tem maior correlação média e menor correlação máxima que os *designs* de AEK09 e POC. Também se observa que erros de estimativa (ϵ), número de mudanças de fatores e contagem de tempo são menores para as abordagens sistemáticas. Esses resultados implicaram menores erros dos tipos I e II e menor correlação com o tempo, ou seja, maior robustez aos efeitos de tendências lineares para AEK09 e POC. As diferenças em termos dos erros tipo I e II entre as três abordagens são mostradas nas Tabelas 9 e 10.

A Tabela 9 apresenta para cada abordagem, a estimativa por intervalo de confiança de 95% (IC 95%) do erro tipo I e a probabilidade de se tomar uma decisão correta ($1-\alpha$) quando o fator produz efeito na variável de resposta. Observe que não há interseção

Tabela 4. Contagem de tempo e correlação das colunas para o *design* 16.5 produzido pela abordagem RAN.

Coluna	TC	ρ
A	0	0,136
B	34	0,461
C	12	0,163
D	30	0,407
E	6	0,081
AB	4	0,054
AC	2	0,027
AD	24	0,325
AE	20	0,271
BC	14	0,190
BD	4	0,054
BE	24	0,325
Média ($\bar{\rho}$)		0,208

Tabela 5. Estimativa dos efeitos principais e de interações de 2ª ordem para o *design* 16.5 produzido pela abordagem AEK09.

Fonte de variação	Estimado		Dif	Detecção de significância estatística	
	Normal	TL		Normal	TL
A	-2,79605	-2,76683	-0,02922	1000	1000
B	-2,27672	-2,17525	-0,10147	1000	1000
C	-0,16492	-0,16587	0,000952	384	355
D	-0,00299	-0,00769	0,004698	69	69
E	-3,79776	-3,76686	-0,0309	1000	1000
AB	1,734983	1,767742	-0,03276	1000	1000
AC	1,004719	1,001869	0,00285	1000	1000
AD	0,002595	-0,00023	0,00282	63	72
AE	2,411162	2,381798	0,029364	1000	1000
BC	0,501172	0,502724	-0,00155	992	994
BD	0,0045	0,001125	0,003375	65	65
BE	0,00219	0,004117	-0,00193	62	63
Média			0,020156		

Tabela 6. Contagem de tempo e correlação para o *design* 16.5 produzido pela abordagem AEK09.

Coluna	TC	ρ
A	0	0,000
B	0	0,000
C	0	0,000
D	0	0,000
E	0	0,000
AB	48	0,651
AC	28	0,380
AD	4	0,054
AE	0	0,000
BC	4	0,054
BD	4	0,054
BE	16	0,217
Média ($\bar{\rho}$)		0,118

Tabela 7. Estimativa dos efeitos principais e de interações de 2ª ordem para o *design* 16.5 produzido pela abordagem POC.

Fonte de variação	Estimado		Dif	Detecção de significância estatística	
	Normal	TL		Normal	TL
A	-2,7879	-2,7888	0,0009	1000	1000
B	-2,2779	-2,268	-0,0099	1000	1000
C	-0,1679	-0,16	-0,0079	329	353
D	-0,0021	0,0021	-0,0042	82	67
E	-3,7992	-3,8	0,0008	1000	1000
AB	1,7371	1,7395	-0,0024	1000	1000
AC	1,0004	0,9999	0,0005	1000	1000
AD	0,0079	0,0011	0,0068	56	66
AE	2,4135	2,3907	0,0228	1000	1000
BC	0,4988	0,5523	-0,0535	995	1000
BD	0,0037	0,0214	-0,0177	58	68
BE	0,0003	-0,0036	0,0039	53	52
Média			-0,00499167		

Tabela 8. Contagem de tempo e correlação das colunas para o *design* 16.5 produzido pela abordagem POC.

Colunas	TC	ρ
A	0	0,000
B	0	0,000
C	0	0,000
D	0	0,000
E	0	0,000
AB	0	0,000
AC	0	0,000
AD	0	0,000
AE	16	0,217
BC	44	0,597
BD	16	0,217
BE	4	0,054
Média ($\bar{\rho}$)		0,090

entre os intervalos de confiança nas condições Normal e TL para o fator D e interação AD para a abordagem RAN (por exemplo, o limite superior da variável D na condição normal, $\alpha=8,47\%$, é menor que o limite inferior na condição TL, $\alpha=39,63\%$). Para AEK09, o

intervalo para a condição normal da variável D [5,33%; 8,47%], na condição TL o intervalo de confiança é igual [5,33%; 8,47%]; para as interações AD, BD e BE são, para a condição normal, respectivamente: [4,49%;7,81%], [4,97%;8,03%] e [4,71%;7,69%].

Tabela 9. IC 95% da estimativa de erros tipo I (α) para os designs do exemplo 16.5 produzidos pelas três abordagens.

Condição	Coluna	RAN				AEK09				POC			
		1- α		α		1- α		α		1- α		α	
		Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL
Média	D	93,10%	57,30%	6,90%	42,70%	61,60%	64,50%	6,90%	6,90%	92,80%	93,30%	8,20%	6,70%
	AD	93,70%	68,50%	6,30%	31,50%	93,70%	92,80%	6,30%	7,20%	94,40%	93,40%	5,60%	6,60%
	BD	93,50%	94,10%	6,50%	5,90%	93,50%	93,50%	6,50%	6,50%	94,20%	93,20%	5,80%	6,80%
	BE	93,80%	92,20%	6,20%	7,80%	93,80%	93,70%	6,20%	6,30%	94,70%	94,80%	5,30%	5,20%
Desvio padrão	D	0,80%	1,56%	0,80%	1,56%	1,54%	1,51%	0,80%	0,80%	0,82%	0,79%	0,87%	0,79%
	AD	0,77%	1,47%	0,77%	1,47%	0,77%	0,82%	0,77%	0,82%	0,73%	0,79%	0,73%	0,79%
	BD	0,78%	0,75%	0,78%	0,75%	0,78%	0,78%	0,78%	0,78%	0,74%	0,80%	0,74%	0,80%
	BE	0,76%	0,85%	0,76%	0,85%	0,76%	0,77%	0,76%	0,77%	0,71%	0,70%	0,71%	0,70%
Limite Superior	D	94,67%	60,37%	8,47%	45,77%	64,61%	67,47%	8,47%	8,47%	94,40%	94,85%	9,90%	8,25%
	AD	95,21%	71,38%	7,81%	34,38%	95,21%	94,40%	7,81%	8,80%	95,83%	94,94%	7,03%	8,14%
	BD	95,03%	95,56%	8,03%	7,36%	95,03%	95,03%	8,03%	8,03%	95,65%	94,76%	7,25%	8,36%
	BE	95,29%	93,86%	7,69%	9,46%	95,29%	95,21%	7,69%	7,81%	96,09%	96,18%	6,69%	6,58%
Limite Inferior	D	91,53%	54,23%	5,33%	39,63%	58,59%	61,53%	5,33%	5,33%	91,20%	91,75%	6,50%	5,15%
	AD	92,19%	65,62%	4,79%	28,62%	92,19%	91,20%	4,79%	5,60%	92,97%	91,86%	4,17%	5,06%
	BD	91,97%	92,64%	4,97%	4,44%	91,97%	91,97%	4,97%	4,97%	92,75%	91,64%	4,35%	5,24%
	BE	92,31%	90,54%	4,71%	6,14%	92,31%	92,19%	4,71%	4,79%	93,31%	93,42%	3,91%	3,82%

Tabela 10. IC 95% da estimativa de erros tipo II (β) para os designs 16.5 produzidos pelas três abordagens.

Condição	Fator	RAN						AEK09						POC											
		1- β		B		1- β		β		1- β		TL		1- β		TL		Normal		TL		Normal		TL	
		Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL	Normal	TL
Média	A	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
	B	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
	C	38,40%	14,20%	61,60%	85,80%	38,40%	35,50%	61,60%	64,50%	32,90%	35,30%	67,10%	64,70%	35,30%	32,90%	67,10%	64,70%	35,30%	32,90%	67,10%	64,70%	35,30%	32,90%	67,10%	64,70%
Desvio Padrão	E	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
	AB	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
	AC	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
Limite Superior	AE	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
	BC	99,20%	96,40%	0,80%	3,60%	99,20%	99,40%	0,80%	0,60%	99,50%	100,00%	0,50%	0,00%	99,50%	100,00%	0,50%	0,00%	99,50%	100,00%	0,50%	0,00%	99,50%	100,00%	0,50%	0,00%
	A	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Limite Inferior	B	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	C	1,54%	1,10%	1,54%	1,10%	1,54%	1,10%	1,54%	1,51%	1,49%	1,51%	1,49%	1,51%	1,49%	1,51%	1,49%	1,51%	1,49%	1,51%	1,49%	1,51%	1,49%	1,51%	1,49%	1,51%
	E	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Limite Superior	AB	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	AC	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	AE	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Limite Inferior	BC	99,75%	97,55%	1,35%	4,75%	99,75%	99,88%	1,35%	1,08%	99,94%	100,00%	0,94%	0,00%	99,94%	100,00%	0,94%	0,00%	99,94%	100,00%	0,94%	0,00%	99,94%	100,00%	0,94%	0,00%
	A	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
	B	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
Limite Superior	C	35,39%	12,04%	58,59%	83,64%	35,39%	32,53%	58,59%	61,53%	29,99%	32,34%	64,19%	61,74%	32,34%	29,99%	64,19%	61,74%	32,34%	29,99%	64,19%	61,74%	32,34%	29,99%	64,19%	61,74%
	E	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
	AB	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
Limite Inferior	AC	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
	AE	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%	100,00%	100,00%	0,00%	0,00%
	BC	98,65%	95,25%	0,25%	2,45%	98,65%	98,92%	0,25%	0,12%	99,06%	100,00%	0,06%	0,00%	99,06%	100,00%	0,06%	0,00%	99,06%	100,00%	0,06%	0,00%	99,06%	100,00%	0,06%	0,00%

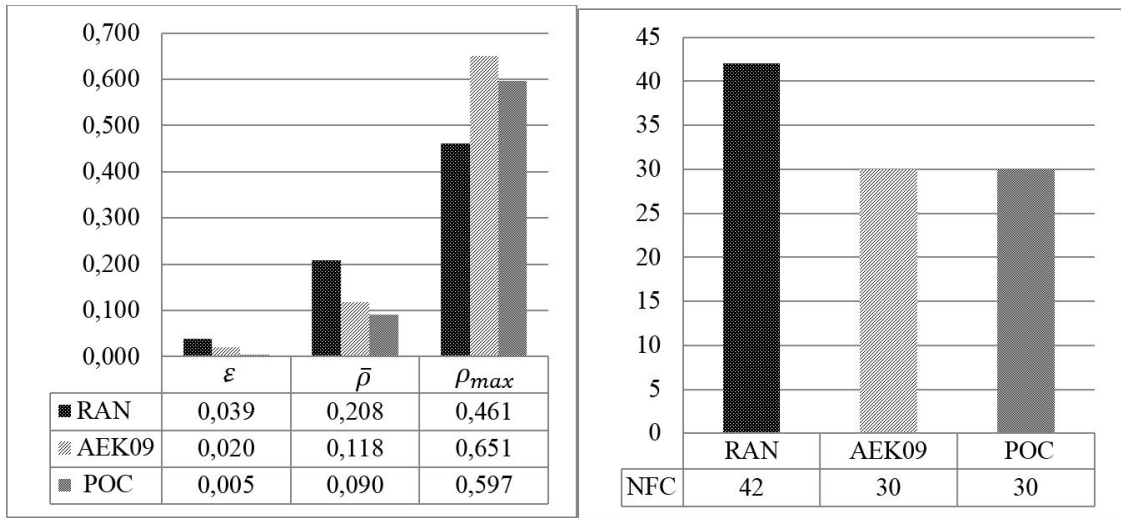


Figura 2. Erros de estimativa, correlação média, correlação máxima e número de mudanças de fatores dos *designs* das abordagens para o exemplo 16.5.

Para a condição TL, considerando ainda a abordagem AEK09, os intervalos de confiança são os seguintes: [5,60%;8,88%], [4,97%;8,03%] e [4,79%;7,81%]. Quando comparamos os IC dessas variáveis entre as condições TL e Normal, há interseção entre os intervalos, indicando não haver diferença estatisticamente significativa entre as duas condições. O mesmo resultado é encontrado na abordagem POC, o que indica maior robustez das abordagens sistemáticas diante da abordagem aleatória na presença de efeitos de tendências lineares. Portanto, a abordagem sistemática tem menor falso positivo que a abordagem clássica de aleatorização das sequências de experimentos quando há presença de efeitos de tendências lineares.

A Tabela 10 mostra os resultados da estimativa por intervalo de confiança de 95% (IC 95%) dos erros tipo II e do poder do teste dado por $(1-\beta)$. Constatamos, nessa análise, que, para o efeito principal C e a interação BC, o erro tipo II é maior quando o *design* está sujeito a efeitos de tendência lineares e a ordem de execução é aleatorizada (procedimento RAN). O mesmo não foi observado para os procedimentos AEK09 e POC. Evidências estatísticas são obtidas quando se comparam os intervalos de confiança. Observa-se que, para a abordagem AEK09 na condição TL, o IC da variável C é [61,53%;67,47%], para a condição Normal, é [58,59%; 64,61%]; há, portanto, interseção entre as duas condições, o que não ocorre para a abordagem aleatória. Há assim fortes evidências estatísticas de que as duas abordagens sistemáticas geram menor erro tipo II, e melhor poder do teste quando comparadas à abordagem aleatória.

Com essas análises, temos evidência estatística de que a ordem sistemática de execução dos experimentos pode apresentar vantagens em relação à aleatorização. No caso do exemplo 16.5, as vantagens observadas

recaem em propriedades estatísticas em termos dos erros tipo I e II, bem como em relação ao número de troca de fatores dos experimentos. Observamos também que, neste exemplo, a abordagem POC apresenta melhor desempenho em termos de número de mudanças de fatores e correlações média e máxima em relação à abordagem AEK09, enquanto AEK09 garante *designs* robustos nos efeitos principais.

6 Conclusões e perspectivas de pesquisa futura

Os livros clássicos em DoE recomendam que a ordem de execução dos experimentos seja aleatorizada a fim de minimizar possíveis efeitos de tendências lineares. Entretanto, já na década de 1960, Daniel & Wilcoxon (1966) e Draper & Stoneman (1968) questionavam essa prática. Mais recentemente, vários autores vêm expondo o problema da inadequação da aleatorização dos experimentos de modo mais enfático e propõem algoritmos para a geração de *designs* segundo diferentes critérios.

Neste trabalho, procuramos contribuir para esse debate, considerando duas abordagens de geração de *designs* em que o sequenciamento dos experimentos é feito de forma sistemática e uma abordagem que aleatoriza o sequenciamento. Comparamos as três propostas com base em seis exemplos de *design* fatorial de dois níveis e verificamos vantagens das abordagens sistemáticas sobre a aleatorização na maioria dos critérios considerados. Em particular, provamos estatisticamente, por meio da simulação de um caso real, que a aleatorização aumenta os erros tipo I e II, o que reduz o poder do experimento em detectar fatores importantes do processo e fazer afirmações corretas sobre a significância dos fatores.

Como perspectiva de pesquisa futura, os bons resultados com o modelo de programação matemática para o sequenciamento de experimentos motivam sua extensão com vistas a incluir a decisão do conjunto de experimentos para compor a matriz. Tal extensão deve considerar simultaneamente critérios como D-eficiência, contagem de tempo e número de mudanças de fatores na escolha dos experimentos. Dada a maior complexidade das decisões envolvidas, um grande desafio será o de desenvolver uma formulação bem resolvida por métodos de otimização.

Referências

- Addelman, S. (1972). Recent development in the designs of factorial experiments. *Journal of the American Statistical Association*, 67(337), 103-111. <http://dx.doi.org/10.1080/001621459.1972.10481211>.
- Adekeye, K. S., & Kunert, J. (2005). On the comparison of run orders of unreplicated 2k-p-design in the process of a time-trend. *Technical Report, Universitat Dortmund*, (3), 475.
- Aggarwal, M. L., Budhraj, V., & Lin, D. K. J. (2003). New class of orthogonal arrays and its applications. *IAPQR Transactions*, 28(1), 23-32.
- Alonso, M. C., Bousbaine, A., Llovet, J., & Malpica, J. A. (2011). Obtaining industrial experimental designs using a heuristic technique. *Expert Systems with Applications*, 38(8), 10094-10098. <http://dx.doi.org/10.1016/j.eswa.2011.02.004>.
- Angelopoulos, P., Evangelaras, H., & Koukouvinos, C. (2009). Run orders for efficient two level experimental plans with minimum factor level changes robust to time trends. *Journal of Statistical Planning and Inference*, 139(10), 3718-3724. <http://dx.doi.org/10.1016/j.jspi.2009.05.002>.
- Atkinson, A. C. (1996). The usefulness of optimum experimental designs. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1), 59-76.
- Atkinson, A. C. (1996). The usefulness of optimum experimental designs. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1), 59-76.
- Atkinson, A. C., & Donev, A. N. (1996). Experimental designs optimally balanced for trend. *Technometrics*, 38(4), 333-341. <http://dx.doi.org/10.1080/00401706.1996.10484545>.
- Atkinson, A. C., Donev, A. N., & Tobias, R. D. (2007). *Optimum experiments design with SAS*. New York: Oxford Press.
- Atkinson, A. C., & Bailey, R. A. (2001). On hundred year of the design of experiments on and off the pages of *Biometrika*. *Biometrika*, 88(1), 53-97. <http://dx.doi.org/10.1093/biomet/88.1.53>.
- Bailey, R. A., Cheng, C. S., & Kipnis, P. (1992). Construction of trend-resistant factorial designs. *Statistica Sinica*, 2, 393-411.
- Bertsimas, D., Johnson, M., & Kallus, N. (2015). The power of optimization over randomization in designing experiments involving small samples. *Operations Research*, 63(4), 868-876. <http://dx.doi.org/10.1287/opre.2015.1361>.
- Bessant, J., Caffyn, S., & Gallagher, M. (2001). An evolutionary model of continuous improvement behavior. *Technovation*, 21(2), 67-77. [http://dx.doi.org/10.1016/S0166-4972\(00\)00023-7](http://dx.doi.org/10.1016/S0166-4972(00)00023-7).
- Bessant, J., & Caffyn, S. (1997). High involvement innovation through continuous improvement. *International Journal of Technology Management*, 14(3), 7-28. <http://dx.doi.org/10.1504/IJTM.1997.001705>.
- Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters*. New York: Wiley.
- Cheng, C.-S. (1985). Run orders of factorial designs. In L. LeCam & R. A. Olshen (Eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (pp. 619-633). Wadsworth.
- Cheng, C.-S. (1990). Construction of run orders of factorial designs. In S. Ghosh (Ed.), *Statistical design and analysis of industrial experiments* (pp. 423-39). New York: Subir Ghosh.
- Cheng, C.-S., & Jacroux, M. (1988). On the construction of trend-free run orders of two level factorial designs. *Journal of the American Statistical Association*, 83(404), 1152-1158. <http://dx.doi.org/10.1080/001621459.1988.10478713>.
- Cook, R. D., & Nachtsheim, C. J. A. (1980). Comparison of algorithms for constructing exact D-optimal designs. *Technometrics*, 22(3), 315-324. <http://dx.doi.org/10.1080/00401706.1980.10486162>.
- Cordier, C., Marchand, H., Laundry, R., & Wolsey, L. A. (1999). bc-opt: a branch-and-cut code for mixed integer programs. *Mathematical Programming*, 86(2), 335-353. <http://dx.doi.org/10.1007/s101070050092>.
- Coster, D. C., & Cheng, C.-S. (1988). Minimum cost trend-free run orders of fractional factorial designs. *Annals of Statistics*, 16(3), 1188-1205. <http://dx.doi.org/10.1214/aos/1176350955>.
- Daniel, C., & Wilcoxon, F. (1966). Factorial 2p-q plans robust against linear and quadratic trends. *Technometrics*, 8, 259-278.
- Davis, O. L. (1956). *The design and analysis of industrial experiments*. London: Longman. 637 p.
- Delbridge, R., & Barton, H. (2002). Organizing for continuous improvement: structures and roles in automotive components plants. *International Journal of Operations & Production Management*, 22(6), 680-692. <http://dx.doi.org/10.1108/01443570210427686>.
- Dickinson, A. W. (1974). Some run orders requirements a minimum number of factor level changes for the 24 and 25 main effect plans. *Technometrics*, 16, 31-37.
- Draper, N. R., & Stoneman, D. M. (1968). Factor changes and linear trends in eight-run two level factorial designs.

- Technometrics*, 10(2), 301-311. <http://dx.doi.org/10.1080/00401706.1968.10490562>.
- Dykstra, O. (1971). The Augmentation of experimental data to maximize $|X'X|$. *Technometrics*, 13(3), 682-688.
- Fisher, R. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*, 33, 503-513.
- Galil, Z., & Kiefer, J. (1980). Time- and space-saving computer methods related to Mitchell's DETMAX for finding D-Optimum designs. *Technometrics*, 22(3), 301-313. <http://dx.doi.org/10.1080/00401706.1980.10486161>.
- Ganju, J., & Lucas, J. M. (2004). Randomized and random run order experiments. *Journal of Statistical Planning and Inference*, 133(1), 199-210. <http://dx.doi.org/10.1016/j.jspi.2004.03.009>.
- Garroi, J. J., Goos, P., & Sorensen, K. (2009). A variable-neighborhood search algorithm for finding optimal run orders in the presence of serial correlation. *Journal of Statistical Planning and Inference*, 139(1), 30-44. <http://dx.doi.org/10.1016/j.jspi.2008.05.014>.
- Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference* (5th ed.). New York: Taylor & Francis.
- Githinji, F., & Jacroux, M. (1998). On the determination and construction of optimal designs for comparing a set of test treatments with a set of controls in the presence of a linear trend. *Journal of Statistical Planning and Inference*, 66(1), 61-74. [http://dx.doi.org/10.1016/S0378-3758\(97\)00067-0](http://dx.doi.org/10.1016/S0378-3758(97)00067-0).
- Harrison, A. (2000). Continuous improvement: the trade-off between self-management and discipline. *Integrated Manufacturing Systems*, 11(3), 180-187. <http://dx.doi.org/10.1108/09576060010320416>.
- Hilow, H. (2013). Comparison among run order algorithms for sequential factorial experiments. *Computational Statistics & Data Analysis*, 58, 397-406. <http://dx.doi.org/10.1016/j.csda.2012.09.013>.
- Hyland, P. W., Soosay, C., & Sloan, T. R. (2003). Continuous improvement and learning in the supply chain. *International Journal of Physical Distribution & Logistics Management*, 33(4), 316-335. <http://dx.doi.org/10.1108/09600030310478793>.
- Imai, M. (1997). *Gemba Kaisen: a common sense, low-cost approach to management*. New York: McGraw-Hill.
- Jacroux, M. (1994). On the construction of trend-resistant fractional factorial row-column designs. *The Indian Journal of Statistics*, 56(Pt. 2), 251-258.
- Joiner, B. L., & Campbell, C. (1976). Designing experiments when run order is important. *Technometrics*, 18(3), 249-259. <http://dx.doi.org/10.1080/00401706.1976.10489445>.
- Kume, H. (1993). *Métodos estatísticos para melhoria da qualidade* (11. ed.). São Paulo: Gente. 245 p.
- Marin-Garcia, J. A., Val, M. P., & Martin, T. B. (2008). Longitudinal study of the results of continuous improvement in an industrial company. *Team Performance Management*, 14(1/2), 56-6.
- Mitchell, T. J. (1974). An algorithm for the construction of D-optimal experimental designs. *Technometrics*, 16, 203-211.
- Montgomery, C. D., Runger, G. C., & Hubele, N. F. (2009). *Engineering statistics*. New York: John Wiley & Sons.
- Montgomery, D. C. (1991). *Design and analysis of experiments*. New York: John Wiley & Sons.
- Oprime, P. C., Monsanto, R., & Donadone, J. C. (2010). Análise da complexidade, estratégias e aprendizagem em projetos de melhoria contínua: estudos de caso em empresas brasileiras. *Gestão & Produção*, 17, 669-682.
- Pureza, V., Oprime, P. C., & Costa, A. F. (2014). *Some experiments on mathematical programming for experiment sequencing* (Documento de pesquisa).
- Savolainen, T. I. (1999). Cycles of continuous improvement: realizing competitive advantages through quality. *International Journal of Operations & Production Management*, 19(11), 1203-1222. <http://dx.doi.org/10.1108/01443579910291096>.
- Street, D. J., & Burgess, L. (2008). Some open combinatorial problems in the design of stated choice experiments. *Discrete Mathematics*, 308(13), 2781-2788. <http://dx.doi.org/10.1016/j.disc.2006.06.042>.
- Suen, C., & Midha, A. C. K. (2013). Optimal fractional factorial designs and their construction. *Journal of Statistical Planning and Inference*, 143(10), 1828-1834. <http://dx.doi.org/10.1016/j.jspi.2013.05.004>.
- Tack, L., & Vandebroek, M. (2004). Trend-resistant and cost-efficient cross-over designs for mixed models. *Computation Statistics & Data Analysis*, (46), 721-746.
- Toledo, J. C. (1986). *Qualidade Industrial: conceitos, sistemas e estratégias*. São Paulo: Editora Atlas.
- Triefenbach, F. (2008). *Design of experiments: the D-Optimal approach and its implementation as a computer algorithm* (Bachelor's thesis). UMEA University, Umea; South Westphalia University of Applied Sciences, Meschede.
- Tsao, H.-S. J., & Liu, H. (2008). Optimal sequencing of test conditions in 2^k factorial experimental design for run-size minimization. *Computers & Industrial Engineering*, 55(2), 450-464. <http://dx.doi.org/10.1016/j.cie.2008.01.006>.
- Wang, P. C. (1991). Symbol changes and trend resistance in orthogonal plans of symmetric factorials. *The Indian Journal of Statistics*, 53(Pt. 3), 297-303.
- Wang, P. C., & Chen, M. H. (1998). Level changes and trend resistance on replacement in asymmetric orthogonal arrays. *Journal of Statistical Planning and Inference*, 69(2), 349-358. [http://dx.doi.org/10.1016/S0378-3758\(97\)00168-7](http://dx.doi.org/10.1016/S0378-3758(97)00168-7).
- Wang, P. C., & Jan, H. W. (1995). Designing two-level factorial experiments using orthogonal arrays when the run order is important. *The Statistician*, 44(3), 379-388. <http://dx.doi.org/10.2307/2348709>.
- Wilmut, M., & Zhou, J. (2011). D-optimal minimax design criterion for two-level fractional factorial designs. *Journal of Statistical Planning and Inference*, 141(1), 576-587. <http://dx.doi.org/10.1016/j.jspi.2010.07.002>.