



**ANDRÉ BARBOSA VENTURA DA SILVA**

**ERRO TIPO I E PODER DO TESTE F EM  
DELINEAMENTOS DO TIPO SUDOKU**

**LAVRAS - MG**

**2012**

**ANDRÉ BARBOSA VENTURA DA SILVA**

**ERRO TIPO I E PODER DO TESTE F EM DELINEAMENTOS DO TIPO  
SUDOKU**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de “Mestre”.

Dr. Júlio Sílvio de Sousa Bueno Filho  
Orientador

**LAVRAS - MG  
2012**

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da  
Biblioteca da UFLA**

Silva, André Barbosa Ventura da.

Erro tipo I e poder do teste F em delineamentos do tipo Sudoku /  
André Barbosa Ventura da Silva. - Lavras : UFLA, 2012.

94 p. : il.

Dissertação (mestrado) - Universidade Federal de Lavras, 2012.

Orientador: Júlio Sílvio de Sousa Bueno Filho.

Bibliografia.

1. Delineamento Experimental. 2. Quadrado Latino. 3. Monte  
Carlo. 4. Simulação. I. Universidade Federal de Lavras. II. Título.

CDD - 519.57

**ANDRÉ BARBOSA VENTURA DA SILVA**

**ERRO TIPO I E PODER DO TESTE F EM DELINEAMENTOS DO TIPO  
SUDOKU**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de “Mestre”.

APROVADA em 29 de fevereiro de 2012.

Dr.<sup>a</sup>. Ana Carla Marques Pinheiro

UFLA

Dr. Daniel Furtado Ferreira

UFLA

Dr. Lucas Monteiro Chaves

UFLA

Dr. Júlio Sílvio de Sousa Bueno Filho  
Orientador

**LAVRAS - MG  
2012**

*Dedico este trabalho primeiramente a Deus, porque todas as coisas foram feitas por intermédio dele, e, sem ele, nada do que foi feito se fez. (João 1:3)*

*A meus pais, Jazon e Desterro, pelo apoio e incentivo ao longo de toda essa trajetória a mim destinado.*

“Não há nada bom nem mau a não ser estas duas coisas:  
a sabedoria que é um bem e a ignorância que é um mal.”

Platão

## AGRADECIMENTOS

A Deus, por dar-me força nos momentos de decisão.

À minha família, pelo carinho e amor que tiveram comigo.

À minha namorada Andréa Barros Candido, pela valiosa participação ao longo desse mestrado.

À Universidade Federal de Lavras, que me proporcionou cursar o mestrado em Estatística e Experimentação Agropecuária.

Ao Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI), pela concessão de bolsa de estudos.

Ao professor Júlio Sílvio de Sousa Bueno Filho, pelo apoio e orientação.

Ao professor Daniel Furtado Ferreira, pela coorientação.

Ao professor Lucas Monteiro Chaves, pelas sugestões dadas na defesa.

À professora Ana Carla, pelas sugestões dadas na defesa.

A todos os meus colegas de curso, que se tornaram especiais e grandes amigos.

A Paulo César Emiliano, Edcalos Miranda de Souza e Janice Ferreira do Nascimento, pela amizade e companheirismo.

A todos os professores do Departamento de Ciências Exatas (DEX), que são exemplos para a minha carreira acadêmica.

A todos os professores da Universidade Estadual da Paraíba, por serem sempre companheiros e incentivadores.

A todas as pessoas que, de alguma forma, participaram na realização deste sonho.

*André Barbosa Ventura da Silva*

## RESUMO

Frequentemente várias fontes de variação indesejadas ao experimento são observadas e precisam ser controladas. Nestes casos, uma possibilidade é empregar delineamentos em quadrado latino e suas variantes. Um quadrado sudoku é um tipo especial de quadrado latino, que além de permitir o controle de fontes de variação no sentido de linhas e colunas, também permite o controle local em subquadrados contendo todos os tratamentos. Contudo, nos sudokus a nova fonte de controle local não é ortogonal com linhas e colunas (como seria em um quadrado greco-latino, por exemplo). Foi corrigido um erro na literatura em relação aos graus de liberdade da análise de variância (ANAVA) em sudoku. Como critério de comparação desses delineamentos, foram avaliados a taxa de erro tipo I e poder do teste F para tratamentos via simulação Monte Carlo. Quanto ao erro tipo I, os sudokus apresentaram mesmo tamanho que os níveis nominais praticados. Os delineamentos do tipo sudoku apresentaram maior poder do que os quadrados latinos comuns de mesmas dimensões. Foi considerado um exemplo da análise sensorial de alimentos em que havia grande número de tratamentos (16, em um fatorial com interações). Este caso é comum e, em geral, demanda quadrados de tamanho grande que induzem potenciais heterogeneidade e podem se beneficiar do controle adicional dos sudokus. Recomenda-se a utilização dos sudokus sempre que se puder identificar três fontes de variação em estrutura adequada para o controle local.

Palavras-chave: Delineamentos experimentais. Erro tipo I. Poder. Quadrado latino. Simulação Monte Carlo. Sudoku.

## ABSTRACT

It is a common situation to have many sources of undesired variability observed and controlled to a better experiment. In these cases we can use latin square designs and designs alike. A sudoku is a special case of latin square that allows for local control in subsquares containing all treatments. The new source of variation is however not orthogonal to rows and columns (as would be the case of a graeco-latin square). Criteria to compare these designs were type I error rate and power of F test for treatment comparisons calculated via Monte-Carlo simulation. We found and fix an error in the literature with respect to the degrees of freedom for the ANOVA. Sudoku designs were more powerfull and more efficient than latin squares in the presence of a source of variation related to squares. An example was given from sensory analysis in which a great number of treatments were compared (16 in a factorial with interactions). This is a rather common situation that usually demands large blocks that induce potential heterogeneity and could benefit from sudoku. Sudoku can be recommended as an efficient and powerful design in situations that we can identify convenient structure to local control.

Keywords: Design of experiments. Latin square. Monte Carlo Simulation. Power. Sudoku. Type I error.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	12
<b>1.1</b>	<b>Objetivos</b>	14
<b>1.1.1</b>	<b>Objetivo geral</b>	14
<b>1.1.2</b>	<b>Objetivos específicos</b>	14
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	15
<b>2.1</b>	<b>Análise Sensorial de Alimentos</b>	15
<b>2.2</b>	<b>Delineamento em Quadrado Latino (QL)</b>	16
<b>2.2.1</b>	<b>Quadrados Latinos Repetidos</b>	19
<b>2.2.2</b>	<b>Quadrados Latinos Ortogonais</b>	22
<b>2.2.3</b>	<b>Quadrado Greco-Latino</b>	23
<b>2.2.4</b>	<b>Quadrado Sudoku (QS)</b>	25
<b>2.2.5</b>	<b>Variantes do Sudoku</b>	29
<b>2.2.6</b>	<b>Modelo Linear</b>	32
<b>2.2.7</b>	<b>Modelo Misto</b>	33
<b>2.2.8</b>	<b>Sistema de Equações Normais (S.E.N.)</b>	35
<b>2.2.9</b>	<b>Análise de Variância</b>	45
<b>2.3</b>	<b>Simulações usando o método Monte Carlo (MMC)</b>	49
<b>2.4</b>	<b>Tipos de erro e poder do teste</b>	49
<b>3</b>	<b>METODOLOGIA</b>	51
<b>3.1</b>	<b>Modelo de Simulação</b>	52
<b>3.2</b>	<b>Modelos de Análise</b>	54
<b>3.3</b>	<b>Erro tipo I e poder</b>	55
<b>3.4</b>	<b>Exemplo de Aplicação em Ciência dos Alimentos</b>	56
<b>4</b>	<b>RESULTADOS</b>	60
<b>5</b>	<b>DISCUSSÃO</b>	72
<b>6</b>	<b>CONCLUSÃO</b>	74
	<b>REFERÊNCIAS</b>	75
	<b>APÊNDICE</b>	78

## LISTA DE FIGURAS

Figura 1	Mesmos provadores repetidos em mesma ordem de prova e em diferentes ocasiões . . . . .	20
Figura 2	Mesmos provadores repetidos em diferentes ordem de prova e em diferentes ocasiões . . . . .	20
Figura 3	Mesmas ordem de prova repetidos para diferentes provadores . . .	21
Figura 4	Diferentes provadores e diferentes ordem de prova repetidos três vezes . . . . .	21
Figura 5	Exemplo de Quadrado Sudoku $9 \times 9$ . . . . .	26
Figura 6	Quadrado Sudoku com 9 linhas, 9 colunas e 9 quadrados cercados com seus respectivos códigos . . . . .	27
Figura 7	JigSaw Sudoku . . . . .	30
Figura 8	Samurai Sudoku . . . . .	30
Figura 9	Kazaguruma . . . . .	31
Figura 10	Tridoku . . . . .	31
Figura 11	Hoshi Sudoku . . . . .	32
Figura 12	Circle Sudoku . . . . .	32
Figura 13	Solução de um Sudoku $4 \times 4$ . . . . .	35
Figura 14	Quadrado Sudoku $16 \times 16$ . . . . .	57
Figura 15	Superfície de resposta para notas de provadores, em função dos graus de conservantes e Açúcar . . . . .	59
Figura 16	Poder do teste F para tratamentos sob $H_0 : QL$ (Gráficos A e B) e sob $H_0 : QS$ (Gráficos C e D) para as dimensões $4 \times 4$ do QL e QS, conforme a ordem dos fatores $\beta, \rho, \gamma$ e $\tau$ no modelo Sudoku	63
Figura 17	Poder do teste F para tratamentos sob $H_0 : QL$ (Gráficos A e B) e sob $H_0 : QS$ (Gráficos C e D) para as dimensões $9 \times 9$ do QL e QS, conforme a ordem dos fatores $\beta, \rho, \gamma$ e $\tau$ no modelo Sudoku	64
Figura 18	Poder do teste F para tratamentos sob $H_0 : QL$ (Gráficos A e B) e sob $H_0 : QS$ (Gráficos C e D) para as dimensões $16 \times 16$ do QL e QS, conforme a ordem dos fatores $\beta, \rho, \gamma$ e $\tau$ no modelo Sudoku . . . . .	65
Figura 19	Poder do teste F para tratamentos sob $H_0 : QL$ (Gráficos A e B) e sob $H_0 : QS$ (Gráficos C e D) para as dimensões $25 \times 25$ do QL e QS, conforme a ordem dos fatores $\beta, \rho, \gamma$ e $\tau$ no modelo Sudoku . . . . .	66

Figura 20 Poder do teste F para tratamentos sob $H_0 : QL$ (Gráficos A e B) e sob $H_0 : QS$ (Gráficos C e D) para as dimensões $36 \times 36$ do QL e QS, conforme a ordem dos fatores $\beta, \rho, \gamma$ e $\tau$ no modelo Sudoku . . . . .	67
Figura 21 Poder do teste F para tratamentos sob $H_0 : QL$ (Gráficos A e B) e sob $H_0 : QS$ (Gráficos C e D) para as dimensões $49 \times 49$ do QL e QS, conforme a ordem dos fatores $\beta, \rho, \gamma$ e $\tau$ no modelo Sudoku . . . . .	68
Figura 22 Poder do teste F para tratamentos sob $H_0 : QL$ (Gráficos A e B) e sob $H_0 : QS$ (Gráficos C e D) para as dimensões $64 \times 64$ do QL e QS, conforme a ordem dos fatores $\beta, \rho, \gamma$ e $\tau$ no modelo Sudoku . . . . .	69
Figura 23 Poder do teste F para tratamentos sob $H_0 : QL$ (Gráficos A e B) e sob $H_0 : QS$ (Gráficos C e D) para as dimensões $81 \times 81$ do QL e QS, conforme a ordem dos fatores $\beta, \rho, \gamma$ e $\tau$ no modelo Sudoku . . . . .	70
Figura 24 Poder do teste F para tratamentos sob $H_0 : QL$ (Gráficos A e B) e sob $H_0 : QS$ (Gráficos C e D) para as dimensões $100 \times 100$ do QL e QS, conforme a ordem dos fatores $\beta, \rho, \gamma$ e $\tau$ no modelo Sudoku . . . . .	71

## LISTA DE TABELAS

Tabela 1	Total de Quadrados Latinos padrão segundo a ordem $t$ . . . . .	18
Tabela 2	Quadrado Latino. . . . .	23
Tabela 3	Quadrado Latino. . . . .	23
Tabela 4	QL ortogonal. . . . .	23
Tabela 5	Quadrado Greco-Latino. . . . .	24
Tabela 6	Dimensões do Sudoku $p \times q$ (DONG; GEN, 2008). . . . .	28
Tabela 7	Modelos utilizados na obtenção das soma de quadrados tipo I das diversas causas de variação de um delineamento em QS $k \times k$ , com suas respectivas somas de quadrados de parâmetros (SQP), equações normais (S.E.N.) e vetor de solução. . . . .	43
Tabela 8	Esquema de análise de variância para um delineamento sudoku $k \times k$ , considerando modelo fixo e modelo aleatório. . . . .	47
Tabela 9	Tabela resumo referente aos Graus de Liberdade (GL) do modelo sudoku $k \times k$ , segundo a ordem dos fatores especificados para os 24 modelos. . . . .	48
Tabela 10	Efeitos de tratamentos simulados segundo uma superfície de resposta. . . . .	58
Tabela 11	Análise de Variância de regressão da superfície de resposta. . . . .	59
Tabela 12	Taxa de erro tipo I para tratamentos em delineamentos do tipo Sudoku e Quadrado Latinos em função do número de dimensões $k$ e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$ sob $H_0 : QL$ , considerando-se distribuição normal. Resultado de 2.000 simulações. . . . .	61
Tabela 13	Taxa de erro tipo I para tratamentos em delineamentos do tipo Sudoku e Quadrado Latinos em função do número de dimensões $k$ e níveis nominais de significância $\alpha = 0,05$ e $\alpha = 0,01$ sob $H_0 : QS$ , considerando-se distribuição normal. Resultado de 2.000 simulações. . . . .	62
Tabela 14	Dados simulados para o exemplo da análise sensorial com ajuste da superfície de resposta. . . . .	87

## 1 INTRODUÇÃO

Algumas das mais importantes contribuições para a teoria e prática da inferência estatística no século XX, foram àquelas realizadas no planejamento experimental. Uma boa parte do desenvolvimento experimental foi adiantado e estimulado por aplicações feitas na agricultura. Os princípios subjacentes da estatística no planejamento experimental foram amplamente desenvolvidos pelo renomado matemático e geneticista *Sir Ronald A. Fisher*, durante o seu trabalho na Estação Experimental *Rothamsted*, nos anos de 1920 e 1930 (NATIONAL RESEARCH COUNCIL, 1995).

As técnicas experimentais desenvolvidas são universais e se aplicam a todas as áreas de conhecimento - agronomia, medicina, engenharia, alimentos, educação, psicologia-, e os métodos estatísticos de análise são os mesmos. De qualquer forma, é na origem agrícola das ciências experimentais que se explica o uso de alguns termos técnicos.

Os planejamentos experimentais mais utilizados são os delineamentos inteiramente casualizados, os em blocos casualizados e os em quadrados latinos. A escolha de um delineamento adequado dependerá das condições onde ocorrerá o ensaio, tais condições podem apresentar características homogêneas ou heterogêneas podendo interferir na análise, contudo, um dos critérios para a escolha de um delineamento ser apropriado ou não, é verificar se esse delineamento selecionado pode identificar e isolar da análise possíveis fontes de variações que poderiam influenciar nos resultados, já que um dos grandes objetivos do pesquisador é eliminar a variabilidade sistemática e, portanto, reduzir o erro experimental.

Neste trabalho, é abordado um novo tipo de delineamento experimental, que é o quadrado sudoku, obtido pela solução de um quadrado latino, em que, além de os tratamentos não se repetirem em linhas e colunas, também não se repetem

dentro de um bloco (quadrado interno).

O Sudoku é um quebra-cabeça que surgiu em Nova York no final dos anos 70, inventado por um norte-americano especialista em quebra cabeças chamado *Howard Garnes*. Sua criação foi publicada na revista *Dell Pencil Puzzles and World Games*, da *Dell Magazines* (editora especializada em jogos de raciocínio), com o nome de *Number Place* - nos EUA, também é chamado de *Nanpure*. Sem muita repercussão, o passatempo foi levado para o Japão em 1984 por uma companhia japonesa chamada *Nikoli*. O presidente da companhia fez pequenas alterações e renomeou o jogo para *Sudoku*. Do japonês “su” (número) e “doku” (único). A popularização só aconteceu, no entanto, depois que o neozelandês *Wayne Gould*, juiz aposentado morando em *Hong Kong*, teve contato com a diversão, em 1997, e dedicou seis anos de sua vida para montar um programa de computador que gerasse novos jogos rapidamente. Com o software em mãos, ele ofereceu a novidade aos jornais britânicos e, em 12 de novembro de 2004, *The Times* deu o pontapé inicial à mania ([www.abril.com.br/noticia/diversao/no\\_168200.shtml](http://www.abril.com.br/noticia/diversao/no_168200.shtml)).

Delineamentos do tipo sudoku são casos particulares de delineamentos em quadrado latinos, com a restrição adicional de que todos os tratamentos ocorrem no mesmo bloco, o que faz do sudoku um delineamento bem estruturado e que pode ser empregado em experimentos planejados com controle local em três dimensões, além da fonte de variação associada aos tratamentos. Desta forma, o controle local pode ser feito em linhas, colunas e sub-quadrados (ou simplesmente quadrados). Esta última fonte não estaria presente em um quadrado latino usual.

Um delineamento, com muitas fontes de controle local, é adequado para situações onde o interesse está em isolar da análise possíveis fontes de variações estranhas aos tratamentos que não devem influenciar nos resultados. Caso a fonte extra de controle local seja importante, isto implica em maior poder para testar

hipóteses sobre efeito de tratamentos. Caso contrário, espera-se que o poder diminua, pois haverá menor número de graus de liberdade para estimar os resíduos do modelo.

Na análise sensorial de alimentos, frequentemente vários fatores de controle local podem ser observados, por exemplo, o provador, a ocasião de prova e a ordem de prova. Neste caso, seria adequado pensar em delineamentos ortogonais ou quase ortogonais envolvendo vários tipos de controle local. Deste modo, será discutida uma possibilidade de aplicação do delineamento do tipo sudoku na área de análise sensorial de alimentos, na qual a utilização de delineamentos mais estruturados para controle local, pode melhorar a avaliação de tratamentos.

## **1.1 Objetivos**

### **1.1.1 Objetivo geral**

O objetivo desse trabalho é avaliar os prós e contras do controle local adicional em delineamentos mais estruturados para controle local.

### **1.1.2 Objetivos específicos**

Avaliar por meio dos reflexos no erro tipo I e no poder do teste F, os prós e contras do controle local adicional em delineamentos do tipo sudoku comparado aos quadrados latinos.

Ilustrar a utilização de delineamento do tipo sudoku em um experimento fictício de dados de análise sensorial de alimentos.

## 2 REFERENCIAL TEÓRICO

### 2.1 Análise Sensorial de Alimentos

A análise sensorial é um método experimental que se baseia na utilização dos sentidos humanos; é um dos métodos mais importantes na pesquisa de análise de alimentos, pois fornece informações relevantes sobre a relação homem x alimento, em razão dos sentidos humanos serem considerados como instrumentos de medida. Portanto, aspectos relacionados como gosto e odores são avaliados em painéis com pessoas especialmente treinadas, dessa forma, a análise sensorial fornece informações exclusivas sobre o grau de aceitação de um determinado alimento, como também é amplamente utilizada para determinar sua qualidade em geral (NATALE *et al.*, 1998).

Os delineamentos mais usuais em análise sensorial, consistem na formação de blocos completos, em que cada provador (pessoa) é um bloco, e esse provador analisa todos os produtos. Neste caso, é natural que cada provador varie muito em seus julgamentos e o bloco seja heterogêneo. Assim, como acontece com qualquer painel não treinado, não há uma garantia de que tais respostas sejam confiáveis ou válidas (STONE; SIDEL, 2004).

A escala hedônica é um dos tipos de escalas utilizadas na análise sensorial por consumidores leigos, segundo Amerine *et al.* (1965 citado por VILLAVICENCIO *et al.*, 2007, pg. 1875), os melhores resultados utilizando a escala hedônica com provadores não treinados vêm de experimentos com muitas repetições ou provadores.

Como em qualquer experimento, na análise sensorial, a escolha do delineamento é fundamental. O delineamento experimental pode ser definido como o processo explícito pelo qual os dados serão coletados durante a investigação. Um

dos delineamentos mais utilizados em análise sensorial, quando se tem grande número de tratamentos, é o delineamento em blocos incompletos (DBI), no qual nem todos os tratamentos entram no mesmo bloco (MORAES, 1993).

Na seção (3.4), será exemplificada a análise de um experimento em que são comparados 16 tratamentos de um fatorial  $4 \times 4$ , com um painel de 16 provadores não treinados em um delineamento sudoku.

## 2.2 Delineamento em Quadrado Latino (QL)

Em certos experimentos, existe a necessidade de instalação de delineamentos para controlar mais de um tipo de fonte de variação, neste caso, delineamentos ortogonais são preferíveis, isso porque a variação entre os efeitos considerados são independentes entre si, isto é, não se confundem.

Os QL são delineamentos em que as linhas são ortogonais com as colunas, sendo usados para controlar fontes de variações em duas direções, ou seja, permite o bloqueio de linhas e colunas, representando assim, um delineamento com duas restrições na casualização. Em geral, um quadrado latino  $p \times p$ , é um quadrado contendo  $p$  linhas e  $p$  colunas, em que cada tratamento é representado por letras, os quais ocorrem apenas uma única vez em cada linha e coluna (MONTGOMERY, 2001).

Os tratamentos são simultaneamente agrupados de duas formas distintas: em linhas e em colunas, sendo que cada linha ou coluna contém uma só repetição de cada tratamento. Desta forma, o número de repetições (número de linhas e colunas) deve ser igual ao número de tratamentos. O delineamento em QL é dito ortogonal, pois permite modelos lineares em que se pode eliminar das comparações entre os tratamentos as diferenças entre linhas e as diferenças entre colunas.

Se, por exemplo, existir um gradiente de fertilidade no solo em duas di-

reções, é adequado formar grupos de unidades experimentais em ambas direções, grupos esses que correspondem às linhas e às colunas do QL.

Segundo Hinkelmann e Kempthorne (2008, p.374), “quadrados latinos e suas propriedades foram primeiro descritos pelo matemático suíço Euler em 1792. Na experimentação os QL foram propostos por Fisher em 1925 e 1926”. Delineamentos em QL se estendem para outros tipos de experimentações, como por exemplo, na zootecnia, em laboratório, na indústria, etc. Nesses experimentos, as letras latinas representam tratamentos, as linhas e as colunas podem representar período, métodos de aplicação de tratamentos, operadores, pacientes, ou outra fonte de variação. Ainda segundo os autores, na prática os QL mais usados são  $5 \times 5$  a  $8 \times 8$ , ou quadrados latinos  $4 \times 4$  repetidos. Quadrados maiores em geral implicam em experimentos muito grandes. Delineamentos incompletos em linhas e colunas podem ser usados quando há muitos tratamentos, em que uma das dimensões não se pode obter blocos completos por razões práticas.

O modelo linear que caracteriza um delineamento em Quadrado Latino é

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \varepsilon_{ijk} \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \end{cases} \quad (2.1)$$

em que,  $y_{ijk}$  representa o valor observado na parcela da  $k$ -ésima coluna;  $j$ -ésimo tratamento submetido a  $i$ -ésima linha;  $\mu$  é a média geral das observações;  $\alpha_i$ ,  $\tau_j$ ,  $\beta_k$  são os efeitos da  $i$ -ésima linha, do  $j$ -ésimo tratamento e da  $k$ -ésima coluna, respectivamente;  $\varepsilon_{ijk}$  representa o erro experimental, supondo  $\varepsilon_{ijk} \sim N(0, \sigma_e^2)$ .

Para o procedimento de casualização dos tratamentos às parcelas de um delineamentos em QL, Fisher e Yates (1963, p.24) definiram quadrados latinos reduzidos ou quadrados latinos padrão, como sendo aqueles em que a primeira

linha e a primeira coluna são organizadas em ordem alfabética, por exemplo, para a ordem:  $t = 3$ ,  $t = 4$ ,  $t = 5$  e  $t = 6$ , tem-se respectivamente:

$3 \times 3$	$4 \times 4$	$5 \times 5$	$6 \times 6$
A B C	A B C D	A B C D E	A B C D E F
B C A	B A D C	B A E C D	B C F A D E
C A B	C D B A	C D A E B	C F B E A D
	D C A B	D E B A C	D E A B F C
		E C D B A	E A D F C B
			F D E C B A

O primeiro passo na casualização dos tratamentos é selecionar um quadrado latino padrão ou reduzido, em seguida, (para QL de ordem 3, 4 ou 5) permuta-se todas as linhas (exceto a primeira) e todas as colunas. Por último, atribui-se aleatoriamente as letras A, B, C, ... aos nomes dos tratamentos. O número total de quadrados latinos padrão de ordem  $t$  denotado por  $T_t$ , é apresentado por Hinkelmann e Kempthorne (2008), conforme Tabela 1:

Tabela 1 Total de Quadrados Latinos padrão segundo a ordem  $t$ .

$t$	$T_t$
2	1
3	1
4	4
5	56
6	9,408
7	16.942.080
8	535.281.401.856

O cálculo para o número de quadrados latinos possíveis derivados do(s) reduzido(s) é dado por  $t!(t-1)!$ . Aplicando-se a fórmula para o caso do QL básico  $3 \times 3$ , haverá  $3!(3-1)! = 12$  quadrados latinos gerados a partir do único quadrado

básico  $3 \times 3$ , ao passo que, para o QL básico  $4 \times 4$  (Tabela 1), haverá 4 QL básicos, gerando um total de  $4!(4-1)! = 144 \times 4 = 576$  diferentes quadrados latinos  $4 \times 4$ . Portanto, a casualização de um quadrado latino  $4 \times 4$  consiste em sortear 1 entre os 576 possíveis com mesma probabilidade. Deste modo, o processo de sorteio atribui a todas as parcelas a mesma probabilidade de receber qualquer tratamento, satisfazendo a condição do quadrado latino de que todos os tratamentos estejam arranjados em cada linha e em cada coluna do delineamento. Quadrados latinos padrões podem ser encontrados nas tabelas apresentadas por Fisher e Yates (1963, p.86). Há também outras variantes de quadrados latinos, como por exemplo, QL replicados, QL retangulares e QL incompletos.

### 2.2.1 Quadrados Latinos Repetidos

Planejamento deste tipo são utilizados em quadrados latinos de tamanho pequeno, geralmente de ordem inferior à 4, com a finalidade de aumentar os graus de liberdade do resíduo, (HINKELMANN; KEMPTHORNE, 2008). No entanto, diferentes métodos de repetição levam a diferentes modelos lineares, em virtude dos fatores estarem cruzados ou aninhados entre si. O número de repetições depende, geralmente, da situação analisada. Considera-se, portanto, as seguintes situações:

i) Linhas e colunas cruzadas dentro das repetições - Seja um experimento planejado em quadrado latino de ordem  $k$ , repetido  $r$  vezes por um fabricante de produtos alimentícios. Neste caso, é considerado a mesma ordem de prova para as  $r$  repetições e mesmos  $k$  provadores são usados em cada repetição. Como ilustrado na Figura 1, as linhas do QL  $2 \times 2$  representam *provadores*, as colunas representam *ordem de prova* e as letras latinas representam os *tratamentos*.

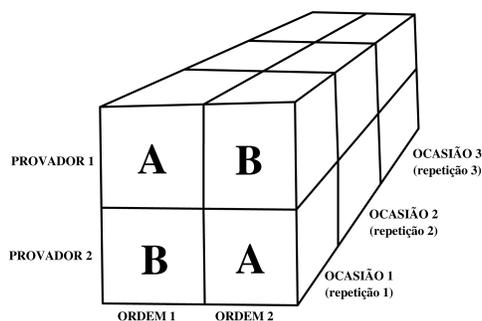


Figura 1 Mesmos provadores repetidos em mesma ordem de prova e em diferentes ocasiões

ii) Colunas cruzadas com linhas aninhadas nas repetições - Os tratamentos são distribuídos aos mesmos  $k$  provadores em diferentes ordem de prova e diferentes ocasiões.

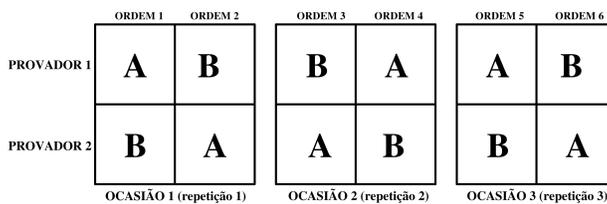


Figura 2 Mesmos provadores repetidos em diferentes ordem de prova e em diferentes ocasiões

iii) Linhas cruzadas com colunas aninhadas nas repetições - Cada uma das  $k$  colunas de ordem de prova são divididas  $r$  vezes, deste modo, cada tratamento é repetido para diferentes provadores.

	ORDEM 1	ORDEM 2	
PROVADOR 1	<b>B</b>	<b>A</b>	OCASIÃO 1 (repetição 1)
PROVADOR 2	<b>A</b>	<b>B</b>	
PROVADOR 3	<b>A</b>	<b>B</b>	OCASIÃO 2 (repetição 2)
PROVADOR 4	<b>B</b>	<b>A</b>	
PROVADOR 5	<b>B</b>	<b>A</b>	OCASIÃO 3 (repetição 3)
PROVADOR 6	<b>A</b>	<b>B</b>	

Figura 3 Mesmas ordem de prova repetidos para diferentes provadores

iv) Linhas e colunas aninhadas dentro das repetições - Neste caso, diferentes ordem de prova e diferentes provadores formam cada repetição.

	ORDEM 1	ORDEM 2	
PROVADOR 1	<b>A</b>	<b>B</b>	OCASIÃO 1 (repetição 1)
PROVADOR 2	<b>B</b>	<b>A</b>	
PROVADOR 3	<b>A</b>	<b>B</b>	OCASIÃO 2 (repetição 2)
PROVADOR 4	<b>B</b>	<b>A</b>	
PROVADOR 5	<b>B</b>	<b>A</b>	OCASIÃO 3 (repetição 3)
PROVADOR 6	<b>A</b>	<b>B</b>	

Figura 4 Diferentes provadores e diferentes ordem de prova repetidos três vezes

Em cada um desses experimentos, o processo de casualização é realizado de forma independente para cada um das situações apresentadas. Os modelos lineares referente a cada situação, bem como a análise de variância e outras classificações de QL são encontrados em Hinkelmann e Kempthorne (2008, p.391).

### 2.2.2 Quadrados Latinos Ortogonais

Dois quadrados preenchidos com as mesmas letras latinas são ditos *ortogonais*, quando sobrepostos satisfazem as seguintes condições:

- i) Cada letra, em cada quadrado, ocorre exatamente uma única vez em cada linha e em cada coluna;
- ii) Ao sobrepor os quadrados, formando pares de letras, cada letra do primeiro quadrado está combinada exatamente uma vez com uma do segundo quadrado.

Como ilustração, considerou-se dois quadrados latinos de ordem  $4 \times 4$  apresentados nas Tabelas 2 e 3. O arranjo na Tabela 4 é obtido após a sobreposição de ambos quadrados latinos. Uma vez que cada letra do primeiro quadrado é combinado com cada letra do segundo quadrado, exatamente uma vez, então os quadrados latinos são ortogonais. Como indicado por Fisher e Yates (1963), há três quadrados latinos que são ortogonais de dimensão  $4 \times 4$ , quatro quadrados ortogonais de dimensão  $5 \times 5$ , nove quadrados ortogonais de dimensão  $6 \times 6$  e seis quadrados ortogonais de dimensão  $7 \times 7$ . Os autores ainda apresentam quadrados latinos mutuamente ortogonais de outros tamanhos.

Tabela 2 Quadrado Latino.

	I	II	III	IV
1	A	B	C	D
2	B	A	D	C
3	C	D	A	B
4	D	C	B	A

Tabela 3 Quadrado Latino.

	I	II	III	IV
1	Z	Y	X	W
2	X	W	Z	Y
3	W	X	Y	Z
4	Y	Z	W	X

Tabela 4 QL ortogonal.

	I	II	III	IV
1	AZ	BY	CX	DW
2	BX	AW	DZ	CY
3	CW	DX	AY	BZ
4	DY	CZ	BW	AX

### 2.2.3 Quadrado Greco-Latino

O resultado da sobreposição dos quadrados, é conhecido como um delineamento em *Quadrado Greco-Latino*, já que nas apresentações originais um quadrado é representado por letras gregas e outro quadrado por letras latinas.

Diaz e Phillips (1995), definem um Quadrado Greco-Latino como sendo

um arranjo de  $r$  letras gregas e  $r$  letras latinas, num quadrado de  $r$  linhas e  $r$  colunas, de tal forma que cada letra Grega e cada letra latina aparecem exatamente uma vez em cada linha e em cada coluna, e também, cada combinação de letras gregas e latinas aparecendo apenas uma vez. Tais quadrados são conhecidos pela sigla MOLS (do Inglês *Mutually Orthogonal Latin Squares*) (DRAKE; VAN REES; WALLIS, 1999).

Arranjos deste tipo são interessantes para se testar a significância de tratamentos sob a presença de três restrições na casualização (linhas, colunas e letras gregas). Por exemplo: deseja-se comparar a “utilidade” de quatro tipos de processadores de textos (A, B, C e D), utilizando quatro computadores diferentes, quatro secretárias e quatro textos diferentes ( $\alpha$ ,  $\beta$ ,  $\gamma$  e  $\delta$ ). Neste caso, elimina-se as diferenças entre os PCs, secretárias e tipos de texto. Um arranjo adequado pode ser o seguinte:

Tabela 5 Quadrado Greco-Latino.

PC	Secretaria			
	1	2	3	4
1	A $\alpha$	B $\gamma$	C $\delta$	D $\beta$
2	B $\beta$	A $\delta$	D $\gamma$	C $\alpha$
3	C $\gamma$	D $\alpha$	A $\beta$	B $\delta$
4	D $\delta$	C $\beta$	B $\alpha$	A $\gamma$

Então, a secretária 1 digita o tipo de texto  $\alpha$  no PC 1 utilizando o processador de texto A, a secretária 2 digita o tipo de texto  $\gamma$  no PC 1 utilizando o processador de texto B, e assim por diante.

O modelo linear correspondente é dado por:

$$Y_{ijkm} = \mu + \beta_i + \tau_j + \gamma_k + \omega_m + \epsilon_{ijkm} \quad (i, j, k, m = 1, \dots, r) \quad (2.2)$$

em que,  $Y_{ijkm}$  representa o valor observado na parcela;  $\mu$  é a média geral das observações;  $\beta_i$ ,  $\tau_j$ ,  $\gamma_k$  e  $\omega_m$  são os efeitos do  $i$ -ésima linha, do  $j$ -ésimo tratamento, da  $k$ -ésima coluna e da  $m$ -ésima letra grega;  $\epsilon_{ijkm}$  representa o erro experimental, supondo  $\epsilon_{ijkm} \sim N(0, \sigma_e^2)$ .

O delineamento em quadrado greco-latino é balanceado (DIAZ; PHILLIPS, 1995), assumindo que os efeitos de linhas, colunas, letras latinas e letras gregas são aditivos (sem interações entre efeitos), sendo possível testá-los separadamente e, de forma independente, cada uma das quatro hipóteses com apenas  $r^2$  observações ao invés de  $r^4$  como seria necessário em um delineamento completo incluindo todas as combinações de níveis.

A vantagem desse tipo de delineamento está na redução do número total de parcelas para grandes experimentos, tornando o experimento mais simples e econômico, porém, para utilização desse tipo de delineamento, é necessário que não haja interação entre os fatores. A desvantagem seria a aleatorização complexa, que nem sempre pode ser posta em prática no experimento.

#### 2.2.4 Quadrado Sudoku (QS)

Delineamentos do tipo Sudoku são casos particulares de delineamentos em Quadrado Latinos, com a restrição adicional de que todos os tratamentos ocorrem juntos no mesmo subquadrado (doravante chamado simplesmente de quadrado). Tais delineamentos consideram três restrições na casualização dos tratamentos às parcelas, assim, o controle local é feito em linha, coluna e quadrado, no que difere

do quadrado latino com a inclusão de quadrado no modelo linear do Sudoku.

O Sudoku, em sua forma padrão, consiste em um quadrado  $9 \times 9$  números, cujos números são dispostos de 1 a 9, os quais são distribuídos em 9 linhas e 9 colunas, distribuídos ainda em 9 quadrados, de modo que cada dígito aparece apenas uma vez em cada linha, coluna e quadrado (DONG; GEN, 2008). Na Figura 5, é apresentado um sudoku  $9 \times 9$ , em que os números na cor preta representam o gerador do sudoku e na cor vermelha sua solução.

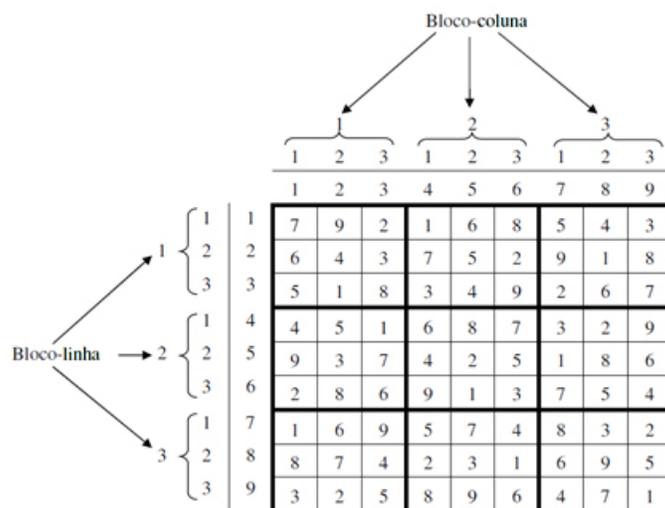
5	8	3	7	2	4	1	9	6
1	2	6	9	3	5	8	7	4
4	9	7	6	1	8	3	5	2
3	5	9	8	4	2	7	6	1
8	1	4	5	6	7	2	3	9
7	6	2	3	9	1	5	4	8
2	3	8	4	5	9	6	1	7
6	4	1	2	7	3	9	8	5
9	7	5	1	8	6	4	2	3

Figura 5 Exemplo de Quadrado Sudoku  $9 \times 9$

Em termos de delineamento experimental,  $k^2$  unidades (parcelas) experimentais são divididas em  $k$  linhas,  $k$  colunas e  $k$  quadrados, em que cada quadrado contém  $k$  unidades experimentais com  $k$  tratamentos.

Segundo Bailey, Cameron e Connelly (2008), a finalidade de um delineamento na experimentação é garantir que todos os tratamentos estejam expostos às diferentes condições experimentais, para que dessa forma haja comparações justas (“gerechte”) entre tratamentos. A palavra alemã “gerechte” significa “justo”, utilizada no contexto da experimentação como *gerechte designs*. Gerechte Designs foram introduzidos por Behrens (1956), como uma especialização de quadrados latinos. Um Gerechte Design de ordem  $n$  é uma matriz  $n \times n$  cujas  $n^2$  células são

particionadas em  $n$  regiões de maneira que cada símbolo ocorre uma única vez em cada linha, uma vez em cada coluna, e uma vez em cada região. Deste modo, soluções de sudoku são delineamentos do tipo “justo”. Geralmente para descrever um quadrado sudoku, a linha, a coluna e o quadrado são codificados. A ilustração na Figura 6 mostra um quadrado sudoku  $9 \times 9$ , de acordo com a estrutura sugerida por Dong e Gen (2008).



$k = pq$  sendo  $p$  ou  $q \geq 2$ . Assim, a ordem  $k$  do quadrado sudoku será um número não primo conforme mostra a Tabela 6.

Tabela 6 Dimensões do Sudoku  $p \times q$  (DONG; GEN, 2008).

Parâmetros			Quadrados
$k$	$p$	$q$	(linha $\times$ coluna)
4	2	2	
6	3	2	$2 \times 3^*$
8	4	2	$2 \times 4^*$
9	3	3	
10	5	2	$2 \times 5^*$
12	4	3	$3 \times 4^*$
12	6	2	$2 \times 6^*$
14	7	2	$2 \times 7^*$
15	5	3	$3 \times 5^*$
16	4	4	
16	8	2	$2 \times 8^*$
18	6	3	$3 \times 6^*$
18	9	2	$2 \times 9^*$
20	5	4	$4 \times 5^*$
20	10	2	$2 \times 10^*$

\*Obtidos a partir da troca de  $p$  por  $q$ .

Neste trabalho, são analisados alguns casos particulares de quadrados sudokus perfeitos, em que,  $\sqrt{k} = p = q$ , sendo a raiz quadrada exata.

Como dito anteriormente, sudokus são casos especiais de quadrados latinos, sendo que o número possível de quadrados latinos  $9 \times 9$  chega a ultrapassar a casa dos cinco octilhões, isto é,  $5.524.751.496.156.892.842.531.225.600$ , correspondendo ao valor aproximado de  $5,5 \times 10^{27}$ . Os autores Bertram Felgenhauer e Frazer Jarvis (2005), também mostraram o número de quadrados sudokus  $9 \times 9$ , chegando ao número aproximado de 830 milhões de vezes menor ao anterior, ou

seja, 6.670.903.752.021.072.936.960, aproximadamente  $6,6 \times 10^{21}$ . Mais tarde, os autores Ed Russell e Frazer Jarvis (2005), eliminaram alguns sudokus que tinham soluções óbvias (reflexões, rotações etc.), assim, o número de sudokus  $9 \times 9$  essencialmente diferentes passou a ser 5.472.730.538.

Ponnuswamy e Subramani (2009) propuseram um método sequencial para construção de sudoku de qualquer ordem. Segundo os autores, a aleatoriedade pode ser alcançada selecionando qualquer um dos quadrados sudokus disponíveis ou construídos a partir do método apresentado. Dong e Gen (2008) apresentaram um método de casualização mais interessante do sudoku  $k \times k$ , com a vantagem de haver várias soluções de sudokus em relação ao método apresentado por Ponnuswamy e Subramani (2009). No apêndice, segue uma rotina implementada em R para realização da casualização do sudoku quadrado perfeito  $k \times k$ , de acordo com as propostas sugeridas pelos autores Ponnuswamy e Subramani (2009) e Dong e Gen (2008).

### 2.2.5 Variantes do Sudoku

Após a invenção do Sudoku como quebra cabeça, surgiram então, vários outros tipos de sudokus com diferentes configurações. No site <http://www.cross-plus-a.com/sudoku.htm> podem ser encontrados alguns deles. Exemplos de variantes de sudokus são:

2	4	6	5	8	3	1	9	7
3	9	2	1	7	6	5	8	4
1	8	7	4	9	5	3	2	6
4	5	9	6	2	1	7	3	8
7	3	8	9	1	2	4	6	5
5	6	1	2	3	4	8	7	9
8	2	3	7	4	9	6	5	1
9	1	5	8	6	7	2	4	3
6	7	4	3	5	8	9	1	2

Figura 7 JigSaw Sudoku

3	5	1	6	8	7	9	2	4			6	4	2	5	1	9	3	7	8	
8	4	7	9	5	2	3	1	6			9	5	7	6	8	3	2	4	1	
2	9	6	3	1	4	7	8	5			8	1	3	7	2	4	6	5	9	
5	6	4	1	2	9	8	7	3			5	2	9	3	7	1	8	6	4	
7	8	2	4	3	5	6	9	1			7	3	8	4	9	6	5	1	2	
1	3	9	8	7	6	5	4	2			4	6	1	8	5	2	7	9	3	
4	7	3	2	6	8	1	5	9	3	4	7	2	8	6	1	4	5	9	3	7
9	1	5	7	4	3	2	6	8	5	1	9	3	7	4	9	6	8	1	2	5
6	2	8	5	9	1	4	3	7	8	6	2	1	9	5	2	3	7	4	8	6
						9	1	6	2	8	5	7	4	3						
						7	4	3	6	9	1	5	2	8						
						8	2	5	7	3	4	9	6	1						
7	5	8	2	4	3	6	9	1	4	7	3	8	5	2	3	1	4	7	9	6
1	3	2	6	8	9	5	7	4	1	2	8	6	3	9	8	5	7	2	4	1
9	4	6	1	7	5	3	8	2	9	5	6	4	1	7	2	6	9	8	3	5
2	7	9	5	1	6	8	4	3				7	9	8	4	3	5	1	6	2
4	8	1	7	3	2	9	6	5				2	6	5	1	9	8	3	7	4
3	6	5	4	9	8	2	1	7				1	4	3	7	2	6	5	8	9
5	9	4	8	2	7	1	3	6				3	2	4	6	8	1	9	5	7
6	1	3	9	5	4	7	2	8				9	7	1	5	4	3	6	2	8
8	2	7	3	6	1	4	5	9				5	8	6	9	7	2	4	1	3

Figura 8 Samurai Sudoku

2	9	6	4	3	5	8	7	1
1	7	5	2	6	8	3	4	9
3	4	8	1	7	9	2	5	6
8	2	4	7	1	6	5	9	3
6	3	1	9	5	2	7	8	4
9	5	7	3	8	4	6	1	2
7	8	2	6	4	1	9	3	5
4	6	3	5	9	7	1	2	8
5	1	9	8	2	3	4	6	7
3	4	8	1	7	6	9	5	2
1	6	7	5	9	2	3	8	4
2	5	9	4	8	3	7	1	6
5	3	2	8	6	4	1	7	9
6	7	4	9	1	5	2	3	8
9	8	1	2	3	7	4	6	5
8	1	3	6	2	9	5	4	7
7	9	5	3	4	8	6	2	1
4	2	6	7	5	1	8	9	3
7	3	4	8	9	6	1	5	2
1	6	2	4	7	5	8	3	9
9	8	5	3	2	1	4	6	7

Figura 9 Kazaguruma

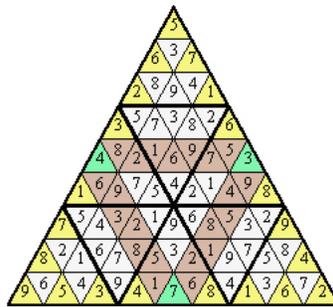


Figura 10 Tridoku

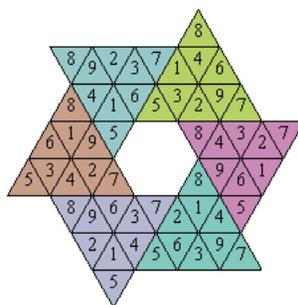


Figura 11 Hoshi Sudoku

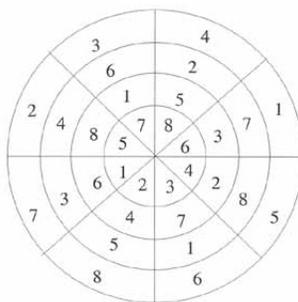


Figura 12 Circle Sudoku

### 2.2.6 Modelo Linear

Segundo Dean e Voss (1999), um modelo deve indicar explicitamente a relação que se acredita existir entre a variável resposta e as principais fontes de variação identificadas no planejamento. O modelo linear que descreve as observações de um delineamento em quadrado sudoku é apresentado por (DONG; GEN, 2008):

$$Y_{ijlm} = \mu + \beta_i + \rho_j + \gamma_l + \tau_m + \epsilon_{ijlm} \quad (i, j, l, m = 1, \dots, k) \quad (2.3)$$

em que,  $Y_{ijlm}$  representa o valor observado na parcela do  $i$ -ésimo quadrado,  $j$ -

ésima linha e  $l$ -ésima coluna submetido ao  $m$ -ésimo tratamento;  $\mu$  é a média geral das observações;  $\beta_i$ ,  $\rho_j$ ,  $\gamma_l$  e  $\tau_m$  são os efeitos do  $i$ -ésimo quadrado, da  $j$ -ésima linha, da  $l$ -ésima coluna e do  $m$ -ésimo tratamento;  $\epsilon_{ijlm}$  representa o erro experimental, supondo  $\epsilon_{ijlm} \sim N(0, \sigma_e^2)$ .

Se os efeitos  $\beta_i$ ,  $\rho_j$ ,  $\gamma_l$  e  $\tau_m$  são parâmetros de efeitos fixos, então, tem-se a seguinte restrição

$$\sum_1^k \beta_i = \sum_1^k \rho_j = \sum_1^k \gamma_l = \sum_1^k \tau_m = 0 \quad (2.4)$$

Por outro lado, se tais parâmetros são de efeitos aleatórios, as seguintes pressuposições são consideradas

$$\beta_i \sim N(0, \sigma_q^2), \rho_j \sim N(0, \sigma_l^2), \gamma_l \sim N(0, \sigma_c^2), \tau_m \sim N(0, \sigma_t^2) \quad (2.5)$$

Ponnuswamy e Subramani (2009), apresentaram mais quatro tipos de modelos sudokus. Além dos fatores apresentados em (2.3), os autores incluem outros fatores no modelo, como por exemplo efeitos de bloco-linha, bloco-coluna, efeitos aninhados e efeitos cruzados. As expressões, para cálculo das somas de quadrados, bem como a análise de variância também são apresentadas.

### 2.2.7 Modelo Misto

Dependendo dos objetivos da pesquisa, os efeitos do modelo (2.3) podem ser classificados como sendo *fixo* ou *aleatório*. Se os efeitos do modelo são especificamente escolhidos pelo pesquisador, os efeitos são *fixos* e as conclusões na análise ficam restritas aos efeitos em teste, porém, em algumas situações experimentais, os fatores são escolhidos aleatoriamente a partir de uma população, e

neste caso, o pesquisador deseja generalizar as conclusões para toda a população de onde tais efeitos foram amostrados, nesta situação, o efeito é chamado *aleatório* (SEARLE; CASELLA; MCCULLOCH, 1992).

Segundo os autores, pode-se considerar que quase todos os modelos estatísticos têm ao menos um intercepto fixo e erro aleatório, podendo ser chamados de casos particulares de modelos mistos. Na prática, o nome “modelo misto” é geralmente reservado para qualquer modelo que tem efeitos fixos (além de  $\mu$ ) e efeitos aleatórios, além do erro.

Em situações reais de aplicação, os modelos mistos têm uso mais amplo do que modelos aleatórios. Por exemplo, um pesquisador está interessado em comparar cinco marcas de iogurtes (tratamentos) comercializados em determinada região utilizando provadores (quadrados). Neste caso, os tratamentos são de efeito fixo, porque o pesquisador estabeleceu a priori as marcas de iogurtes que irá comparar. Quanto aos efeitos dos provadores, mais provavelmente devem ser tomados como aleatórios, supondo que foram sorteados de um conjunto de provadores para os quais se deseja generalizar a inferência.

Se, porém, ao invés do pesquisador especificar as marcas de iogurtes e tomasse, por exemplo, algumas marcas existentes no mercado, então o efeito de iogurte também pode ser considerado como sendo aleatório e neste caso a hipótese a ser testada é de que a componente da variância, para “marcas de iogurte”, é nula.

Para análise do modelo misto, o seguinte modelo é apresentado por Robinson (1991):

$$y = X\beta + Zb + \varepsilon \quad (2.6)$$

em que,

$y$  é um vetor de  $n$  observações, com dimensões  $n \times 1$ ;

$X$  é a matrix de delineamento dos efeitos fixos, com dimensões  $n \times p$ ,  $p$  é o número

de parâmetros associados aos efeitos fixos;

$\beta$  é um vetor de parâmetros de efeitos fixos com dimensões  $p \times 1$ ;

$Z$  é a matrix de delineamento dos efeitos aleatórios, com dimensões  $n \times q$ ,  $q$  é o número de parâmetros associados aos efeitos aleatórios;

$b$  é um vetor de parâmetros de efeitos aleatórios com dimensões  $q \times 1$ ;

$\varepsilon$  é um vetor de resíduos com dimensões  $n \times 1$ ;

### 2.2.8 Sistema de Equações Normais (S.E.N.)

A série de equações que descrevem as observações de um delineamento são chamadas de *Equações Normais*, sendo definida como (SEARLE, 1987):

$$X'X\theta^0 = X'Y \quad (2.7)$$

em que  $X$  representa a matriz de delineamento,  $\theta^0$  representa o vetor de parâmetros a serem estimados e  $Y$  o vetor de observações.

Para exemplificar o S.E.N. de um delineamento em quadrado sudoku, considera-se um caso particular de um sudoku  $4 \times 4$  como segue:

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

Figura 13 Solução de um Sudoku  $4 \times 4$

Os números entre parênteses representam a ordem das parcelas de 1 à 16,

os maiores representam a ordem dos quadrados de 1 à 4, e os demais representam os tratamentos de 1 à 4, sendo o número de tratamentos igual ao número de quadrados (repetições).

Para este caso particular, considerou-se a ordem dos fatores do modelo (2.3) da esquerda para direita (quadrado, linha, coluna e tratamento), assim por exemplo, observa-se na parcela 6 a ocorrência do tratamento 4, no quadrado 1, na linha 2 e coluna 2. Ao aplicar o valor esperado no modelo (2.3), tem-se as seguintes equações normais para o caso particular considerado:

$$E(Y_{1113}) = \mu + \beta_1 + \rho_1 + \gamma_1 + \tau_3$$

$$E(Y_{1121}) = \mu + \beta_1 + \rho_1 + \gamma_2 + \tau_1$$

$$E(Y_{1212}) = \mu + \beta_1 + \rho_2 + \gamma_1 + \tau_2$$

$$E(Y_{1224}) = \mu + \beta_1 + \rho_2 + \gamma_2 + \tau_4$$

$$E(Y_{2311}) = \mu + \beta_2 + \rho_3 + \gamma_1 + \tau_1$$

$$E(Y_{2323}) = \mu + \beta_2 + \rho_3 + \gamma_2 + \tau_3$$

$$E(Y_{2414}) = \mu + \beta_2 + \rho_4 + \gamma_1 + \tau_4$$

$$E(Y_{2422}) = \mu + \beta_2 + \rho_4 + \gamma_2 + \tau_2$$

$$E(Y_{3134}) = \mu + \beta_3 + \rho_1 + \gamma_3 + \tau_4$$

$$E(Y_{3142}) = \mu + \beta_3 + \rho_1 + \gamma_4 + \tau_2$$

$$E(Y_{3233}) = \mu + \beta_3 + \rho_2 + \gamma_3 + \tau_3$$

$$E(Y_{3241}) = \mu + \beta_3 + \rho_2 + \gamma_4 + \tau_1$$

$$E(Y_{4332}) = \mu + \beta_4 + \rho_3 + \gamma_3 + \tau_2$$

$$E(Y_{4344}) = \mu + \beta_4 + \rho_3 + \gamma_4 + \tau_4$$

$$E(Y_{4431}) = \mu + \beta_4 + \rho_4 + \gamma_3 + \tau_1$$

$$E(Y_{4443}) = \mu + \beta_4 + \rho_4 + \gamma_4 + \tau_3$$

Dessa forma, a matriz  $X$  de delineamento terá a seguinte estrutura:

Parcela	$Xm$	Quadrado ( $\beta$ )				Linha ( $\rho$ )				Coluna ( $\gamma$ )				Tratamento ( $\tau$ )			
	$\mu$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\rho_1$	$\rho_2$	$\rho_3$	$\rho_4$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\tau_1$	$\tau_2$	$\tau_3$	$\tau_4$
1	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0
2	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0
3	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1
4	1	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0
5	1	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0
6	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1
7	1	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0
8	1	0	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0
9	1	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	0
10	1	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0
11	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0
12	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1
13	1	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1
14	1	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0	0
15	1	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0
16	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0

$X$

em que os valores binários “0” e “1” na matrix  $X$  de delineamento, representam respectivamente a ausência ou presença dos efeitos nas parcelas.

De um modo geral, as submatrizes da matrix  $X$  são representadas respectivamente por  $\beta_{n \times k}$ ,  $\rho_{n \times k}$ ,  $\gamma_{n \times k}$  e  $\tau_{n \times k}$ , em que  $n$  é o total de observações no experimento e  $k$  é a ordem do quadrado sudoku, dessa forma, cada uma das  $n$  linhas da matrix correspondem a uma das observações. O vetor  $Xm$  é composto de elementos todos iguais a 1 e corresponde a média  $\mu$ ; a matrix  $\beta$  corresponde aos efeitos de quadrados; a matrix  $\rho$  corresponde aos efeitos de linhas; a matrix  $\gamma$  corresponde aos efeitos de colunas; a matrix  $\tau$  corresponde aos efeitos de tratamentos.

Consideremos um exemplo numérico com observações simuladas para o vetor  $Y$  segundo uma distribuição normal. As identidades matriciais úteis para a análise que são apresentadas a seguir são apenas ilustrativas do procedimento geral que pode ser encontrado em Searle (1987).



referente ao vetor de 1's.

Neste caso, a matriz  $X'X$  não tem inversa clássica ( $\det = 0$ ), sendo essa denominada *singular*. Então, para obter solução única para o S.E.N., pode-se impor restrições paramétricas para eliminar as referidas dependências lineares. Uma maneira de resolver esse problema, é completar o posto da matrix  $X$  de delineamento, produzindo a matrix reduzida  $X_r$ , como se segue:

- i) Para o fator  $\beta$ , elimina-se a última coluna ( $\beta_4$ );
- ii) Para o fator  $\rho$ , elimina-se a segunda e quarta coluna ( $\rho_2$  e  $\rho_4$ );
- iii) Para o fator  $\gamma$ , elimina-se a segunda e quarta coluna ( $\gamma_2$  e  $\gamma_4$ );
- iv) Para o fator  $\tau$ , elimina-se a última coluna ( $\tau_4$ ).

Então, a matrix reduzida denotada por  $X_r$ , é apresentada a seguir:

Parcela	Xm	Quadrados ( $\beta_r$ )			Linha ( $\rho_r$ )		Coluna ( $\gamma_r$ )		Tratamento ( $\tau_r$ )		
	$\mu$	$\beta_1$	$\beta_2$	$\beta_3$	$\rho_1$	$\rho_3$	$\gamma_1$	$\gamma_3$	$\tau_1$	$\tau_2$	$\tau_3$
1	1	1	0	0	1	0	1	0	0	0	1
2	1	1	0	0	1	0	0	0	1	0	0
3	1	0	0	1	1	0	0	1	0	0	0
4	1	0	0	1	1	0	0	0	0	1	0
5	1	1	0	0	0	0	1	0	0	1	0
6	1	1	0	0	0	0	0	0	0	0	0
7	1	0	0	1	0	0	0	1	0	0	1
8	1	0	0	1	0	0	0	0	1	0	0
9	1	0	1	0	0	1	1	0	1	0	0
10	1	0	1	0	0	1	0	0	0	0	1
11	1	0	0	0	0	1	0	1	0	1	0
12	1	0	0	0	0	1	0	0	0	0	0
13	1	0	1	0	0	0	1	0	0	0	0
14	1	0	1	0	0	0	0	0	0	1	0
15	1	0	0	0	0	0	0	1	1	0	0
16	1	0	0	0	0	0	0	0	0	0	1

$\underbrace{\hspace{15em}}_{X_r}$

Deste modo tem-se o modelo restrito na forma  $E(y) = X_r\theta_r$ , em que  $X_r$  tem posto coluna completo. O S.E.N. reduzido é apresentado em (2.9).

$$\underbrace{\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|} \hline 16 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\ \hline 4 & 4 & 0 & 0 & 2 & 0 & 2 & 0 & 1 & 1 & 1 \\ \hline 4 & 0 & 4 & 0 & 0 & 2 & 2 & 0 & 1 & 1 & 1 \\ \hline 4 & 0 & 0 & 4 & 2 & 0 & 0 & 2 & 1 & 1 & 1 \\ \hline 4 & 2 & 0 & 2 & 4 & 0 & 1 & 1 & 1 & 1 & 1 \\ \hline 4 & 0 & 2 & 0 & 0 & 4 & 1 & 1 & 1 & 1 & 1 \\ \hline 4 & 2 & 2 & 0 & 1 & 1 & 4 & 0 & 1 & 1 & 1 \\ \hline 4 & 0 & 0 & 2 & 1 & 1 & 0 & 4 & 1 & 1 & 1 \\ \hline 4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 4 & 0 & 0 \\ \hline 4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 4 & 0 \\ \hline 4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 4 \\ \hline \end{array}}_{X_r'X_r} \cdot \underbrace{\begin{array}{|c|} \hline \hat{\mu} \\ \hline \hat{\beta}_1 \\ \hline \hat{\beta}_2 \\ \hline \hat{\beta}_3 \\ \hline \hat{\rho}_1 \\ \hline \hat{\rho}_3 \\ \hline \hat{\gamma}_1 \\ \hline \hat{\gamma}_3 \\ \hline \hat{\tau}_1 \\ \hline \hat{\tau}_2 \\ \hline \hat{\tau}_3 \\ \hline \end{array}}_{\hat{\theta}_r} = \underbrace{\begin{array}{|c|} \hline 480,6 \\ \hline 118,1 \\ \hline 121,7 \\ \hline 123,0 \\ \hline 118,8 \\ \hline 121,0 \\ \hline 121,0 \\ \hline 121,9 \\ \hline 117,2 \\ \hline 119,3 \\ \hline 121,9 \\ \hline \end{array}}_{X_r'Y} \quad (2.9)$$

A solução única do S.E.N. reduzido é dada pelo método de mínimos quadrados, que consiste em minimizar a diferença entre o valor observado e o estimado, ou seja,  $\hat{\theta}_r = (X_r'X_r)^{-1}X_r'y$ . Então, para o exemplo numérico, tem-se os seguintes resultados:

$$\hat{\theta}_r = \begin{array}{|c|} \hline 29,27 \\ \hline 0,67 \\ \hline 0,84 \\ \hline 2,05 \\ \hline -0,87 \\ \hline 0,63 \\ \hline 1,05 \\ \hline 0,76 \\ \hline -1,25 \\ \hline -0,74 \\ \hline -0,09 \\ \hline \end{array} \begin{array}{l} \rightarrow \hat{\mu} \\ \rightarrow \hat{\beta}_1 \\ \rightarrow \hat{\beta}_2 \\ \rightarrow \hat{\beta}_3 \\ \rightarrow \hat{\rho}_1 \\ \rightarrow \hat{\rho}_3 \\ \rightarrow \hat{\gamma}_1 \\ \rightarrow \hat{\gamma}_3 \\ \rightarrow \hat{\tau}_1 \\ \rightarrow \hat{\tau}_2 \\ \rightarrow \hat{\tau}_3 \end{array}$$

Durante o processo de imposição de restrições na matriz  $X$  de delineamento, alguns efeitos (colunas) foram eliminados. No exemplo em questão, foram eliminados os efeitos:  $\hat{\beta}_4$ ,  $\hat{\rho}_2$ ,  $\hat{\rho}_4$ ,  $\hat{\gamma}_2$ ,  $\hat{\gamma}_4$  e  $\hat{\tau}_4$ . Entretanto, após a estimação dos demais efeitos, pode-se então recuperar (estimar) os efeitos faltantes, somando

simplesmente as estimativas dos efeitos que não foram eliminados e colocando um sinal negativo no resultado. Deste modo, teremos:

$$\hat{\beta}_4 = -(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3) = -(0,67 + 0,84 + 2,05) = -3,56$$

$$\hat{\rho}_2 = -(\hat{\rho}_1) = -(-0,87) = 0,87$$

$$\hat{\rho}_4 = -(\hat{\rho}_3) = -(0,63) = -0,63$$

$$\hat{\gamma}_2 = -(\hat{\gamma}_1) = -(1,05) = -1,05$$

$$\hat{\gamma}_4 = -(\hat{\gamma}_3) = -(0,76) = -0,76$$

$$\hat{\tau}_4 = -(\hat{\tau}_1 + \hat{\tau}_2 + \hat{\tau}_3) = -(-1,25 - 0,74 - 0,09) = 2,08$$

Dessa forma, essa solução atende à restrição previamente imposta, segundo a qual:  $\sum_1^k \beta_i = \sum_1^k \rho_j = \sum_1^k \gamma_l = \sum_1^k \tau_m = 0$

De acordo com Bailey (2008), a ortogonalidade é um conceito importante na estatística, porque vetores ortogonais correspondem a variáveis aleatórias com correlação zero. Na análise de variância, a decomposição do vetor de observações  $\mathbf{Y}$  é feita com vetores mutuamente ortogonais, baseada no teorema de Pitágoras. Embora delineamento do tipo sudoku não seja completamente ortogonal, devido a colinearidade existente entre bloco $\times$ linha e bloco $\times$ coluna, porém, sua análise é realizada de maneira ortogonal através de projeções ortogonais. No sudoku, a decomposição ortogonal acontece do seguinte modo: cada vez que um subespaço é linearmente dependente a um subespaço anterior, então, uma ou mais colunas desse atual subespaço é excluída, eliminando dessa forma a dependência linear existente, assim, cada subespaço adicionado fica ortogonal ao anterior. Por exemplo: Seja a matrix  $X$  de delineamento, o espaço vetorial e as submatrizes  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  os subespaços de  $X$ . Então, os subespaços dos quadrados é ortogonal a

média geral (quando excluído umas de suas colunas), os subespaços das linhas é ortogonal aos quadrados (quando excluído duas colunas de  $\rho$ ), os subespaços das colunas é ortogonal aos quadrados e as linhas (quando excluído duas colunas de  $\gamma$ ), os subespaços dos tratamentos é ortogonal aos demais (quando excluído uma das colunas de  $\tau$ ). Deste modo, as somas de quadrados são obtidas por projeções ortogonais utilizando inversa simples da seguinte forma:

$$\begin{aligned} P &= X_r[(X_r)'X_r]^{-1}(X_r)' \\ P_1 &= X_{r_1}[(X_{r_1})'X_{r_1}]^{-1}(X_{r_1})' \\ P_2 &= X_{r_2}[(X_{r_2})'X_{r_2}]^{-1}(X_{r_2})' \\ P_3 &= X_{r_3}[(X_{r_3})'X_{r_3}]^{-1}(X_{r_3})' \\ P_4 &= X_{r_4}[(X_{r_4})'X_{r_4}]^{-1}(X_{r_4})' \\ P_5 &= X_{r_5}[(X_{r_5})'X_{r_5}]^{-1}(X_{r_5})' \end{aligned}$$

em que,  $P$  é o projetor ortogonal sobre a matrix  $X_r$  e  $P_i$ ,  $i = 1, \dots, 5$  são projetores ortogonais em  $X_{r_i}$ ,  $i = 1, \dots, 5$  respectivamente. Sendo as seguintes submatrizes envolvidas:

$X_{r_1}$  : coluna 1 da matrix  $X_r$ ;

$X_{r_2}$  : coluna 1 até a  $i$ -ésima coluna de  $\beta_r$ ;

$X_{r_3}$  : coluna 1 até a  $j$ -ésima coluna de  $\rho_r$ ;

$X_{r_4}$  : coluna 1 até a  $l$ -ésima coluna de  $\gamma_r$ ;

$X_{r_5}$  : coluna 1 até a  $m$ -ésima coluna de  $\tau_r$ .

Os respectivos graus de liberdades dos fatores no modelo linear do sudoku são dados pela dimensão do espaço coluna de  $\beta_r$ ,  $\rho_r$ ,  $\gamma_r$  e  $\tau_r$ . Os graus de liberdade associados ao resíduo é dado pela diferença entre o número de observações do vetor  $\mathbf{Y}$  e o posto de  $X_r$ , ou seja,  $n - p(X_r)$ . Os graus de liberdade total é dado pela diferença entre o número de observações do vetor  $\mathbf{Y}$  e o posto de  $X_{r_1}$ , isto é,  $n - p(X_{r_1})$ . As somas de quadrados são obtidas da seguinte forma:

$$SQ_{quadrado} = y'(P_2 - P_1)y$$

$$SQ_{linha} = y'(P_3 - P_2)y$$

$$SQ_{coluna} = y'(P_4 - P_3)y$$

$$SQ_{tratamento} = y'(P_5 - P_4)y$$

$$SQ_{residuo} = y'(I_{(n)} - P)y$$

$$SQ_{total} = y'(I_{(n)} - P_1)y$$

sendo  $I_{(n)}$  uma matriz identidade de ordem  $n$ .

Isto é equivalente a utilizar uma inversa generalizada para a resolução do S.E.N.. No caso apresentado, as somas de quadrados são obtidas de maneira sequencial, conforme metodologia apresentada por (SEARLE, 1987).

Tabela 7 Modelos utilizados na obtenção das soma de quadrados tipo I das diversas causas de variação de um delineamento em QS  $k \times k$ , com suas respectivas somas de quadrados de parâmetros (SQP), equações normais (S.E.N.) e vetor de solução.

Modelo	SQP	S.E.N.	Vetor Solução
$y_{ijlm} = \mu + e_{ijlm}$	$R(\mu)$	$X'_1 X_1 \theta_1 = X'_1 y$	$\theta_1^0 = [\mu]$
$y_{ijlm} = \mu + \beta_i + e_{ijlm}$	$R(\beta \mu)$	$X'_2 X_2 \theta_2 = X'_2 y$	$\theta_2^{0'} = [\mu^0 \beta^0]$
$y_{ijlm} = \mu + \beta_i + \rho_j + e_{ijlm}$	$R(\rho \mu, \beta)$	$X'_3 X_3 \theta_3 = X'_3 y$	$\theta_3^{0'} = [\mu^0 \beta^0 \rho^0]$
$y_{ijlm} = \mu + \beta_i + \rho_j + \gamma_l + e_{ijlm}$	$R(\gamma \mu, \beta, \rho)$	$X'_4 X_4 \theta_4 = X'_4 y$	$\theta_4^{0'} = [\mu^0 \beta^0 \rho^0 \gamma^0]$
$y_{ijlm} = \mu + \beta_i + \rho_j + \gamma_l + \tau_m + e_{ijlm}$	$R(\tau \mu, \beta, \rho, \gamma)$	$X'_5 X_5 \theta_5 = X'_5 y$	$\theta_5^{0'} = [\mu^0 \beta^0 \rho^0 \gamma^0 \tau^0]$

As submatrizes envolvidas são:

$X_1$  : coluna 1 da matriz X de delineamento;

$X_2$  : coluna 1 até a  $i$ -ésima coluna do fator quadrado na matriz X de delineamento;

$X_3$  : coluna 1 até a  $j$ -ésima coluna do fator linha na matriz X de delineamento;

$X_4$  : coluna 1 até a  $l$ -ésima coluna do fator coluna na matriz X de delineamento;

$X_5$  : coluna 1 até a  $m$ -ésima coluna do fator tratamento na matriz X de delineamento.

Assim, as somas de quadrados tipo I são obtidas como segue:

$$R(1) = R(\mu) = (\theta_1^0)' \times (Xm)'y$$

$$R(2) = R(\beta|\mu) = (\theta_2^0)' \times (X_1)'y - R(1)$$

$$R(3) = R(\rho|\mu, \beta) = (\theta_3^0)' \times (X_2)'y - R(1) - R(2)$$

$$R(4) = R(\gamma|\mu, \beta, \rho) = (\theta_4^0)' \times (X_3)'y - R(1) - R(2) - R(3)$$

$$R(5) = R(\tau|\mu, \beta, \rho, \gamma) = (\theta_5^0)' \times (X_4)'y - R(1) - R(2) - R(3) - R(4)$$

em que  $R(2)$ ,  $R(3)$ ,  $R(4)$  e  $R(5)$ , representam as somas de quadrados de quadrado ( $\beta$ ), linha ( $\rho$ ), coluna ( $\gamma$ ) e tratamento ( $\tau$ ), respectivamente. O vetor de soluções de cada submatriz de equações normais é obtido por

$$\theta_1^0 = [(X_1)'X_1]^0 \times (X_1)'y$$

$$\theta_2^0 = [(X_2)'X_2]^0 \times (X_2)'y$$

$$\theta_3^0 = [(X_3)'X_3]^0 \times (X_3)'y$$

$$\theta_4^0 = [(X_4)'X_4]^0 \times (X_4)'y$$

$$\theta_5^0 = [(X_5)'X_5]^0 \times (X_5)'y$$

sendo que,  $[(X_i)'X_i]^0$ ,  $i = 1, 2, 3, 4, 5$ , são matrizes inversas generalizada de Moore-Penrose da matriz  $X_i'X_i$ , podendo ser obtida com o uso do programa “R” por meio da biblioteca “MASS” com a função “*ginv()*”. Qualquer outra inversa generalizada poderia ser utilizada. A notação  $R(i)$ , com  $i = 1, \dots, 5$  representa a redução na soma de quadrados total e de resíduos após o ajuste dos parâmetros no modelo, isto é:

$R(1)$  : redução na soma de quadrados total após ajustar  $\mu$ ;

$R(2)$  : redução na soma de quadrados do resíduo após ajustar  $\beta$  no modelo que contém  $\mu$ ;

$R(3)$  : redução na soma de quadrados do resíduo após ajustar  $\rho$  no modelo que contém  $\mu$  e  $\beta$ ;

$R(4)$  : redução na soma de quadrados do resíduo após ajustar  $\gamma$  no modelo que contém  $\mu$ ,  $\beta$  e  $\rho$ ;

$R(5)$  : redução na soma de quadrados do resíduo após ajustar  $\tau$  no modelo que contém  $\mu$ ,  $\beta$ ,  $\rho$  e  $\gamma$ .

A soma de quadrados total e de resíduos, são obtidas respectivamente por:

$$SQ_{total} = y'y - R_1$$

$$SQ_{residuo} = SQ_{total} - (R_2 + R_3 + R_4 + R_5)$$

### 2.2.9 Análise de Variância

Comumente na pesquisa científica, o pesquisador ao trabalhar com amostras de observações, se depara com variações inerentes aos seus dados, por motivos conhecidos ou não. A análise de variância (ANAVA) é um método paramétrico introduzido por *Fisher*, na década de 1920. O método funciona basicamente na decomposição da variação total das observações do experimento em partes dessa variação. Tais variações podem ser de causas conhecidas (fatores controlados) e causas desconhecidas (fatores não controlados), de natureza aleatória. O método permite a comparação de vários grupos através da estatística  $F$ , que é dada pela razão entre duas distribuições qui-quadrado independentes (SEARLE; CASELLA; MCCULLOCH, 1992).

Segundo (SEARLE; CASELLA; MCCULLOCH, 1992), a ANAVA foi desenvolvida inicialmente para considerar diferenças entre as médias de grupos, entretanto, mais tarde esta análise foi adaptada para estimar componentes da variância. Para realização da ANAVA, as seguintes pressuposições são consideradas:

- Aditividade dos efeitos - os efeitos devem ser aditivos;

- Independência dos erros - os erros ou resíduos devem ser não-correlacionados o que pode ser garantido na casualização dos tratamentos no experimento;
- Homogeneidade das variâncias dos erros - a variabilidade da observação de um tratamento deve ser semelhante aos demais tratamentos;
- Normalidade dos erros - os erros devem ter distribuição normal com média zero e variância comum, isto é,  $\epsilon_{ijlm} \sim N(0, \sigma_e^2)$

Dong e Gen (2008), em uma revisão sobre o uso de arranjos em delineamentos do tipo sudoku na experimentação de campo, deixaram passar um erro importante de análise ao não levar em conta a dependência linear que existia entre blocos  $\times$  linha e bloco  $\times$  coluna. O quadro-resumo da ANAVA apresentado é o mesmo do delineamento em quadrado Greco-latino, com  $k - 1$  graus de liberdade para linhas, colunas e quadrados (letras gregas). Isto é incorreto, pois no quadrado sudoku, quadrados não são ortogonais com linhas e colunas. Ao completar três quadrados em um sudoku  $9 \times 9$ , também completam-se três colunas, ao completar três linhas, também completam-se três quadrados, etc. Na tabela 8, são apresentados os graus de liberdades corretos (segundo a ordem especificada no modelo, ver Tabela 9) para análise de variância, sendo essa, um caso geral para quadrados sudokus perfeitos, isto é, quando se tem  $\sqrt{k} = p = q$ , em que a raiz quadrada é exata.

Tabela 8 Esquema de análise de variância para um delineamento sudoku  $k \times k$ , considerando modelo fixo e modelo aleatório.

<i>FV</i>	<i>GL</i>	<i>SQ</i>	<i>QM</i>	<i>E(QM)</i>	
				Fixo	Aleatório
Quadrado	$k - 1$	$SQ_q$	$QM_q$	$\sigma_e^2 + k \sum \beta_i^2 / (k - 1)$	$\sigma_e^2 + k\sigma_q^2$
Linha	$k - \sqrt{k}$	$SQ_l$	$QM_l$	$\sigma_e^2 + k \sum \rho_j^2 / (k - \sqrt{k})$	$\sigma_e^2 + k\sigma_l^2$
Coluna	$k - \sqrt{k}$	$SQ_c$	$QM_c$	$\sigma_e^2 + k \sum \gamma_l^2 / (k - \sqrt{k})$	$\sigma_e^2 + k\sigma_c^2$
Tratamento	$k - 1$	$SQ_t$	$QM_t$	$\sigma_e^2 + k \sum \tau_m^2 / (k - 1)$	$\sigma_e^2 + k\sigma_t^2$
Resíduo	$k(k - 4) + 2\sqrt{k} + 1$	$SQ_r$	$QM_r$	$\sigma_e^2$	$\sigma_e^2$
Total	$k^2 - 1$	$SQ_T$			

Para se obter os graus de liberdades na ANAVA de um delineamento em QS, os mesmos dependem da ordem dos fatores em que foram especificados na matrix  $X$  de delineamento. Como há quatro fatores (quadrado “ $\beta$ ”, linha “ $\rho$ ”, coluna “ $\gamma$ ” e tratamento “ $\tau$ ”), então o total de permutações (disposições) dos fatores é dado por  $P(n) = n!$ . Deste modo, existem  $P(4) = 4! = 24$  modelos diferentes para análise, entretanto, a escolha de um desses, dependerá da prioridade (importância) dos fatores a serem ordenados no modelo. Na Tabela 9, são apresentados os graus de liberdades segundo a ordem de especificação dos fatores.

Tabela 9 Tabela resumo referente aos Graus de Liberdade (GL) do modelo sudoku  $k \times k$ , segundo a ordem dos fatores especificados para os 24 modelos.

Ordem	Modelo								GL
	1	2	3	4	5	6	7	8	
1 <sup>a</sup>	$\beta$	$\beta$	$\rho$	$\gamma$	$\tau$	$\tau$	$\tau$	$\tau$	$k - 1$
2 <sup>a</sup>	$\tau$	$\tau$	$\tau$	$\tau$	$\gamma$	$\beta$	$\beta$	$\rho$	$k - 1$
3 <sup>a</sup>	$\rho$	$\gamma$	$\beta$	$\beta$	$\beta$	$\gamma$	$\rho$	$\beta$	$k - \sqrt{k}$
4 <sup>a</sup>	$\gamma$	$\rho$	$\gamma$	$\rho$	$\rho$	$\rho$	$\gamma$	$\gamma$	$k - \sqrt{k}$
	9	10	11	12	13	14			
1 <sup>a</sup>	$\rho$	$\rho$	$\gamma$	$\gamma$	$\tau$	$\tau$			$k - 1$
2 <sup>a</sup>	$\gamma$	$\tau$	$\rho$	$\tau$	$\gamma$	$\rho$			$k - 1$
3 <sup>a</sup>	$\tau$	$\gamma$	$\tau$	$\rho$	$\rho$	$\gamma$			$k - 1$
4 <sup>a</sup>	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$			$k - 2\sqrt{k} + 1$
	15	16	17	18					
1 <sup>a</sup>	$\beta$	$\beta$	$\rho$	$\gamma$					$k - 1$
2 <sup>a</sup>	$\rho$	$\gamma$	$\beta$	$\beta$					$k - \sqrt{k}$
3 <sup>a</sup>	$\tau$	$\tau$	$\tau$	$\tau$					$k - 1$
4 <sup>a</sup>	$\gamma$	$\rho$	$\gamma$	$\rho$					$k - \sqrt{k}$
	19	20	21	22					
1 <sup>a</sup>	$\beta$	$\beta$	$\rho$	$\gamma$					$k - 1$
2 <sup>a</sup>	$\rho$	$\gamma$	$\beta$	$\beta$					$k - \sqrt{k}$
3 <sup>a</sup>	$\gamma$	$\rho$	$\gamma$	$\rho$					$k - \sqrt{k}$
4 <sup>a</sup>	$\tau$	$\tau$	$\tau$	$\tau$					$k - 1$
	23	24							
1 <sup>a</sup>	$\rho$	$\gamma$							$k - 1$
2 <sup>a</sup>	$\gamma$	$\rho$							$k - 1$
3 <sup>a</sup>	$\beta$	$\beta$							$k - 2\sqrt{k} + 1$
4 <sup>a</sup>	$\tau$	$\tau$							$k - 1$

Os graus de liberdade total é dado por  $k^2 - 1$  e de resíduos por  $k(k - 4) + 2\sqrt{k} + 1$ .

### 2.3 Simulações usando o método Monte Carlo (MMC)

Freqüentemente, para avaliar um teste estatístico, é mais fácil obter informações numéricas do que analíticas sobre o poder e taxas de erro tipo I. Uma maneira de se obter as informações desejadas é por meio de simulações. O método Monte Carlo é um método utilizado para que sejam simulados números aleatórios de distribuição uniforme (0,1). A partir disto, geram-se amostras de outras distribuições, o que permite testar hipóteses estatísticas complexas e obter a distribuição de estimadores sob a lógica frequentista (MUN, 2006).

A origem do nome Monte Carlo está relacionado com a cidade de mesmo nome, no Principado de Mônaco, onde ocorriam jogos de azar. Na época, o dispositivo mecânico mais simples, para geração de números aleatórios, era a roleta (com distribuição uniforme discreta entre zero e 35). No entanto, o uso do método de Monte Carlo como técnica de simulação universal difundiu-se logo após do advento dos computadores (SOBOL, 1994).

Em síntese, qualquer processo simulado que envolva um componente aleatório de qualquer distribuição é considerado como pertencente ao método Monte Carlo (CARARI, 2004).

### 2.4 Tipos de erro e poder do teste

Na lógica frequentista dos problemas de experimentação, o interesse está em tomar a decisão de aceitar ou rejeitar determinada afirmação (em geral hipóteses sobre contrastes entre tratamentos), baseando-se em um conjunto de evidências. Um exemplo seria o problema de se decidir sobre a preferência ou não de certo tipo de iogurte a ser utilizado no comércio alimentício.

O pesquisador formula então as hipóteses “ $H_0$ : a preferência pelos io-

gurtas é a mesma” e “ $H_1$ : a preferência varia com os iogurtas”. Com base nos resultados de um experimento, o pesquisador terá então que se decidir por  $H_0$  ou  $H_1$ , e neste caso, poderá incorrer em dois tipos de erros probabilísticos, o erro tipo I, que ocorre com probabilidade  $\alpha$  e está associado a rejeitar uma  $H_0$  verdadeira e o erro tipo II, que ocorre com probabilidade  $\beta$  e está associado a aceitar uma  $H_0$  falsa.

Como exemplo, um pesquisador da área de ciências de alimentos poderia simplesmente administrar um mesmo tipo de iogurte de mesma marca para todos os provadores submetidos a uma análise sensorial, os quais receberiam, de maneira aleatória, o iogurte distribuído em diferentes recipientes codificados com as letras A, B, C, ..., etc. Neste caso, sabemos que a hipótese  $H_0$  é verdadeira e se ela for rejeitada na análise, então, incorreu-se no erro tipo I. A probabilidade de isso acontecer é  $\alpha$  e pode ser facilmente controlada ao estabelecer uma regra de decisão. Por outro lado, se fosse possível realizar o experimento com iogurtas com diferença de preferência conhecida, no caso de a hipótese  $H_0$  não ser rejeitada, incorre-se no erro tipo II. A probabilidade de isso acontecer é  $\beta$ , que é difícil de controlar em uma regra de decisão. A probabilidade de que seja rejeitada uma  $H_0$  falsa é denominada poder do teste ( $1 - \beta$ ). É claro que é desejável ter alta probabilidade de acertos e baixa probabilidade de erros, no entanto, a probabilidade de incorrer no erro do tipo I e do tipo II estão de tal forma associados, em que, ao reduzir a probabilidade de erro do tipo I (por ser a mais simples de controlar) aumenta-se a probabilidade de erro do tipo II. Para limitar  $\beta$ , procuramos aumentar o tamanho amostral (ou a informação sobre os contrastes de interesse) no experimento (MOOD; GRAYBILL; BOES, 1974). O controle do erro do tipo I, é feito por meio da fixação de  $\alpha$  em um “nível de significância” tolerável para o teste. Esta é a probabilidade máxima estabelecida pelo pesquisador antes da realização

da pesquisa, geralmente  $\alpha = 0,05$ . Ao se planejar o experimento, no entanto, só é possível diminuir  $\beta$  sem aumentar  $\alpha$ , pela utilização de tamanhos amostrais maiores ou reduzindo a “variância do erro” (qualquer variação na variável dependente  $y$ , causada por fatores desconhecidos), quando se identifica uma nova fonte de variação sobre  $y$  e a inclui no modelo.

Neste trabalho, será verificado se a utilização do sudoku como delineamento experimental controla bem a taxa de erro tipo I e se apresenta poder, quando comparado ao quadrado latino, tanto em situações em que o controle local adicional deve ser exercido (há efeito de quadrado na simulação), quanto em situações em que seria melhor utilizar o modelo do quadrado latino (não há efeito de quadrado na simulação).

### 3 METODOLOGIA

Neste trabalho, foram comparados delineamentos do tipo quadrado sudoku perfeito e quadrado latino nas dimensões  $4 \times 4$ ,  $9 \times 9$ ,  $16 \times 16$ ,  $25 \times 25$ ,  $36 \times 36$ ,  $49 \times 49$ ,  $64 \times 64$ ,  $81 \times 81$  e  $100 \times 100$ .

Consideraram-se duas diferentes situações de simulação. Na primeira situação, a simulação foi realizada sob a hipótese de que não havia efeito de quadrado no modelo teórico, isto é, considerou-se o modelo linear do quadrado latino. Na segunda situação, a simulação foi realizada sob a hipótese de que havia efeito de quadrado no modelo teórico, ou seja, considerou-se o modelo linear do quadrado sudoku.

Em ambas situações, a simulação visou detectar se o delineamento em sudoku perdia muita informação ao restringir a casualização sem necessidade.

As medidas tomadas para avaliar esta perda de informação foram o poder e o erro tipo I do teste F para tratamentos em ambos os modelos de análise.

Os dados simulados foram obtidos utilizando o software *R* (R Development Core Team, 2011) e a rotina utilizada encontra-se no apêndice. Foram gerados 2.000 experimentos para cada situação apresentada.

### 3.1 Modelo de Simulação

Foram adotados os modelos lineares descritos nas seguintes equações:

$$Y = M + Ll + Cc + Tt + e \quad (3.1)$$

$$Y = M + Qq + Ll + Cc + Tt + e \quad (3.2)$$

em que,

$Y$ : representa o vetor de observações de dimensões  $k^2 \times 1$ ;

$M$ : o vetor com uma constante comum a todas as observações de dimensões  $k^2 \times 1$ ;

$Q$ : a matrix de delineamento para o fator quadrado de dimensões  $k^2 \times (k - 1)$ ;

$L$ : a matrix de delineamento para o fator linha de dimensões  $k^2 \times (k - 1)$ ;

$C$ : a matrix de delineamento para o fator coluna de dimensões  $k^2 \times (k - 1)$ ;

$T$ : a matrix de delineamento para o fator tratamento de dimensões  $k^2 \times (k - 1)$ ;

$q$ : o vetor de efeitos para o fator quadrado de dimensões  $k \times 1$ ;

$l$ : o vetor de efeitos para o fator linha de dimensões  $k \times 1$ ;

$c$ : o vetor de efeitos para o fator coluna de dimensões  $k \times 1$ ;

$t$ : o vetor de efeitos para o fator tratamento de dimensões  $k \times 1$ ;

$e$ : o vetor de resíduos.

Os efeitos dos fatores foram simulados a partir de uma distribuição normal, sendo o valor paramétrico para variância das fontes de variação igual a 2 e

valor igual a 1 para variância do erro.

A simulação Monte Carlo foi realizada sob duas situações distintas e para cada caso simulado dois modelos de análise foram considerados. A primeira situação simulada não considera o quadrado (QL) como fonte de variação, enquanto que a segunda situação considera (QS). Os modelos analisados, em cada caso, referem-se ao delineamento em QL e ao delineamento em QS, os quais receberam o mesmo vetor com diferentes efeitos de tratamentos para verificação do poder do teste F. Para a verificação do erro tipo I, os modelos receberam um vetor com efeito nulo de tratamentos.

Para cada experimento simulado, calculou-se o teste F para tratamentos nos níveis de significância de 5% e de 1%. Em síntese, tem-se a seguinte estrutura para a simulação Monte Carlo:

Primeira situação simulada:

1.  $H_0 : QL$  (SEM efeito de quadrado)

1.1. Simulando COM efeito de tratamento (poder)

$$1.1.1. y = L + C + T + e \quad \longrightarrow \quad (QL)$$

$$1.1.2. y = Q + L + C + T + e \quad \longrightarrow \quad (QS)$$

1.2. Simulando SEM efeito de tratamento (erro tipo I)

$$1.2.1. y = L + C + T + e \quad \longrightarrow \quad (QL)$$

$$1.2.2. y = Q + L + C + T + e \quad \longrightarrow \quad (QS)$$

Segunda situação simulada:

2.  $H_0 : QS$  (COM efeito de quadrado)

2.1. Simulando COM efeito de tratamento (poder)

$$2.1.1. y = L + C + T + e \quad \longrightarrow \quad (QL)$$

$$2.1.2. y = Q + L + C + T + e \quad \longrightarrow \quad (QS)$$

## 2.2. Simulando SEM efeito de tratamento (erro tipo I)

$$2.2.1. y = L + C + T + e \quad \longrightarrow \quad (\text{QL})$$

$$2.2.2. y = Q + L + C + T + e \quad \longrightarrow \quad (\text{QS})$$

Para verificação do poder do teste F em ambos delineamentos, foram considerados os seguintes efeitos (erros padrões da média):  $1/8\sigma_{\bar{x}}$ ,  $1/4\sigma_{\bar{x}}$ ,  $1/2\sigma_{\bar{x}}$ ,  $1\sigma_{\bar{x}}$ ,  $2\sigma_{\bar{x}}$  e  $4\sigma_{\bar{x}}$ . O erro padrão da média, definido como sendo o desvio padrão da distribuição das médias amostrais de uma população, é dado por (FERREIRA, 2005):

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{k}} = \frac{\sigma}{\sqrt{k}} \quad (3.3)$$

em que,  $\sigma^2$  representa a variância do erro experimental e  $k$  o tamanho da amostra.

## 3.2 Modelos de Análise

O modelo de análise do quadrado latino e do quadrado sudoku são dados respectivamente por (2.1) e (2.3). A hipótese estatística a ser testada é a mesma para ambos os modelos, que é a hipótese de não existência dos efeitos de tratamentos sobre a variável resposta  $Y$ , sendo equivalente a testar a hipótese da igualdade de médias de tratamentos por meio do teste F na ANAVA, isto é:

$$\left\{ \begin{array}{l} H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0 \\ H_1 : \tau_1 \neq \tau_2 \neq \dots \neq \tau_k \neq 0 \end{array} \right. \implies \left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 = \dots = \mu_k = 0 \\ H_1 : \mu_1 \neq \mu_2 \neq \dots \neq \mu_k \neq 0 \end{array} \right.$$

Deste modo, cada modelo de análise tem sua hipótese testada através do teste F da ANAVA.

### 3.3 Erro tipo I e poder

O processo de simulação foi repetido 2.000 vezes para cada configuração (delineamento, tamanho, vetor de efeitos). Em cada repetição foi feita a análise do modelo QL e do modelo QS nos dois níveis de significância considerados.

As taxas de erro tipo I foram estimadas no processo de simulação para cada delineamento, sendo que a proporção de vezes em que a hipótese nula foi rejeitada (falsamente), foi comparada com o nível nominal de significância estabelecido. A taxa de erro tipo I assim estimada é dada pela proporção:

$$\hat{\alpha} = \frac{n^*}{N} \quad (3.4)$$

em que,  $n^*$  representa o número de resultados simulados significativos do teste F para tratamentos no modelo de análise, considerando os valores nominais descritos anteriormente, isto é,  $\alpha = 1\%$  e  $\alpha = 5\%$ ;  $N$  representa o número total de simulações realizadas.

Uma vez que essas taxas de erro tipo I foram estimadas utilizando simulações de Monte Carlo, elas não estão livres de erro (OLIVEIRA; FERREIRA, 2010). Foi utilizado o intervalo de confiança exato para a proporção de rejeições de  $H_0$ , tanto para o erro tipo I quanto para a geração de curvas de poder dos delineamentos. Foi verificado se os níveis nominais estabelecidos a priori pertenciam ao intervalo para o erro tipo I (testes de tamanho exato), ou se ficavam abaixo do estabelecido (testes rigorosos ou conservativos) ou acima do estabelecido (testes liberais).

O intervalo de confiança exato para proporção baseado na distribuição F e distribuição binomial, é dado por (FERREIRA, 2005):

$$IC_{1-\alpha}(p) : \left[ P_I = \frac{1}{1 + \frac{(n-y+1)F_{\alpha/2;v_1=2(n-y+1),v_2=2y}}{y}}; P_S = \frac{1}{1 + \frac{(n-y)}{(y+1)F_{\alpha/2;v_1=2(y+1),v_2=2(n-y)}}} \right]$$

em que,  $P_I$  é o limite inferior e  $P_S$  o limite superior e  $F_{\alpha/2}$  é o quantil superior da distribuição F com  $v_1$  e  $v_2$  graus de liberdade. Se  $y = 0$ , então  $P_I = 0$  e  $P_S$  é obtido pelo intervalo de confiança dado acima. Se  $y = n$ ,  $P_S = 1$  e  $P_I$  é obtido pelo intervalo de confiança exato apresentado anteriormente. Segundo Ferreira (2005), o termo exato é usado no contexto de que o valor real da significância é no máximo igual ao valor nominal de significância ( $\alpha$ ) devida a natureza discreta da distribuição binomial.

O teste F para tratamentos, sob o delineamento considerado, será classificado como conservativo se o intervalo de confiança apresentado estiver abaixo do nível nominal ( $\alpha$ ), será classificado como liberal se o intervalo de confiança estiver acima de  $\alpha$  e será classificado como sendo de tamanho *exato*, se  $\alpha$  estiver contido no intervalo de confiança ao nível de significância estabelecido.

### 3.4 Exemplo de Aplicação em Ciência dos Alimentos

Consideremos um experimento hipotético que consistiu em determinar a relação ótima entre conservantes a serem utilizados na composição de doce de goiaba associado a quatro graus de adição de açúcar. O fator A envolveu as seguintes doses de adição de conservantes: 0,6%, 0,8%, 1,0% e 1,2%. O fator B representa a porcentagem de adição de açúcar, nas doses a seguir: 30%, 40%, 50% e 60%.

Foi simulada uma superfície polinomial com base nestes níveis. A esta superfície foi superimposto um delineamento sudoku com efeitos de linha, coluna e quadrados e um erro experimental como os da seção anterior. A resposta contínua foi analisada segundo o modelo linear do sudoku.

O objetivo do estudo hipotético seria verificar qual a melhor relação entre conservantes e níveis de açúcar que fornece a maior predileção entre os provadores. Para controlar os efeitos de provadores, dias de prova e ocasiões de prova, arranjaram-se os tratamentos em um quadrado sudoku  $16 \times 16$  com linhas representando provadores, colunas representando dias de prova e quadrados representando ocasiões de prova. Na Figura 14, é apresentada a solução de um Sudoku  $16 \times 16$ , em que, os 16 valores distribuídos em cada quadrado representam os 16 tratamentos conforme a Tabela 10.

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

Figura 14 Quadrado Sudoku  $16 \times 16$

Por simplicidade de análise, as respostas simuladas foram obtidas utilizando uma escala não estruturada de notas, consistindo entre “0” (desgostei muito) e “5” (gostei muito). Para descrever o delineamento, considerou-se o modelo (2.3) com os fatores especificados na ordem: quadrado ( $\beta$ ), linha ( $\rho$ ), coluna ( $\gamma$ ) e tratamento ( $\tau$ ).

Tabela 10 Efeitos de tratamentos simulados segundo uma superfície de resposta.

Tratamento	Conservantes	Açúcar	Resposta simulada
1	0,6% (-15)	30% (-15)	0,0
2	0,6% (-15)	40% (-5)	1,5
3	0,6% (-15)	50% (5)	1,7
4	0,6% (-15)	60% (15)	0,6
5	0,8% (-5)	30% (-15)	2,7
6	0,8% (-5)	40% (-5)	4,3
7	0,8% (-5)	50% (5)	4,7
8	0,8% (-5)	60% (15)	3,7
9	1,0% (5)	30% (-15)	2,8
10	1,0% (5)	40% (-5)	4,5
11	1,0% (5)	50% (5)	5,0
12	1,0% (5)	60% (15)	4,1
13	1,2% (15)	30% (-15)	0,2
14	1,2% (15)	40% (-5)	2,1
15	1,2% (15)	50% (5)	2,7
16	1,2% (15)	60% (15)	1,9

Os dados completos simulados da Tabela 10, se encontram na Tabela 14 do apêndice dessa dissertação, como também o programa R utilizado nas análises.

Na Tabela 11, são apresentados os resultados da análise de variância do QS para o exemplo simulado. Nota-se que o efeito de quadrado foi significativo, implicando dizer que foi oportuna a utilização do QS como delineamento experimental. Para o ajuste da superfície de resposta, foi utilizada a função básica  $lm()$  do R. O resultado desta análise está apresentado na Tabela 14 do apêndice.

Tabela 11 Análise de Variância de regressão da superfície de resposta.

F.V	G.L	SQ	QM	F	P-valor
Ocasão ( $\beta$ )	15	636,06	42,40	43,65	2,2e-16 ***
Provador ( $\rho$ )	12	437,76	36,48	37,55	2,2e-16 ***
Dias ( $\gamma$ )	12	305,34	25,44	26,19	2,2e-16 ***
Tratamento ( $\tau$ )	(15)	598,45	39,90	41,00	2,2e-16 ***
<i>a</i>	1	14,73	14,73	15,16	0,0001 ***
<i>b</i>	1	55,67	55,67	57,31	1,1e-12 ***
<i>ab</i>	1	2,51	2,51	2,58	0,1097
<i>a</i> <sup>2</sup>	1	435,30	435,30	448,09	2,2e-16 ***
<i>b</i> <sup>2</sup>	1	80,84	80,84	83,22	2,2e-16 ***
<i>desvios</i>	10	9,4	0,94	0,97	0,4743
Resíduos	201	195,58	0,97		
<b>Total</b>	<b>255</b>	<b>2173,19</b>			

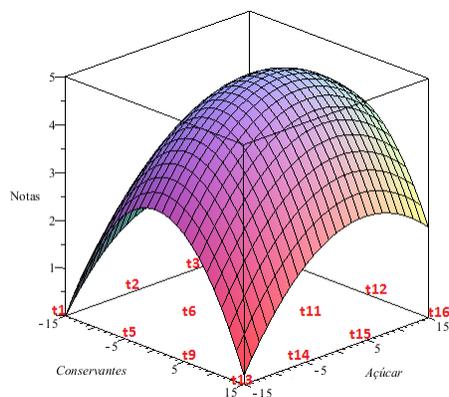


Figura 15 Superfície de resposta para notas de provadores, em função dos graus de conservantes e Açúcar

Para fins práticos, o processo importante é a interpretação dos efeitos de tratamentos. Neste sentido, pode-se observar na Figura 15, que a maior predileção dos provadores, está entre os níveis médios dos fatores conservantes e açúcar.

## 4 RESULTADOS

Os resultados serão apresentados em tabelas para o erro tipo I e em gráficos para as curvas de poder (que incluem, incidentalmente o erro tipo I quando os efeitos de tratamentos são nulos).

Na Tabela 12, são apresentadas as taxas de erro tipo I para os delineamentos em quadrado latino e quadrado sudoku, sob a hipótese de não haver efeito de quadrado no modelo de simulação, isto é,  $H_0 : QL$ .

Observa-se que para todas as situações (diferentes  $k$  e  $\alpha$ ) realizadas sob  $H_0 : QL$ , os QS têm erro tipo I ligeiramente maior do que os QL, mas, todos os intervalos de confiança indicam que ambos os delineamentos sempre apresentam tamanho exato.

Tabela 12 Taxa de erro tipo I para tratamentos em delineamentos do tipo Sudoku e Quadrado Latinos em função do número de dimensões  $k$  e níveis nominais de significância  $\alpha = 0,05$  e  $\alpha = 0,01$  sob  $H_0 : QL$ , considerando-se distribuição normal. Resultado de 2.000 simulações.

k	$\alpha = 0,05$		$\alpha = 0,01$	
	Q.Latino	Q.Sudoku	Q.Latino	Q.Sudoku
4	0,0480	0,0505	0,0065	0,0080
	[0,0391 ; 0,0583]	[0,0413 ; 0,0610]	[0,0028 ; 0,0127]	[0,0038 ; 0,0147]
9	0,0455	0,0490	0,0120	0,0120
	[0,0368 ; 0,0556]	[0,0400 ; 0,0594]	[0,0066 ; 0,0198]	[0,0066 ; 0,0198]
16	0,0520	0,0505	0,0105	0,0100
	[0,0427 ; 0,0627]	[0,0413 ; 0,0610]	[0,0055 ; 0,0179]	[0,0052 ; 0,0173]
25	0,0505	0,0510	0,0120	0,0110
	[0,0413 ; 0,0610]	[0,0418 ; 0,0616]	[0,0066 ; 0,0198]	[0,0059 ; 0,0185]
36	0,0435	0,0430	0,0080	0,0070
	[0,0350 ; 0,0534]	[0,0345 ; 0,0528]	[0,0038 ; 0,0147]	[0,0031 ; 0,0134]
49	0,0540	0,0530	0,0100	0,0110
	[0,0445 ; 0,0648]	[0,0436 ; 0,0637]	[0,0052 ; 0,0173]	[0,0059 ; 0,0185]
64	0,0540	0,0540	0,0115	0,0120
	[0,0445 ; 0,0648]	[0,0445 ; 0,0648]	[0,0063 ; 0,0192]	[0,0066 ; 0,0198]
81	0,0595	0,0595	0,0125	0,0125
	[0,0495 ; 0,0708]	[0,0495 ; 0,0708]	[0,0070 ; 0,0204]	[0,0070 ; 0,0204]
100	0,0455	0,0450	0,0085	0,0080
	[0,0368 ; 0,0556]	[0,0363 ; 0,0550]	[0,0041 ; 0,0153]	[0,0038 ; 0,0147]

Na Tabela 13, são apresentadas as taxas de erro tipo I para os delineamentos em quadrado latino e quadrado sudoku, sob a hipótese de haver efeito de quadrado no modelo de simulação, isto é,  $H_0 : QS$ .

Observa-se que para todas as situações (diferentes  $k$  e  $\alpha$ ) realizadas sob  $H_0 : QS$ , os QS foram de tamanho exato e os QL foram conservadores (sendo inferiores ao nível de significância nominal especificado).

Este resultado já era esperado, devido ao modelo de simulação ser o sudoku. Neste caso, a variância estimada do erro pelo QL é maior do que a esperada.

Com o aumento do tamanho dos delineamentos ( $k$ ), os modelos em QL ficam sendo cada vez mais rigorosos.

Tabela 13 Taxa de erro tipo I para tratamentos em delineamentos do tipo Sudoku e Quadrado Latinos em função do número de dimensões  $k$  e níveis nominais de significância  $\alpha = 0,05$  e  $\alpha = 0,01$  sob  $H_0 : QS$ , considerando-se distribuição normal. Resultado de 2.000 simulações.

k	$\alpha = 0,05$		$\alpha = 0,01$	
	Q.Latino	Q.Sudoku	Q.Latino	Q.Sudoku
4	0,0215 * [0,0156 ; 0,0289]	0,0500 [0,0409 ; 0,0605]	0,0035 * [0,0010 ; 0,0085]	0,0100 [0,0052 ; 0,0173]
9	0,0025 * [0,0008 ; 0,0058]	0,0400 * [0,0318 ; 0,0495]	0,0005 * [0,0000 ; 0,0037]	0,0065 [0,0028 ; 0,0127]
16	0,0000 * [0,0000 ; 0,0018]	0,0530 [0,0436 ; 0,0637]	0,0000 * [0,0000 ; 0,0026]	0,0160 [0,0097 ; 0,0247]
25	0,0000 * [0,0000 ; 0,0018]	0,0515 [0,0422 ; 0,0621]	0,0000 * [0,0000 ; 0,0026]	0,0105 [0,0055 ; 0,0179]
36	0,0000 * [0,0000 ; 0,0018]	0,0520 [0,0427 ; 0,0627]	0,0000 * [0,0000 ; 0,0026]	0,0100 [0,0052 ; 0,0173]
49	0,0000 * [0,0000 ; 0,0018]	0,0455 [0,0368 ; 0,0556]	0,0000 * [0,0000 ; 0,0026]	0,0085 [0,0041 ; 0,0153]
64	0,0000 * [0,0000 ; 0,0018]	0,0485 [0,0395 ; 0,0588]	0,0000 * [0,0000 ; 0,0026]	0,0060 [0,0025 ; 0,0120]
81	0,0000 * [0,0000 ; 0,0018]	0,0545 [0,0450 ; 0,0654]	0,0000 * [0,0000 ; 0,0026]	0,0115 [0,0063 ; 0,0192]
100	0,0000 * [0,0000 ; 0,0018]	0,0575 [0,0477 ; 0,0686]	0,0000 * [0,0000 ; 0,0026]	0,0105 [0,0055 ; 0,0179]

\* conservativo.

Nos gráficos das Figuras 16 à 24, estão apresentadas as curvas de poder do teste F para tratamentos sob as hipóteses  $H_0 : QL$  e  $H_0 : QS$  dos delineamentos

em QL e QS nos níveis nominais de 5% e 1% de significância, considerando distribuição normal, sendo o valor paramétrico para a variância das demais causas de variação igual a 2 e valor igual a 1 para a variância do erro. Então, sob a hipótese  $H_0 : QL$  (Gráficos A e B), verifica-se que, à medida em que aumentam as diferenças entre tratamentos para delineamentos pequenos (em especial, para  $k = 4$ ), o QL apresenta poder ligeiramente maior do que QS, mas, conforme aumenta o tamanho dos delineamentos, ambos se tornam equivalentes em poder. Por outro lado, quando a hipótese considerada é  $H_0 : QS$  (Gráficos C e D), observa-se que o QS é sempre superior ao QL para todo  $k$ .

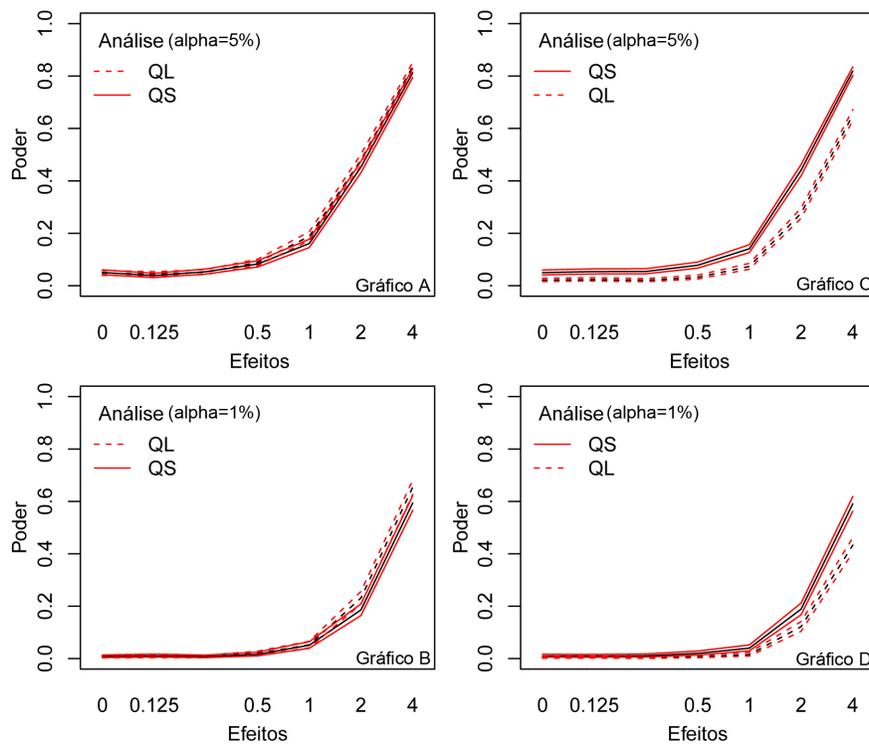


Figura 16 Poder do teste F para tratamentos sob  $H_0 : QL$  (Gráficos A e B) e sob  $H_0 : QS$  (Gráficos C e D) para as dimensões  $4 \times 4$  do QL e QS, conforme a ordem dos fatores  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  no modelo Sudoku

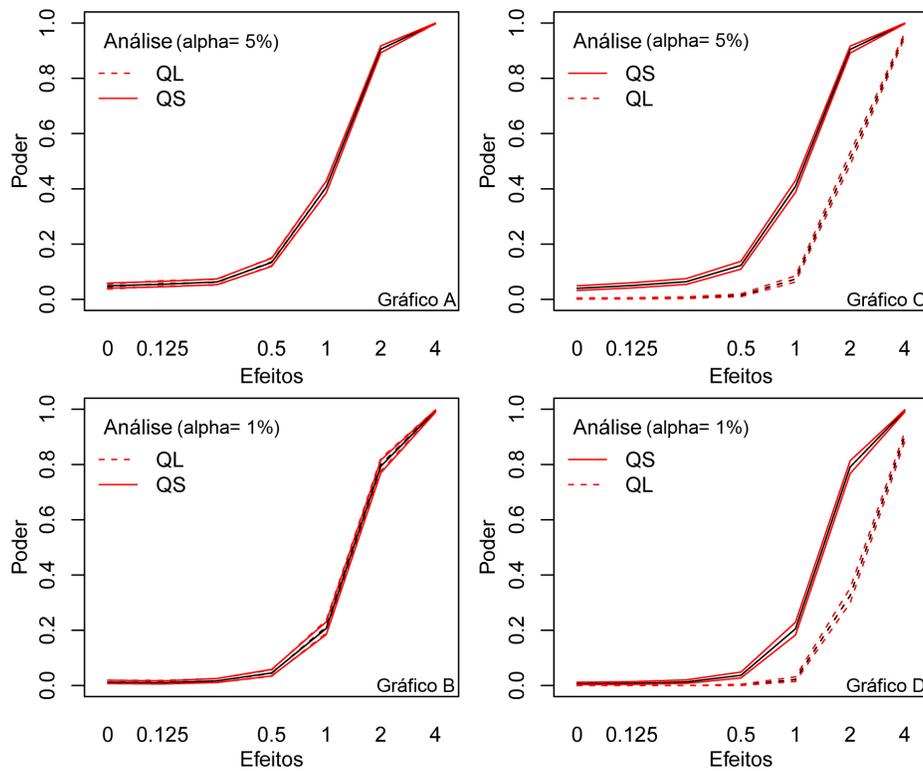


Figura 17 Poder do teste F para tratamentos sob  $H_0 : QL$  (Gráficos A e B) e sob  $H_0 : QS$  (Gráficos C e D) para as dimensões  $9 \times 9$  do QL e QS, conforme a ordem dos fatores  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  no modelo Sudoku

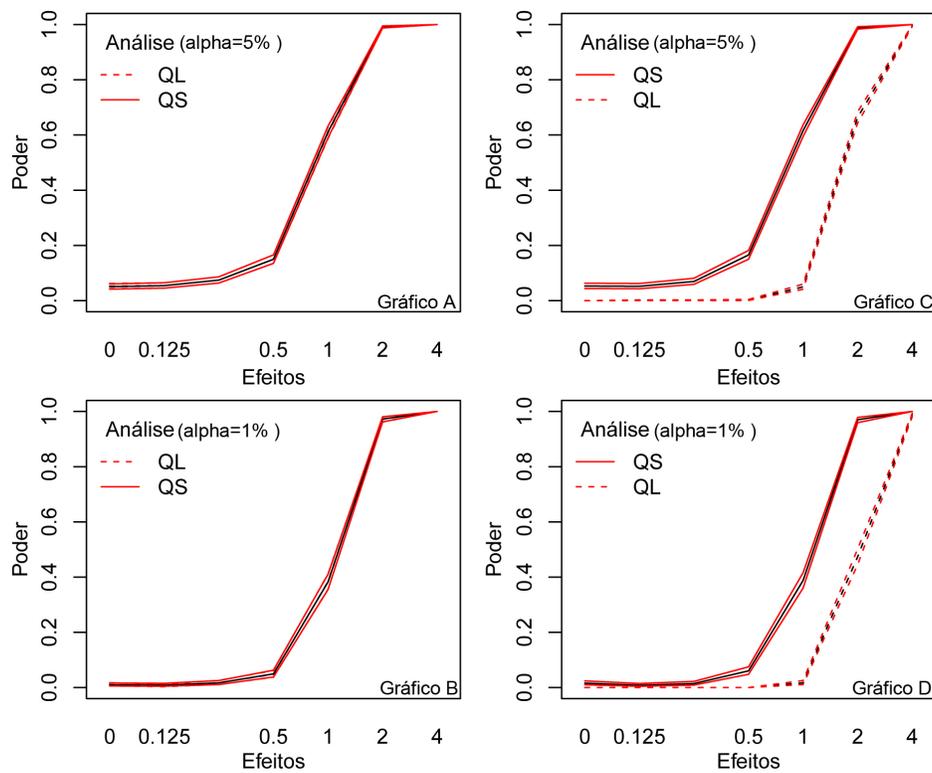


Figura 18 Poder do teste F para tratamentos sob  $H_0 : QL$  (Gráficos A e B) e sob  $H_0 : QS$  (Gráficos C e D) para as dimensões  $16 \times 16$  do QL e QS, conforme a ordem dos fatores  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  no modelo Sudoku

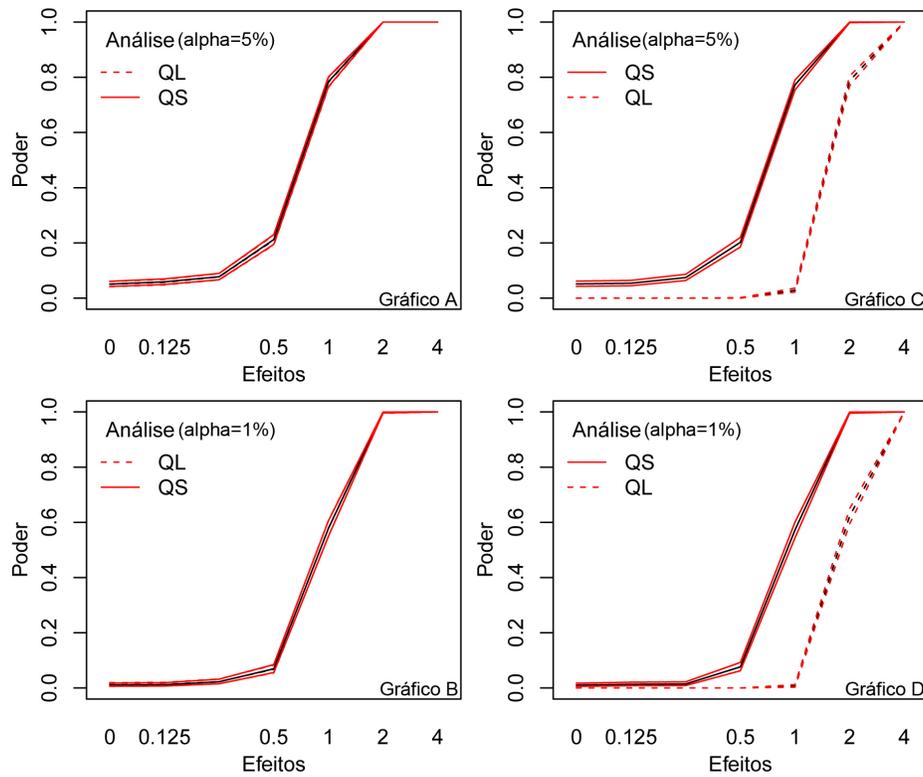


Figura 19 Poder do teste F para tratamentos sob  $H_0 : QL$  (Gráficos A e B) e sob  $H_0 : QS$  (Gráficos C e D) para as dimensões  $25 \times 25$  do QL e QS, conforme a ordem dos fatores  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  no modelo Sudoku

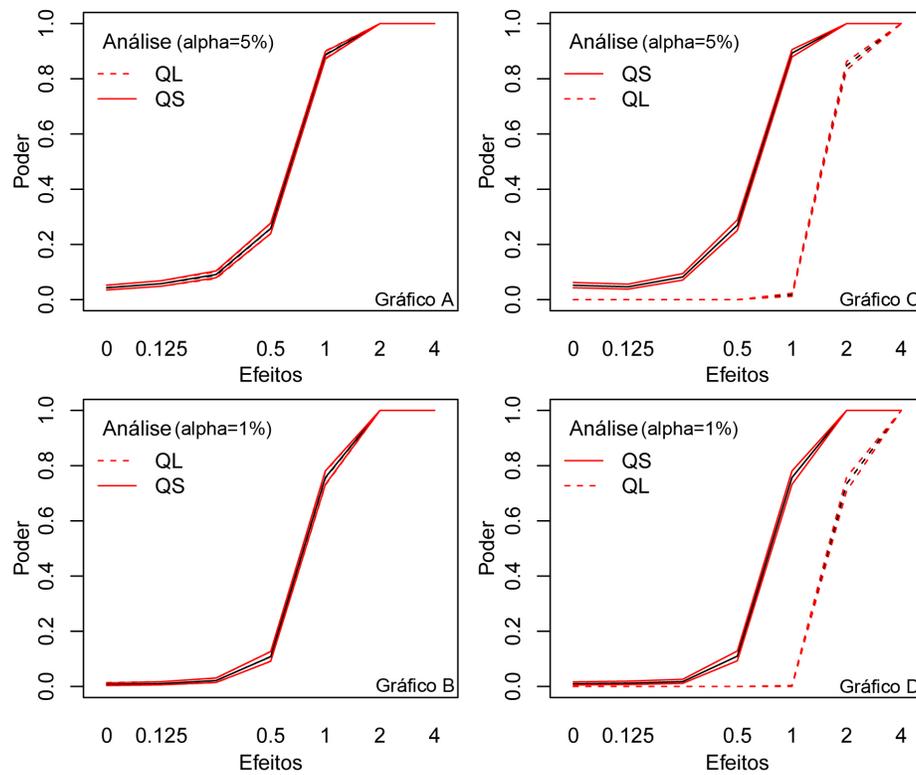


Figura 20 Poder do teste F para tratamentos sob  $H_0 : QL$  (Gráficos A e B) e sob  $H_0 : QS$  (Gráficos C e D) para as dimensões  $36 \times 36$  do QL e QS, conforme a ordem dos fatores  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  no modelo Sudoku

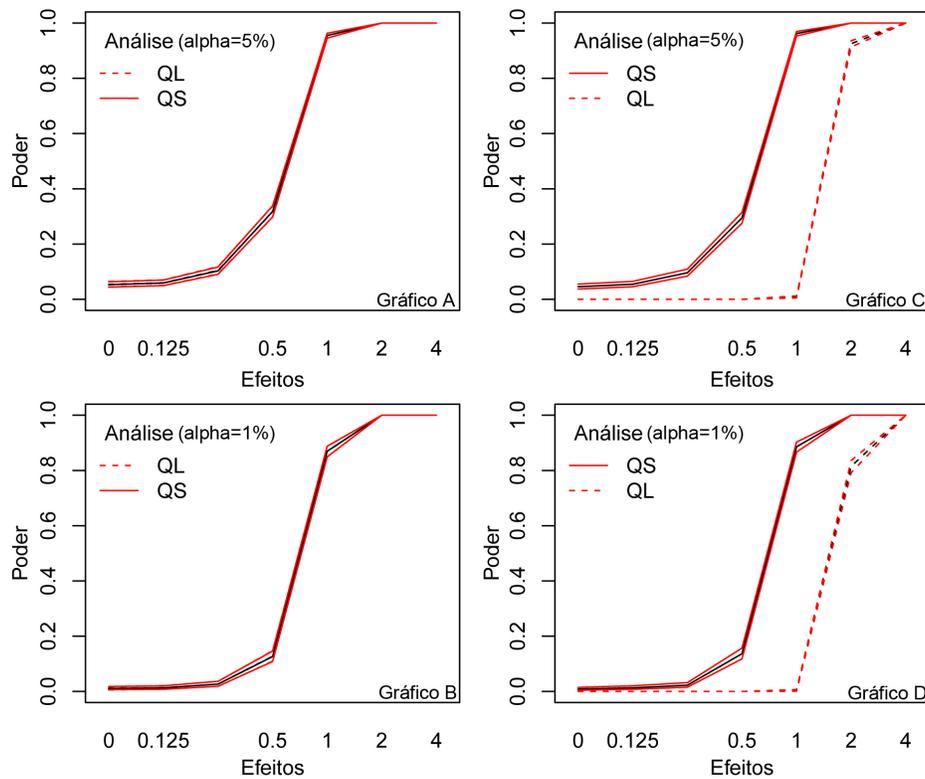


Figura 21 Poder do teste F para tratamentos sob  $H_0 : QL$  (Gráficos A e B) e sob  $H_0 : QS$  (Gráficos C e D) para as dimensões  $49 \times 49$  do QL e QS, conforme a ordem dos fatores  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  no modelo Sudoku

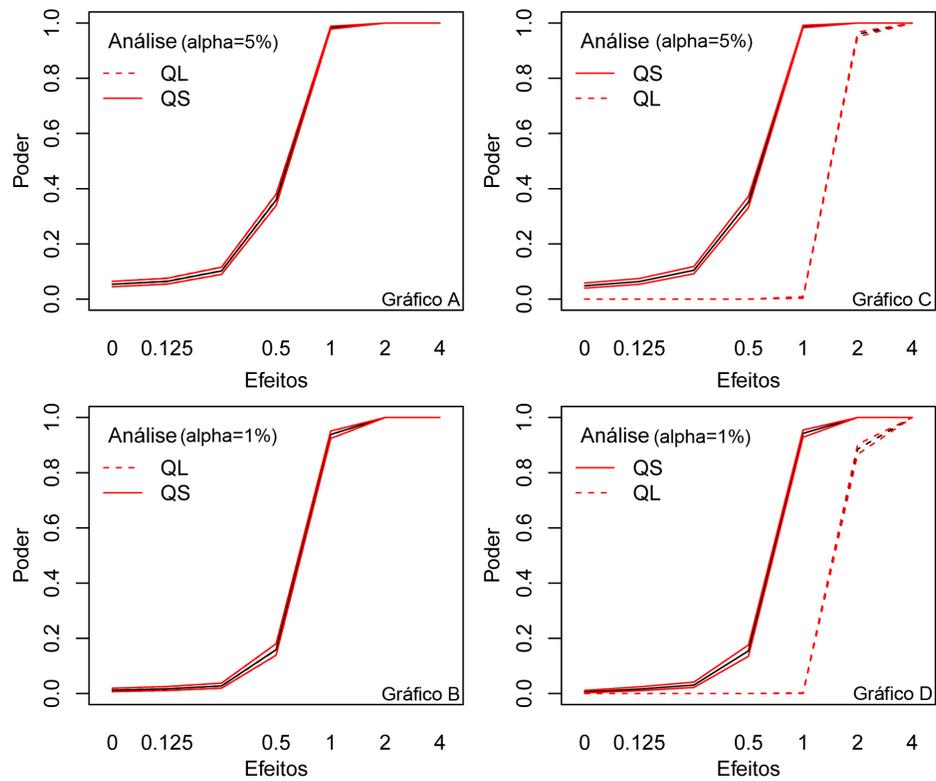


Figura 22 Poder do teste F para tratamentos sob  $H_0 : QL$  (Gráficos A e B) e sob  $H_0 : QS$  (Gráficos C e D) para as dimensões  $64 \times 64$  do QL e QS, conforme a ordem dos fatores  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  no modelo Sudoku

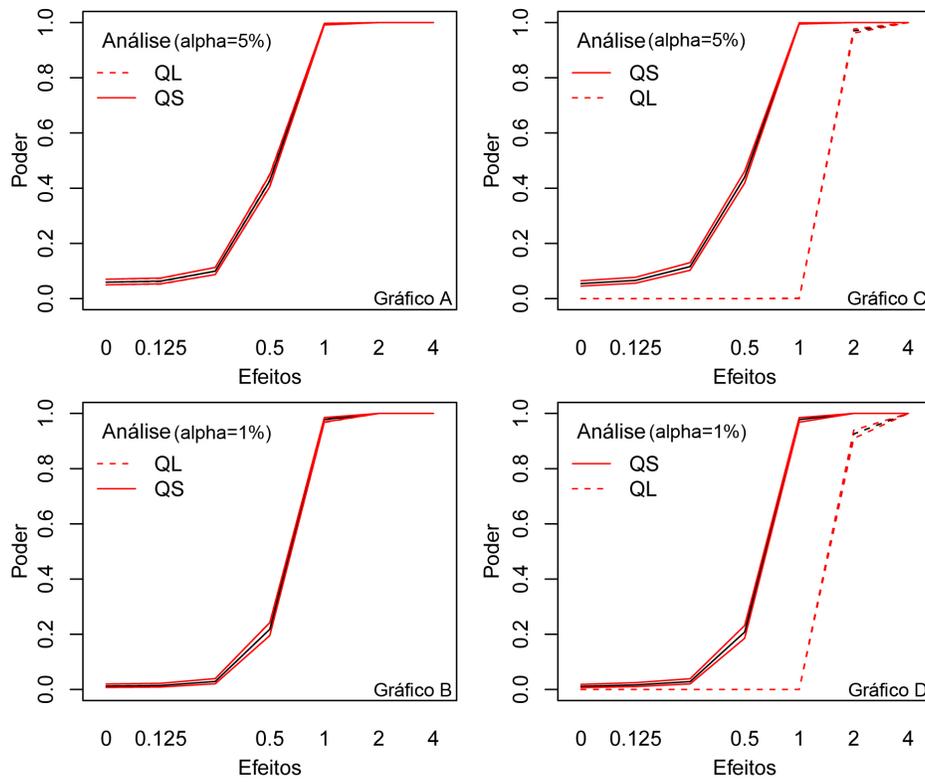


Figura 23 Poder do teste F para tratamentos sob  $H_0 : QL$  (Gráficos A e B) e sob  $H_0 : QS$  (Gráficos C e D) para as dimensões  $81 \times 81$  do QL e QS, conforme a ordem dos fatores  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  no modelo Sudoku

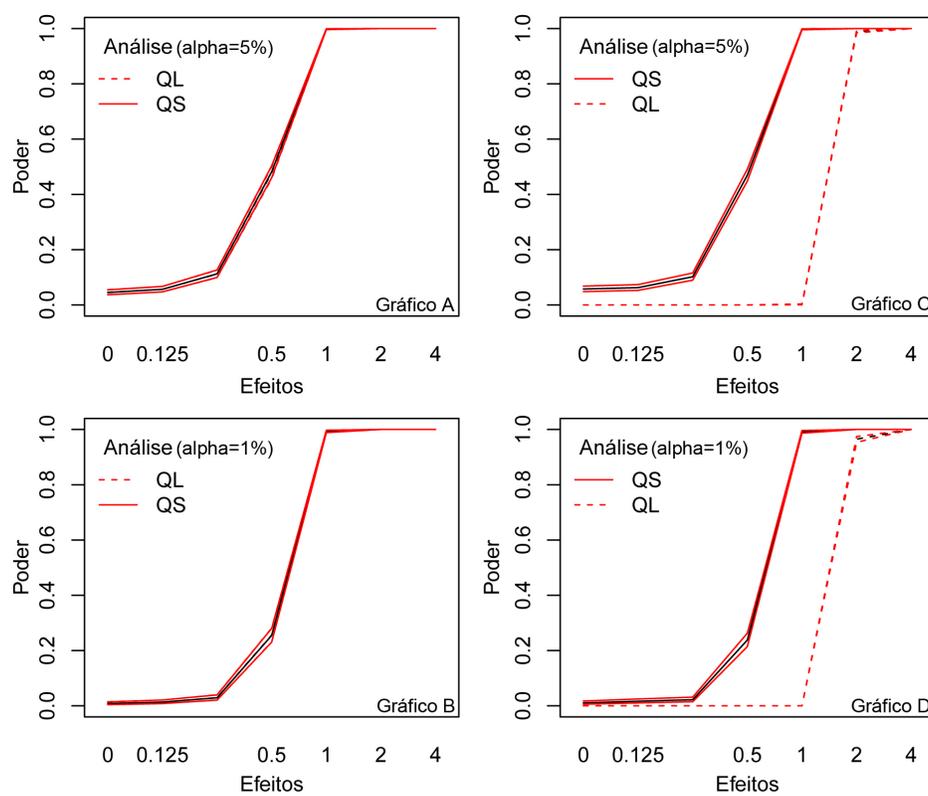


Figura 24 Poder do teste F para tratamentos sob  $H_0 : QL$  (Gráficos A e B) e sob  $H_0 : QS$  (Gráficos C e D) para as dimensões  $100 \times 100$  do QL e QS, conforme a ordem dos fatores  $\beta$ ,  $\rho$ ,  $\gamma$  e  $\tau$  no modelo Sudoku

## 5 DISCUSSÃO

Nas situações analisadas, em que o sudoku é o modelo adequado, as taxas de erro tipo I são rigorosas no delineamento em quadrado latino, como esperado. Não sabemos se isto chega a comprometer a escolha deste delineamento para experimentos grandes.

Para o caso de o delineamento simulado não ter efeito de quadrado, o sudoku mostrou-se surpreendentemente robusto. Provavelmente, isto significa que a restrição extra da aleatorização envolvida no sudoku é em geral desejável, sempre que se puder identificar fontes de variação associadas ao efeito de quadrado.

Quanto ao poder, no caso de o modelo de simulação ser o quadrado latino, as curvas de poder foram ligeiramente diferentes e apenas quando o delineamento era muito pequeno ( $k = 4$ ). Como o QS é um caso especial de QL, esperava-se que a curva de poder do QS fosse semelhante à curva de poder do QL, mas não igual, como ocorreu com delineamentos grandes.

No caso do modelo de simulação envolver efeito de quadrado, o sudoku foi bem mais poderoso que o quadrado latino.

Esse resultado é ótimo, porque pode-se planejar um experimento em QS sem a preocupação de perda de eficiência, caso não haja efeito associado ao quadrado. Ou seja, caso apenas se considere que possa existir efeito de quadrado, o sudoku é um ótimo delineamento para ser usado, com maior poder, e sem perder o rigor em relação aos QL de mesma dimensão.

Em suma, os sudokus preservam a taxa de erro tipo I, proporcionando testes de tamanho exato e em geral com maior poder do que o quadrado latino. Parece razoável recomendar que, sempre que possível, é preferível utilizar o sudoku ao QL.

Pode-se argumentar que uma desvantagem dos sudokus é que o número de repetições é muito grande para sudokus grandes. Este também é o caso dos quadrados latinos. No caso do quadrado latino, a solução, para experimentos com muitos tratamentos, é fazer delineamentos em que linhas (ou colunas) sejam incompletos. No caso dos sudokus, o mesmo poderia ser feito com a vantagem adicional de haver outra forma de controle local completa, por exemplo, colunas de quadrados.

Assim, um experimento com 6561 parcelas poderia ser reduzido para 729 parcelas, preservando dessa forma as suas propriedades gerais de controle local e análise.

Para os casos mais comuns relatados de quadrados latinos, haveria a alternativa de sudoku  $9 \times 9$  com 81 parcelas (ou subgrupos com 27 ou 54 parcelas).

Esta situação pode ser interessante se se quiser reduzir a fadiga em delineamentos em análise sensorial de alimentos em que não se dispunha de tempo e, ou, dinheiro para utilizar o sudoku completo.

Como exemplo, o pesquisador poderia planejar um sudoku  $81 \times 81$  e utilizar apenas nove dos 81 quadrados. Desta forma, poderia utilizar 81 pessoas. Ao invés de fazer com que, provem 81 chocolates, apenas nove seriam testados em cada ocasião de prova (dias diferente, por exemplo).

O número de dias que se pode repetir o experimento também não precisa ser nove (quadrado sudoku completo), mas poderia ser qualquer número menor, por exemplo, dois dias (1458 parcelas), com um sudoku completo para colunas e quadrados, mas incompleto para linhas, sendo no entanto, parcialmente balanceado e ainda bastante eficiente.

## 6 CONCLUSÃO

Sudokus são delineamentos poderosos e de tamanho exato quanto ao erro tipo I. Desta forma, podem e devem ser utilizados em delineamentos práticos.

Na análise sensorial de alimentos, há situações em que se podem identificar fontes de variação associáveis aos quadrados de um sudoku, nestas situações, tais delineamentos são uma alternativa promissora, podendo significar uma melhoria em relação aos delineamentos usuais em que muitos tratamentos são oferecidos em uma só ocasião de prova a cada provador.

## REFERÊNCIAS

- ABRIL. A história do Sudoku. Disponível em <[http://www.abril.com.br/noticia/diversao/no\\_168200.shtml](http://www.abril.com.br/noticia/diversao/no_168200.shtml)>. Acesso em: 6 maio, 2011.
- BAILEY, R. A.; CAMERON, P. J.; CONNELLY, R. Sudoku, gerechte designs, resolutions, affine space, spreads, reguli, and Hamming codes. **Amer. Math. Monthly**, 115 (2008), 383-403 p.
- BAILEY, R.A. **Design of Comparative experiments**. Cambridge University Press, 2008.
- BEHRENS, W. U. **Feldversuchsanordnungen mit verbessertem Ausgleich der Bodenunterschiede, Zeitschrift für Landwirtschaftliches Versuchswesen**. 2 (1956), 176-193.
- CARARI, M. L. **Intervalo de confiança para a diferença de duas proporções binomiais utilizando Bootstrap infinito**. 66 p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Lavras, 2004.
- DEAN, A.; VOSS, D. **Design and Analysis of Experiments**. 1st ed. Springer-Verlag, 1999.
- DIAZ, A.G.; PHILLIPS, D.T. **Principles of Experimental Design and Analysis**. Chapman & Hall, 1995.
- DONG, M.H.; GEN, X.R. Sudoku Square - a New Design in Field Experiment. **Acta Agronomica Sinica**. v. 34, Issue 9, p. 1489-1493, 2008.
- DRAKE D.A.; VAN REES, G.H.J.; WALLIS, W. D. Maximal sets of mutually orthogonal Latin squares. **Discrete Mathematics**. v.194 n.1-3, p.87-94, Jan. 6, 1999.

FELGENHAUER, B.; JARVIS, F. **Enumerating possible Sudoku grids**, 2005. Disponível em <<http://www.afjarvis.staff.shef.ac.uk/sudoku/sudoku.pdf>>. Acesso em: 2 junho, 2011.

FERREIRA, D.F. **Estatística Básica**. Lavras: Editora UFLA, 2005, 664p.

FISHER, R.A.; YATES, F. **Statistical Tables for Biological, Agricultural and Medical Research**. 6rd. Edition. London: Oliver and Boyd, 1963.

HINKELMANN, K; KEMPTHORNE, O. **Design and Analysis of Experiments**. Volume I: Introduction to Experimental Design (Second ed.). Wiley. ISBN 978-0-471-72756-9, 2008.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 5 Edition, John Wiley & Sons, New York, NY, 2001.

MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the Theory of Statistics** (3 edition). McGraw-Hill International, 1974.

MORAES, M.A.C. **Métodos para avaliação sensorial dos alimentos**. 8 edição, Campinas, SP. Editora da Unicamp, p.93, 1993.

MUN, J. **Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization Techniques**. John Wiley & Sons, Inc. 2006.

NATALE, C. D.; MACAGNANO, A.; PAOLESSE, R.; MANTINI, A.; TARIZZO, E. **Electronic nose and sensorial analysis: comparison of performances in selected cases**. Measurement, v. 50, p. 246-252, 1998.

NATIONAL RESEARCH COUNCIL. Appendix B: A Short History of Experimental Design, with Commentary for Operational Testing. Statistical Methods for Testing and Evaluating Defense Systems: Interim Report. Washington, DC: **The National Academies Press**, 1995.1. Print.

OLIVEIRA, I.R.C.; FERREIRA, D.F. Multivariate extension of chi-squared univariate normality test. **Journal of Statistical Computation and Simulation**, 80, 513-526, 2010.

PONNUSWAMY, K.N.; SUBRAMANI, J. Construction and analysis of sudoku designs. **Model Assisted Statistics and Applications**, DOI 10.3233/MAS-2009-0139, IOS Press, 2009.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2011. Disponível em <<http://www.r-project.org>>. Acesso em: 1 mar. 2011.

ROBINSON, G.K. That BLUP is a good thing: the estimation of random effects (with discussion). **Statistical Science**, 6, 15-51, 1991.

RUSSELL, E.; JARVIS F. **There are 5472730538 essentially different Sudoku grids**, 2005. Disponível em <<http://www.afjarvis.staff.shef.ac.uk/sudoku/sudgroup.html>>. Acesso em: 2 junho, 2011.

SEARLE, S.R. **Linear models for unbalanced data**. New York: John Wiley, 536p, 1987.

SEARLE, S.R., CASELLA, G.; MCCULLOCH, C.E. **Variance Components**. New York: John Wiley, 1992.

SOBOL, I.M. **A Primer for the Monte Carlo Method**. CRC Press LLC, (from the 4th ed. in Russian, 1985), 107 pp., 1994.

STONE, H.; SIDEL, J.L. **Sensory evaluation practices**. 3rd ed. London: Academic Press, Inc., 408p, 2004.

VILLAVIVENCIO, A. H.; ARAÚJO, M.M.; FANARO, G.B.; RELA, P.R.; RELA, J. M; MANCINI, F. J. **Sensorial analysis evaluation in cereal bars preserved by ionizing radiation processing: 1875-1877**, 2007.

## APÊNDICE

### Programas em linguagem R empregados na análise e simulação desta dissertação:

```
# Função que gera sudoku k*k, com k = p*q
gera.sudoku <- function(p,q)
{
  k <- p*q
  indice <- NULL
  sudoku <- matrix(0,k,k)
  for(i in 1:p)
  {
    for(j in 1:q)
    {
      indice <- (i-1)*p+j
      sudoku[indice,] <- 1:k +(j-1)*q + (i-1)
    }
  }
  sudoku <- sudoku%%k
  sudoku[which(sudoku==0)] <- k
  return(sudoku=sudoku)
}

# Função que sorteia sudoku k*k, com k = p*q

aleat.SDK <- function(p,q)
{
  grupo.coluna <- kronecker((sample(1:p)-1), rep(1,q))
  grupo.linha <- kronecker((sample(1:q)-1), rep(1,p))
  dentro.linha <- NULL
  dentro.coluna <- NULL
  for(i in 1:q)
```

```

        {
            dentro.linha <- c(dentro.linha,sample(1:p))
        }
    for(i in 1:p)
        {
            dentro.coluna <- c(dentro.coluna,sample(1:q))
        }
    sorteio.linha <- as.real(grupo.linha*p+dentro.linha)
    sorteio.coluna <- as.real(grupo.coluna*q+dentro.coluna)
    return(list(sorteio.linha=sorteio.linha,
               sorteio.coluna=sorteio.coluna))
}

# Utilizando as funções acima para gerar e aleatorizar um sudoku:

SDKpq <- function(p,q)
{
    sudoku <- gera.sudoku(p,q)
    sorteio <- aleat.SDK(p,q)
    sudoku <- sudoku[sorteio$sorteio.linha,]
    sudoku <- sudoku[,sorteio$sorteio.coluna]
    return(sudoku=sudoku)
}

# Grupos de linhas: p
# Grupos de colunas: q
# Linhas por grupo: q
# Colunas por grupo: p

p <- 3 # funciona para QS perfeitos, isto é, para p=q
q <- 3

```

```
sudoku <- SDKpq(p,q)

a <- rep(c(1:(p*q)), (p*q))
S1 <- sudoku
S2 <- S1
mascara <-sample(1:(p*q))

for(i in 1:(p*q))
{
  indice <- which(S1==mascara[i])
  S2[indice] <- i
}

sudoku9 <- S2

# Fontes de variação do Sudoku 9x9:

Quadrado <- 0*sudoku9
Quadrado[1:3,1:3] <- 1
Quadrado[1:3,4:6] <- 2
Quadrado[1:3,7:9] <- 3
Quadrado[4:6,1:3] <- 4
Quadrado[4:6,4:6] <- 5
Quadrado[4:6,7:9] <- 6
Quadrado[7:9,1:3] <- 7
Quadrado[7:9,4:6] <- 8
Quadrado[7:9,7:9] <- 9
Quadrado <- as.vector(Quadrado)

Linha <- kronecker(1:9,rep(1,9))
Coluna <- kronecker(rep(1,9),1:9)
```

```
Trat <- as.vector(sudoku9)

# Estrutura do delineamento:

design <- cbind(Quadrado,Linha,Coluna,Trat)

# Matriz X do delineamento:

Q <- factor(Quadrado)
L <- factor(Linha)
C <- factor(Coluna)
Trt <- factor(Trat)

k <- 9
Xm <- rep(1,k*k)          # coluna de 1's referente a média geral
XQ <- model.matrix(~-1+Q) # matrix de Quadrados
XL <- model.matrix(~-1+L) # matrix de Linhas
XC <- model.matrix(~-1+C) # matrix de Colunas
XT <- model.matrix(~-1+Trt) # matrix de Tratamentos
X <- cbind(Xm,XQ,XL,XC,XT) # matrix X do delineamento

Vq <- 2 # variância de Quadrado
Vl <- 2 # variância de Linha
Vc <- 2 # variância de Coluna
Vr <- 1 # variância do resíduo (definida como valor 1 no R)

# Primeira situação analisada: Simulação supondo H0:QL

simula.y.QL <- function(k,XL,XC,XT,ep)
{#função que recebe diferentes efeitos de tratamentos
  et <- ep*rnorm(k)      # efeito de tratamento
  lp <- sqrt(Vl)*rnorm(k) # efeito de linha
```

```

cp <- sqrt(Vc)*rnorm(k) # efeito de coluna
erro <- sqrt(Vr)*rnorm(k*k)
media <- c(rep(1,k*k))
y <- media + XL%*%lp + XC%*%cp + XT%*%et + erro
return(y=y)
}

saida <- matrix(0,7*Nsim,2)

for(j in 1:Nsim)
{
  for(i in 1:6)
  {
    ep <- (2^(i-1))/(8*sqrt(k))
    y <- simula.y.QL(k,XL,XC,XT,ep)#Simulando COM efeito de trat.
    saida[((j-1)*7+i),1] <- anova(lm(y~L+C+Trt))[3,5] # QL
    saida[((j-1)*7+i),2] <- anova(lm(y~Q+L+C+Trt))[4,5] # QS
  }
  y <- simula.y.QL(k,XL,XC,XT,0)#Simulando SEM efeito de trat.
  saida[(7*j),1] <- anova(lm(y~L+C+Trt))[3,5] # QL
  saida[(7*j),2] <- anova(lm(y~Q+L+C+Trt))[4,5] # QS
}

# Segunda situação analisada: Simulação supondo H0:QS

simula.y.QS <- function(k,XQ,XL,XC,XT,ep)
{#função que recebe diferentes efeitos de tratamentos
  et <- ep*rnorm(k) # efeito de tratamento
  qp <- sqrt(Vq)*rnorm(k) # efeito de quadrado
  lp <- sqrt(Vl)*rnorm(k) # efeito de linha
  cp <- sqrt(Vc)*rnorm(k) # efeito de coluna
  erro <- sqrt(Vr)*rnorm(k*k)

```

```

media <- c(rep(1,k*k))
y <- media + XQ%*%qp + XL%*%lp + XC%*%cp + XT%*%et + erro
return(y=y)
}

saida <- matrix(0,7*Nsim,2)

for(j in 1:Nsim)
{
  for(i in 1:6)
  {
    ep <- (2^(i-1))/(8*sqrt(k))
    y <- simula.y.QS(k,XQ,XL,XC,XT,ep)#Simulando COM efeito de trat.
    saida[((j-1)*7+i),1] <- anova(lm(y~L+C+Trt))[3,5] # QL
    saida[((j-1)*7+i),2] <- anova(lm(y~Q+L+C+Trt))[4,5] # QS
  }
  y <- simula.y.QS(k,XQ,XL,XC,XT,0)#Simulando SEM efeito de trat.
  saida[(7*j),1] <- anova(lm(y~L+C+Trt))[3,5] # QL
  saida[(7*j),2] <- anova(lm(y~Q+L+C+Trt))[4,5] # QS
}

# Resultados:

rej5 <- saida<0.05 # rejeições à 5%
rej1 <- saida<0.01 # rejeições à 1%

RQL5 <- matrix(0,Nsim,7)
RQS5 <- RQL5
RQL1 <- RQL5
RQS1 <- RQS5

for(i in 1:Nsim)

```

```

{
  for(j in 1:7)
  {
    RQL5[i, j] <- rej5[((i-1)*7+j), 1]
    RQS5[i, j] <- rej5[((i-1)*7+j), 2]
    RQL1[i, j] <- rej1[((i-1)*7+j), 1]
    RQS1[i, j] <- rej1[((i-1)*7+j), 2]
  }
}

NXSQL5 <- apply(RQL5, 2, sum)
NXSQS5 <- apply(RQS5, 2, sum)
NXSQL1 <- apply(RQL1, 2, sum)
NXSQS1 <- apply(RQS1, 2, sum)

ResultQS <- c(k, NXSQL5, NXSQS5, NXSQL1, NXSQS1)
save(ResultQS, file="ResultQS.rda")

```

### **Simulando o conjunto de dados para o exemplo de análise sensorial:**

# Definindo os tratamentos segundo uma superfície de resposta:

```

trat <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16)
A <- c(-15, -15, -15, -15, -5, -5, -5, -5, 5, 5, 5, 5, 15, 15, 15, 15)
B <- c(-15, -5, 5, 15, -15, -5, 5, 15, -15, -5, 5, 15, -15, -5, 5, 15)
Resp <- (390 + 2*A + 3*B - A^2 - 0.5*B^2 + 0.1*A*B) / 79

a <- 0*Trat
b <- 0*Trat
for(i in 1:16)
{
  a[which(T==i)] <- A[i]
  b[which(T==i)] <- B[i]
}

```

```

    }
    a;b

# Fontes de variação:

Quadrado <- 0*sudoku16 # obtido a partir de "gera.sudoku" (p=q=4)
Quadrado[1:4,1:4] <- 1
Quadrado[1:4,5:8] <- 2
Quadrado[1:4,9:12] <- 3
Quadrado[1:4,13:16] <- 4
Quadrado[5:8,1:4] <- 5
Quadrado[5:8,5:8] <- 6
Quadrado[5:8,9:12] <- 7
Quadrado[5:8,13:16] <- 8
Quadrado[9:12,1:4] <- 9
Quadrado[9:12,5:8] <- 10
Quadrado[9:12,9:12] <- 11
Quadrado[9:12,13:16] <- 12
Quadrado[13:16,1:4] <- 13
Quadrado[13:16,5:8] <- 14
Quadrado[13:16,9:12] <- 15
Quadrado[13:16,13:16] <- 16
Linha <- kronecker(1:16,rep(1,16))
Coluna <- kronecker(rep(1,16),1:16)
Trat <- as.vector(t(sudoku16))

# Matrix X do delineamento:

Q = factor(Quadrado)
L = factor(Linha)
C = factor(Coluna)
T = factor(Trat)

```

```
Xm <- rep(1,16*16)
XQ <- model.matrix(~-1+Q)
XL <- model.matrix(~-1+L)
XC <- model.matrix(~-1+C)
XT <- model.matrix(~-1+T)
X <- cbind(Xm,XQ,XL,XC,XT)

# Efeitos:

mp <- 1 # constante
qp <- sqrt(Vq)*rnorm(16) # efeito do quadrado
lp <- sqrt(Vl)*rnorm(16) # efeito da linha
cp <- sqrt(Vc)*rnorm(16) # efeito da coluna
ep <- rnorm(16*16) # efeito do erro

# Vetor de observações simulados segundo a superfície:

y <- mp + XQ%*%qp + XL%*%lp + XC%*%cp + XT%*%Resp+ ep
```

Tabela 14 Dados simulados para o exemplo da análise sensorial com ajuste da superfície de resposta.

Ordem	Quadrados	Linhas	Colunas	Tratamentos	Resposta observada
1	1	1	1	7	0,78
2	1	1	2	15	-0,52
3	1	1	3	14	0,02
4	1	1	4	10	3,42
5	5	1	5	3	-1,63
6	5	1	6	16	-2,82
7	5	1	7	4	-3,29
8	5	1	8	12	1,61
9	9	1	9	9	1,89
10	9	1	10	13	-3,26
11	9	1	11	5	1,98
12	9	1	12	2	2,85
13	13	1	13	11	2,90
14	13	1	14	8	1,29
15	13	1	15	6	3,56
16	13	1	16	1	-1,40
17	1	2	1	6	0,83
18	1	2	2	11	4,73
19	1	2	3	13	-0,02
20	1	2	4	1	0,03
21	5	2	5	5	1,02
22	5	2	6	14	-0,69
23	5	2	7	9	1,71
24	5	2	8	2	-1,63
25	9	2	9	16	-0,25
26	9	2	10	4	-0,24
27	9	2	11	15	1,59
28	9	2	12	8	8,15
29	13	2	13	10	3,03
30	13	2	14	7	2,06
31	13	2	15	3	1,75
32	13	2	16	12	5,56
33	1	3	1	16	0,42
34	1	3	2	12	2,46
35	1	3	3	5	0,90
36	1	3	4	8	4,39
37	5	3	5	11	4,66
38	5	3	6	13	-1,70
39	5	3	7	6	1,48
40	5	3	8	15	-0,42

Ordem	Quadrados	Linhas	Colunas	Tratamentos	Resposta observada
41	9	3	9	3	2,09
42	9	3	10	1	-1,53
43	9	3	11	10	5,35
44	9	3	12	7	5,18
45	13	3	13	9	3,90
46	13	3	14	4	1,65
47	13	3	15	2	1,58
48	13	3	16	14	1,85
49	1	4	1	3	0,29
50	1	4	2	4	-0,46
51	1	4	3	2	-0,12
52	1	4	4	9	3,11
53	5	4	5	8	2,51
54	5	4	6	7	1,13
55	5	4	7	10	2,01
56	5	4	8	1	-2,09
57	9	4	9	12	2,85
58	9	4	10	6	0,16
59	9	4	11	11	6,54
60	9	4	12	14	4,40
61	13	4	13	13	-0,06
62	13	4	14	5	1,29
63	13	4	15	15	0,29
64	13	4	16	16	3,56
65	2	5	1	2	-2,00
66	2	5	2	7	1,50
67	2	5	3	10	1,33
68	2	5	4	15	1,01
69	6	5	5	12	2,56
70	6	5	6	3	-1,53
71	6	5	7	5	0,86
72	6	5	8	6	0,56
73	10	5	9	4	-1,11
74	10	5	10	11	2,55
75	10	5	11	8	3,44
76	10	5	12	16	5,88
77	14	5	13	14	1,85
78	14	5	14	1	-1,43
79	14	5	15	13	-0,38
80	14	5	16	9	1,71

Ordem	Quadrados	Linhas	Colunas	Tratamentos	Resposta observada
81	2	6	1	9	-1,55
82	2	6	2	1	-4,60
83	2	6	3	12	-0,95
84	2	6	4	4	-1,73
85	6	6	5	2	-0,35
86	6	6	6	11	1,47
87	6	6	7	14	0,70
88	6	6	8	16	-2,85
89	10	6	9	7	4,00
90	10	6	10	5	-0,21
91	10	6	11	6	3,52
92	10	6	12	13	3,23
93	14	6	13	8	3,83
94	14	6	14	3	-0,56
95	14	6	15	10	4,63
96	14	6	16	15	3,34
97	2	7	1	5	4,09
98	2	7	2	14	1,31
99	2	7	3	3	2,90
100	2	7	4	6	6,15
101	6	7	5	9	4,14
102	6	7	6	1	0,14
103	6	7	7	13	0,53
104	6	7	8	8	6,04
105	10	7	9	10	7,45
106	10	7	10	15	3,09
107	10	7	11	2	2,38
108	10	7	12	12	10,27
109	14	7	13	16	3,16
110	14	7	14	11	7,11
111	14	7	15	4	2,16
112	14	7	16	7	8,40
113	2	8	1	13	-0,33
114	2	8	2	16	2,20
115	2	8	3	8	3,36
116	2	8	4	11	5,31
117	6	8	5	10	6,88
118	6	8	6	4	2,47
119	6	8	7	15	3,68
120	6	8	8	7	6,33

Ordem	Quadrados	Linhas	Colunas	Tratamentos	Resposta observada
121	10	8	9	1	0,93
122	10	8	10	14	4,04
123	10	8	11	9	6,88
124	10	8	12	3	7,56
125	14	8	13	6	7,37
126	14	8	14	12	6,36
127	14	8	15	5	4,15
128	14	8	16	2	5,76
129	3	9	1	15	1,22
130	3	9	2	13	-1,19
131	3	9	3	11	3,20
132	3	9	4	12	4,98
133	7	9	5	16	1,62
134	7	9	6	6	3,57
135	7	9	7	7	4,73
136	7	9	8	4	1,76
137	11	9	9	14	3,00
138	11	9	10	9	-0,29
139	11	9	11	3	-0,06
140	11	9	12	5	5,48
141	15	9	13	2	3,26
142	15	9	14	10	4,54
143	15	9	15	1	2,14
144	15	9	16	8	6,85
145	3	10	1	1	1,73
146	3	10	2	10	5,74
147	3	10	3	6	4,58
148	3	10	4	7	7,65
149	7	10	5	14	7,62
150	7	10	6	5	5,33
151	7	10	7	2	4,96
152	7	10	8	13	2,22
153	11	10	9	11	5,69
154	11	10	10	8	4,69
155	11	10	11	16	3,69
156	11	10	12	15	7,44
157	15	10	13	3	7,55
158	15	10	14	9	7,24
159	15	10	15	12	7,89
160	15	10	16	4	6,06

Ordem	Quadrados	Linhas	Colunas	Tratamentos	Resposta observada
161	3	11	1	14	1,09
162	3	11	2	5	2,01
163	3	11	3	9	1,28
164	3	11	4	3	1,19
165	7	11	5	15	3,79
166	7	11	6	8	2,06
167	7	11	7	12	4,44
168	7	11	8	11	5,34
169	11	11	9	2	1,94
170	11	11	10	10	2,07
171	11	11	11	1	1,12
172	11	11	12	4	2,09
173	15	11	13	7	8,76
174	15	11	14	6	5,08
175	15	11	15	16	3,96
176	15	11	16	13	1,09
177	3	12	1	8	2,71
178	3	12	2	2	-1,52
179	3	12	3	4	-0,70
180	3	12	4	16	1,63
181	7	12	5	1	3,31
182	7	12	6	10	1,67
183	7	12	7	3	3,21
184	7	12	8	9	1,53
185	11	12	9	13	-1,33
186	11	12	10	7	1,85
187	11	12	11	12	4,55
188	11	12	12	6	8,40
189	15	12	13	5	4,81
190	15	12	14	15	4,05
191	15	12	15	14	4,34
192	15	12	16	11	8,84
193	4	13	1	10	1,84
194	4	13	2	9	1,95
195	4	13	3	7	5,06
196	4	13	4	5	3,20
197	8	13	5	6	6,66
198	8	13	6	12	5,11
199	8	13	7	1	2,51
200	8	13	8	14	1,62

Ordem	Quadrados	Linhas	Colunas	Tratamentos	Resposta observada
201	12	13	9	15	3,84
202	12	13	10	2	2,33
203	12	13	11	13	3,98
204	12	13	12	11	11,02
205	16	13	13	4	4,83
206	16	13	14	16	5,12
207	16	13	15	8	7,61
208	16	13	16	3	4,90
209	4	14	1	11	6,64
210	4	14	2	3	3,54
211	4	14	3	16	4,56
212	4	14	4	14	4,40
213	8	14	5	13	3,19
214	8	14	6	9	4,02
215	8	14	7	8	4,96
216	8	14	8	10	5,89
217	12	14	9	6	8,52
218	12	14	10	12	5,09
219	12	14	11	4	4,66
220	12	14	12	1	7,49
221	16	14	13	15	9,70
222	16	14	14	2	4,73
223	16	14	15	7	10,10
224	16	14	16	5	8,72
225	4	15	1	12	1,20
226	4	15	2	6	1,96
227	4	15	3	15	1,29
228	4	15	4	13	-1,30
229	8	15	5	4	1,27
230	8	15	6	2	0,55
231	8	15	7	16	1,13
232	8	15	8	5	2,59
233	12	15	9	8	6,32
234	12	15	10	3	1,01
235	12	15	11	7	5,78
236	12	15	12	9	7,82
237	16	15	13	1	2,23
238	16	15	14	14	3,64
239	16	15	15	11	7,12
240	16	15	16	10	7,19

Ordem	Quadrados	Linhas	Colunas	Tratamentos	Resposta observada
241	4	16	1	4	-0,60
242	4	16	2	8	1,81
243	4	16	3	1	-3,75
244	4	16	4	2	0,89
245	8	16	5	7	1,40
246	8	16	6	15	0,22
247	8	16	7	11	2,94
248	8	16	8	3	-0,21
249	12	16	9	5	3,16
250	12	16	10	16	1,26
251	12	16	11	14	4,04
252	12	16	12	10	8,76
253	16	16	13	12	6,21
254	16	16	14	13	0,42
255	16	16	15	9	4,50
256	16	16	16	6	6,88

```
# Modelos analisados:
```

```
modelo <- lm(y ~ Q + L + C + T) # modelo completo para
anova(modelo) # os tratamentos.
```

```
modelo.SR <- lm(y ~ Q + L + C + I(a)+I(b)+I(a*b)+I(a*a)+I(b*b))
anova(modelo.SR) # superfície de resposta
```

```
anova(modelo.SR,modelo) # desvios da regressão.
```

```
# Respectivas saídas R:
```

```
Analysis of Variance Table
```

```
Response: y
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Q	15	636.06	42.404	43.580	< 2.2e-16 ***
L	12	437.76	36.480	37.492	< 2.2e-16 ***
C	12	305.34	25.445	26.151	< 2.2e-16 ***
T	15	598.45	39.896	41.003	< 2.2e-16 ***
Residuals	201	195.58	0.973		

## Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Q	15	636.06	42.40	43.6499	< 2.2e-16 ***
L	12	437.76	36.48	37.5524	< 2.2e-16 ***
C	12	305.34	25.44	26.1927	< 2.2e-16 ***
I(a)	1	14.73	14.73	15.1648	0.0001322 ***
I(b)	1	55.67	55.67	57.3054	1.145e-12 ***
I(a * b)	1	2.51	2.51	2.5795	0.1097494
I(a * a)	1	435.30	435.30	448.0879	< 2.2e-16 ***
I(b * b)	1	80.84	80.84	83.2183	< 2.2e-16 ***
Residuals	211	204.98	0.97		

# Análise dos desvios da regressão:

## Analysis of Variance Table

Model 1: y ~ Q + L + C + I(a)+I(b)+I(a\*b)+I(a\*a)+I(b\*b)

Model 2: y ~ Q + L + C + T

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	211	204.98				# desvios não
2	201	195.58	10	9.3996	0.966	0.4743 # significativos.