

Universidade Federal do Paraná

GUILHERME PARREIRA DA SILVA

ISADORA APARECIDA SPRENGOSKI DO NASCIMENTO

**GRÁFICOS DE CONTROLE PARA A MÉDIA
VIA *UNIFIED RANKED SET SAMPLING***

Curitiba

2015

GUILHERME PARREIRA DA SILVA
ISADORA APARECIDA SPRENGOSKI DO NASCIMENTO

**GRÁFICOS DE CONTROLE PARA A MÉDIA VIA
*UNIFIED RANKED SET SAMPLING***

Trabalho de Conclusão de Curso apresentado
à disciplina Laboratório de Estatística do
Curso de Graduação em Estatística da Uni-
versidade Federal do Paraná, como requisito
para elaboração do Trabalho de Conclusão
de Curso.

Orientador: Prof. Dr. Cesar Augusto Taconeli

Curitiba
2015

Agradecimentos

Agradecemos primeiramente a Deus pela capacidade que nos concedeu e por sempre estar nos guiando em mais uma etapa de nossas vidas.

Aos nossos pais, que muitas vezes abriram mão de seus sonhos em pró dos nossos, mesmo nos momentos de dificuldades estiveram ao nosso lado nos dando todo apoio necessário para seguir em frente e concluir esta etapa.

Aos nossos companheiros que caminham ao nosso lado de mãos dadas, nos apoiando nesta jornada e tiveram muita paciência, sabendo compreender os momentos difíceis.

Ao Professor Doutor Cesar Augusto Taconeli que foi nosso orientador, pelos ensinamentos e estar nos acompanhando não somente ao longo deste trabalho e sim por todo período acadêmico, compartilhando seu conhecimento e acreditando em nossa capacidade.

Aos nossos familiares e amigos que estiveram sempre torcendo por nós, ansiosos por essa conquista e pela compressão nos momentos de ausências.

À Professora Doutora Suely Ruiz Giolo e ao Professor Doutor Gustavo Valentim Loch pela disponibilidade em aceitarem o convite de participar da banca deste trabalho, além de dividirem seus conhecimentos.

Por fim agradecemos um ao outro, pelos aprendizados, respeito, companheirismo e compreensão necessárias para realização deste trabalho.

*"Dar o melhor de si,
é mais importante que ser o melhor".*

Mike Lermer

Resumo

O uso de amostras ao invés de uma população é uma prática comum em diversas áreas do conhecimento. Uma das formas de obtê-las é via Amostragem Aleatória Simples (AAS), que é um delineamento com larga aplicação em diferentes contextos. No entanto, existem outras formas de se realizar a amostragem, sendo uma delas, a Amostragem por Conjuntos Ordenados (ACO). Esse tipo de amostragem é útil em situações em que a variável de interesse é cara ou de difícil mensuração, mas é possível ordenar as unidades amostrais, segundo a variável de interesse com base em alguma variável concomitante, ou ainda, por algum julgamento pessoal, que sejam acessíveis e de baixo custo. Existem vários delineamentos que foram criados baseados no delineamento original da ACO, incluindo o *Extreme Ranked Set Sampling (ERSS)*, *Median Ranked Set Sampling (MRSS)*, além do *Unified Ranked Set Sampling (URSS)*. A aplicação de tais delineamentos é discutida e apresentada, como por exemplo, no contexto de Controle Estatístico de Processos. Tais delineamentos podem ser úteis, por exemplo, para a construção de gráficos de controle para média, dada a maior eficiência em relação à AAS e consequente detecção de uma alteração na média do processo mais rapidamente. Este trabalho, tem como objetivo, comparar o desempenho do *URSS* em relação aos demais delineamentos baseados em conjuntos ordenados e à AAS na construção de gráficos de controle para a média, através de um estudo por simulação em que foram considerados diferentes tamanhos de amostras e níveis de correlação entre a variável de interesse e uma variável concomitante na identificação da alteração na média do processo. Também, é feita uma simulação de aplicação do *URSS*, ACO e AAS em dados reais. Ainda, é proposto um estimador para a variância da média via *URSS*. Os resultados revelaram que o *URSS* foi o delineamento que teve melhor desempenho na construção de gráficos de controle para a média, e em média, proporcionou maior rapidez na detecção da alteração da média do processo, segundo o estudo por simulação. Sendo que o *URSS* foi mais eficiente em 190 dos 210 cenários considerados em relação aos delineamentos baseados em conjuntos ordenados e em praticamente todos os cenários se comparados à AAS. Já na aplicação em dados reais, o *URSS* também acusou mais rapidamente um descontrole em 7 dos 8 cenários considerados. Portanto, sugere-se o delineamento *URSS* para a construção de gráficos de controle para a média. Ademais, verificou-se vício desprezível para o estimador proposto para a variância da média via *URSS*.

Palavras-chaves: Unified Ranked Set Sampling (*URSS*). Amostragem por Conjuntos Ordenados (ACO). Comprimento Médio de Sequência (CMS). Gráficos de controle para a média. Estudo por simulação.

Lista de abreviaturas e siglas

UFPR	Universidade Federal do Paraná
CMS	Comprimento Médio da Sequência
URSS	<i>Unified Ranked Set Sampling</i>
MRSS	<i>Median Ranked Set Sampling</i>
ERSS	<i>Extremed Ranked Set Sampling</i>
ACO	Amostragem por Conjuntos Ordenados
RSS	<i>Ranked Set Sampling</i>
AAS	Amostragem Aleatória Simples
LC	Límite de Controle
LIC	Límite Inferior de Controle
LSC	Límite Superior de Controle
DQRSS	<i>Double Quartile-Ranked Set Sampling</i>
QDRSS	<i>Quartile Double-Ranked Set Sampling</i>
DERSS	<i>Double Extreme-Ranked Set Sampling</i>
EDRSS	<i>Extreme Double-Ranked Set sampling</i>

Listas de ilustrações

Figura 1 – Representação da seleção de uma amostra de tamanho $k = 5$ via ACO	16
Figura 2 – Representação da seleção de uma amostra por conjuntos ordenados de tamanho 9, com 3 ciclos.	17
Figura 3 – Representação da seleção de uma amostra de tamanhos $k = 4$ e 5 via MRSS	19
Figura 4 – Representação da seleção de uma amostra de tamanhos $k = 4$ e 5 via ERSS	21
Figura 5 – Representação de um gráfico de controle para a média, para um processo sob controle.	29
Figura 6 – Representação de um gráfico de controle para a média, considerando um processo fora de controle.	30
Figura 7 – Representação da mensuração da resistência do concreto.	38
Figura 8 – Gráficos descritivos para a variável resistência do concreto	51
Figura 9 – Gráficos de controle para a variável resistência do concreto com $k = 3$ e processo sob controle ($\delta = 0$)	52
Figura 10 – Gráficos de controle para a variável resistência do concreto com $k = 3$ e $\delta = 1, 2$	53
Figura 11 – Gráficos de controle para a variável resistência do concreto com $k = 3$ e $\delta = 1, 2$	54
Figura 12 – Gráficos de controle para a variável resistência do concreto com $k = 5$ e processo sob controle ($\delta = 0$).	55
Figura 13 – Gráficos de controle para a variável resistência do concreto com $k = 5$ e $\delta = 0, 8$	56
Figura 14 – Gráficos de controle para a variável resistência do concreto com $k = 5$ e $\delta = 1, 2$	57

Listas de tabelas

Tabela 1 – Representação das posições das unidades amostrais via <i>URSS</i> , a serem selecionadas de uma amostra de tamanho k^2 , após ordenação.	24
Tabela 2 – Valores do CMS para gráficos de controle para a média, para amostras de tamanho $k=3$, utilizando AAS, ACO, <i>ERSS</i> , <i>DERSS</i> , <i>EDRSS</i> , <i>DQRSS</i> , <i>QDRSS</i> , <i>MRSS</i> e <i>URSS</i> , considerando ordenação perfeita	41
Tabela 3 – Valores do CMS para gráficos de controle para a média, para amostras de tamanho $k=4$, utilizando AAS, ACO, <i>ERSS</i> , <i>DERSS</i> , <i>EDRSS</i> , <i>DQRSS</i> , <i>QDRSS</i> , <i>MRSS</i> e <i>URSS</i> , considerando ordenação perfeita	41
Tabela 4 – Valores do CMS para gráficos de controle para a média, para amostras de tamanho $k=5$, utilizando AAS, ACO, <i>ERSS</i> , <i>DERSS</i> , <i>EDRSS</i> , <i>DQRSS</i> , <i>QDRSS</i> , <i>MRSS</i> e <i>URSS</i> , considerando ordenação perfeita	42
Tabela 5 – CMS para gráficos de controle para a média baseados em amostras de tamanho $k = 3$, obtidos via simulação sob ordenação imperfeita.	45
Tabela 6 – CMS para gráficos de controle para a média baseados em amostras de tamanho $k = 4$, obtidos via simulação sob ordenação imperfeita.	46
Tabela 7 – CMS para gráficos de controle para a média baseados em amostras de tamanho $k = 5$, obtidos via simulação sob ordenação imperfeita.	47
Tabela 8 – CMS para gráficos de controle para a média baseados em amostras de tamanho $k = 6$, obtidos via simulação sob ordenação imperfeita.	48
Tabela 9 – Medidas resumo para as variáveis quantidade de cimento e resistência do concreto	50

Sumário

1	INTRODUÇÃO	10
1.1	Problema e Justificativa	13
1.1.1	Objetivo Geral	13
1.1.2	Objetivos Específicos	14
2	MATERIAL E MÉTODOS	15
2.1	Delineamentos Amostrais por Conjuntos Ordenados	15
2.1.1	Amostragem por Conjuntos Ordenados	15
2.1.2	<i>Median Ranked Set Sampling</i>	18
2.1.3	<i>Extreme Ranked Set Sampling</i>	20
2.1.4	<i>Double-Ranked Set Sampling</i>	22
2.1.5	<i>Unified Ranked Set Sampling</i>	22
2.2	Estatísticas de Ordem	24
2.2.1	Definição	25
2.2.2	Distribuição Marginal	25
2.2.3	Distribuição conjunta	26
2.2.4	Momentos e Momentos Produtos	26
2.3	Controle Estatístico de Processos	27
2.3.1	Gráficos de Controle para a Média do Processo	28
2.3.2	Comprimento Médio da Sequência	31
2.4	Metodologia do Estudo	32
2.4.1	Proposta para a Utilização do <i>URSS</i> para Gráficos de Controle	32
2.4.2	Estudo Via Simulação do Desempenho dos Gráficos de Controle para a Média Via <i>URSS</i>	32
2.4.3	Obtenção de $Var(\bar{Y}_{URSS})$	34
2.4.3.1	Obtenção de $Var(\bar{Y}_{URSS})$ sob Ordenação Perfeita	34
2.4.3.2	Obtenção de $Var(\bar{Y}_{URSS})$ sob Ordenação Imperfeita	34
2.4.4	Proposta e Estudo do Viés de um Estimador de $Var(\bar{Y}_{URSS})$ Aplicado à Construção de Gráficos de Controle	35
2.5	Materiais	37
2.5.1	Conjunto de Dados	37
2.5.2	Descrição da Aplicação do <i>URSS</i> , ACO e a AAS no Conjunto de Dados	38
2.5.3	Recursos Computacionais	39
3	RESULTADOS E DISCUSSÃO	40

3.1	Desempenho e Comparação dos Gráficos de Controle entre os De-lineamentos Apresentados	40
3.2	Análise do Viés do Estimador Proposto para a Variância da Média via URSS	49
3.3	Simulação de Aplicação em Dados Reais	49
4	CONCLUSÕES	59
	REFERÊNCIAS	61
	APÊNDICES	64
	APÊNDICE A – SELEÇÃO DE UMA AMOSTRA URSS	65
	APÊNDICE B – ESTATÍSTICAS DE ORDEM	67
	APÊNDICE C – ESTIMAÇÃO DA $Var(URSS)$ PARA ORDENAÇÃO IMPERFEITA	70
	APÊNDICE D – ESTIMAÇÃO DE $Var(URSS)$ PARA ORDENAÇÃO PERFEITA	72
	APÊNDICE E – OBTENDO CMS PARA ORDENAÇÃO IMPERFEITA	74
	APÊNDICE F – OBTENDO CMS PARA ORDENAÇÃO PERFEITA	77
	APÊNDICE G – ESTUDO SOBRE O VIÉS DO ESTIMADOR PROPOSTO PARA $Var(URSS)$	79
	APÊNDICE H – ELABORAÇÃO DAS CARTAS CONTROLE EM UM CONJUNTO DE DADOS	81
	APÊNDICE I – OBTENÇÃO DO CMS PARA MRSS	90

1 INTRODUÇÃO

Para se estudar uma população de interesse na maioria das vezes existem limitações quanto a considerar (ou incluir) todos os elementos, por isso geralmente trabalha-se com amostragem. A seleção de uma amostra exige cuidado, pois tal amostra deve ser representativa da população, para que os resultados amostrais reflitam adequadamente a população, sendo possível produzir inferências. A Amostragem Aleatória Simples (AAS), por exemplo, é um método usual para a seleção de amostras, servindo de base também para delineamentos amostrais mais elaborados, como amostragem em dois ou mais estágios ([BOLFARINE; BUSSAB, 2005](#)). A AAS é usualmente aplicada, por exemplo, na extração de unidades amostrais de linhas de produção, a fim de realizar o monitoramento de parâmetros de processos industriais considerando alguma variável de interesse. Porém, em algumas situações, podem existir dificuldades quanto a seleção de grandes amostras, pois as unidades amostrais podem ser pouco acessíveis ou a mensuração da variável de interesse envolver altos custos. Para contornar isso, busca-se por delineamentos amostrais com maior eficiência para amostras pequenas, o que pode remeter em redução de custos e tempo ([BARRETO, 2001](#)).

[McIntyre \(1952\)](#) propôs a Amostragem por Conjuntos Ordenados (ACO), cujo termo vem do inglês *Ranked Set Sampling (RSS)*. A ACO é uma alternativa mais eficiente e econômica em relação à AAS em situações em que a mensuração da variável de interesse é de alto custo ou de difícil acesso, o que impossibilitaria obter uma amostra numerosa, mas é possível ordenar as unidades amostrais, via julgamento pessoal, seja por meio de alguma variável concomitante, ou algum outro critério de baixo custo, sem precisar de fato mensurar a variável de interesse. As diferentes etapas para obter uma amostra de tamanho k via ACO consistem em, primeiramente, selecionar k^2 unidades da população via AAS dividindo-as, aleatoriamente, em k conjuntos de k unidades cada; na sequência, ranqueiam-se as unidades amostrais dentro de cada conjunto de acordo com a variável de interesse, mas através do critério de ordenação pré-estabelecido; finalmente, seleciona-se a unidade amostral ranqueada na primeira posição (aquele que se julga a menor, ou, caso a ordenação seja realizada de forma decrescente, aquela que se julga a maior) do 1º conjunto, bem como do 2º conjunto a unidade amostral ranqueada na segunda posição, e assim sucessivamente até selecionar a observação ranqueada na última posição do k -ésimo conjunto, para então mensurar a variável de interesse nesse subconjunto composto por k das k^2 unidades amostrais originais ([TACONELI; BARRETO, 2005](#)).

A ACO foi introduzida por [McIntyre \(1952\)](#) para estimar a produção média de pastagens, uma vez que se tinha alto custo no que diz respeito à mensuração de tal variável, pois é necessário lavrar, secar para então pesar a matéria seca. Ao invés disso, uma

pessoa experiente poderia ranquear amostras de parcelas do pasto de maneira eficiente por inspeção visual, ou com base em registros, fotos, vídeos, e após isso, selecionar unidades amostrais de acordo com a ordenação delas, para então fazer a mensuração.

Dentre as formas de ordenar as amostras já citadas, a ordenação por julgamento pessoal da variável de interesse pode ser realizada, por exemplo, com base na avaliação visual das unidades, ou através de fotos, vídeos, entre outros. Já na ordenação por variável concomitante, ordenam-se as unidades amostrais em relação à variável de interesse, mas com base nos valores de alguma variável concomitante, correlacionada com a variável de interesse. Independente do mecanismo utilizado para ordenação, se o critério de ordenação não apresentar erros, tem-se a ordenação perfeita; caso contrário, a ordenação imperfeita ([CHEN; BAI; SINHA, 2004](#)).

Existem várias publicações que comprovam o ganho da ACO em relação à AAS em diferentes contextos e na estimação de diferentes parâmetros, como a média e a variância. [Takahasi e Wakimoto \(1968\)](#) e [Dell e Clutter \(1972\)](#) comprovaram o ganho em precisão na estimação da média decorrente da aplicação da ACO, sendo que o ganho em precisão aumenta conforme a eficácia do critério de ordenação utilizado; [Stokes \(1980\)](#), [MacEachern et al. \(2002\)](#) e [Al-Hadrami \(2010\)](#) mostraram maior eficiência na estimação da variância. Diversas aplicações da ACO e delineamentos relacionados encontram-se publicados. Dentre eles, [Bowerman e O'Connell \(1993\)](#) utilizaram a ACO para estimar a variância e a média do preço de casas à venda. [Nussbaum e Sinha \(1997\)](#) utilizaram ACO para estimar a poluição atmosférica média causada pela combustão da gasolina liberada por escapamentos de carros. [Hatefi e Jozani \(2014\)](#) utilizaram a ACO e o PROS (outro delineamento baseado em conjuntos ordenados) em um estudo de câncer de mama com objetivo de estimar a proporção de pacientes com tumores malignos de câncer na população estudada; além desses, uma revisão de literatura sobre a ACO e outros delineamentos baseados em ranqueamentos de amostras são descritos em [Al-Omari e Bouza \(2014\)](#).

No Controle Estatístico de Qualidade (CEQ), utilizam-se de amostras para garantir a qualidade de um determinado produto. O conceito de qualidade, segundo [Juran \(1990\)](#), refere-se ao desempenho do produto quanto à percepção do cliente ou, ainda, à ocorrência mínima de falhas no produto. Segundo [Garvin \(1987\)](#), a qualidade pode ser sumarizada em 8 dimensões: desempenho, confiabilidade, durabilidade, assistência técnica, estética, características, qualidade percebida e conformidade com as especificações destinadas para determinado produto.

A qualidade de um produto nem sempre é garantida, pois podem haver alguns fatores, se não controlados, podem comprometê-la. Por isso, busca-se reduzir as causas de variação, que acarretam na redução da qualidade dos produtos. Há dois tipos de variabilidade inerentes a processos industriais, sendo suas causas classificadas como aleatórias ou atribuíveis ([NOGUEIRA, 2011](#)). As causas aleatórias são inerentes ao processo e configuram a variabilidade natural do processo contemplando, por exemplo, variações pequenas

nas condições do ambiente (temperatura, umidade, ...), da matéria-prima utilizada, rotação de mão de obra, dentre outros. Por outro lado, as causas atribuíveis não são naturais do processo e devem ser diagnosticadas e corrigidas para garantir um maior nível de qualidade do produto. Dentre as principais causas destacam-se a má regulagem de máquinas, a utilização de matéria prima defeituosa e erros de operação por parte de funcionários mal treinados, dentre outros. Uma vez que a qualidade do produto é inversamente proporcional à variabilidade, busca-se então controlar e reduzir as fontes de variabilidade, obtendo-se assim um processo classificado sob controle, que opera somente com causas aleatórias de variação ([MONTGOMERY, 2011](#)).

A principal forma de detectar a ocorrência de causas atribuíveis de variação é por meio de gráficos de controle. Os gráficos de controle (também conhecidos como cartas controles), propostos por [Shewhart \(1924\)](#), são compostos por uma linha central (LC) e um par de limites (limite inferior de controle (LIC) e limite superior de controle (LSC)), além de pontos que representam o valor de alguma estatística calculada em amostras que são coletadas regularmente do processo. A linha central e os limites de controle representam a distribuição da estatística utilizada, sendo calculados a partir de considerações probabilísticas na condição em que o processo opera sob controle. Padrões sistemáticos e valores extremos observados nos pontos dispostos em um gráfico de controle podem ser indicativos de descontrole. Em particular, pontos fora dos limites ou pontos que apareçam de forma sequencial acima ou abaixo da LC ([MONTGOMERY, 2011](#)).

A importância de um gráfico de controle é substancial, pois permite identificar rapidamente uma situação de descontrole. Um indicador que permite estudar o desempenho do gráfico é o Comprimento Médio da Sequência (CMS), que consiste no número médio de amostras observadas antes que algum ponto exceda os limites de controle ([SAMOHYL, 2009](#)). Particularmente, se o processo opera sob controle têm-se CMS_0 , o que representaria a taxa de falsos alarmes; já se o processo está fora de controle têm-se CMS_1 , que representaria o número médio de amostras até detecção do descontrole. Via simulação de gráficos de controle para o monitoramento da média, [Muttlak e Al-Sabah \(2010\)](#) demonstraram que o CMS é menor quando as amostras são retiradas via ACO se comparado à AAS, quando o processo não opera sob controle (considerando diferentes níveis de alteração na média do processo), para diferentes tamanhos de amostras e níveis de correlação entre a variável concomitante e a de interesse. Os autores comprovaram também menor CMS para outros dois delineamentos baseados em ranqueamento de amostras: *Median Ranked Set Sampling* (MRSS - ([MUTTLAK, 1997](#))) e *Extreme Ranked Set Sampling* (ERSS - ([SAMAWI; AHMED; ABU-DAYYEH, 1996](#))).

Ainda, outros delineamentos alternativos à ACO foram propostos e aplicados no monitoramento da média via gráficos de controle. Esses delineamentos são conhecidos por *double-ranked set sampling methods*, ou simplesmente *doubles*, pois são realizados em dois estágios de ordenação, sendo necessário selecionar uma amostra inicial de tamanho k^3

via AAS para obter uma amostra final de tamanho k . Dentre os delineamentos da classe dos *doubles* têm-se o *Double Quartile-Ranked Set Sampling* (*DQRSS* - ([JEMAIN; AL-OMARI, 2006](#))), *Quartile Double-Ranked Set Sampling* (*QDRSS* - ([JEMAIN; AL-OMARI, 2006](#))), *Double Extreme-Ranked Set Sampling* (*DERSS* - ([SAMAWI, 2002](#))) e *Extreme Double-Ranked Set Sampling* (*EDRSS*- ([ABUJIYA; MUTTLAK, 2007](#))). [Al-Omari e Haq \(2012\)](#) apresentaram resultados de um estudo por simulação e uma aplicação dos *doubles*, comparando-os com os delineamentos oriundos da ACO e a AAS na construção de gráficos de controle para a média. A simulação consistiu em comparar o CMS dos gráficos de controle produzidos desses delineamentos com os obtidos via AAS, ACO e *ERSS* na alteração da média do processo considerando diferentes tamanhos de amostra e ordenação perfeita. Em um segundo momento, todos os delineamentos foram utilizados em dados reais de resistência elétrica de isolamento. Em ambas as situações, os *double-ranked set sampling methods* foram superiores aos seus concorrentes, que são baseados em um único estágio de ordenação, além da AAS.

[Zamanzade e Al-Omari \(2015\)](#) propuseram um novo delineamento amostral baseado em ranqueamento de amostras, configurando uma alteração no delineamento original da ACO, que consiste na ordenação (após seleção de uma amostra de tamanho k^2 via AAS) das k^2 unidades amostrais em um único conjunto, ao invés de k conjuntos de k unidades. Esse novo delineamento é denominado *Unified Ranked Set Sampling* (*URSS*). Os autores verificaram via simulação Monte Carlo, considerando diferentes distribuições, níveis de correlação entre a variável de interesse e a variável concomitante e tamanhos de amostras, que o estimador da média populacional via *URSS* é mais eficiente em relação aos estimadores da média via ACO e AAS; e não viesado quando a distribuição da variável de interesse é simétrica e se tenha ordenação perfeita, sendo essas propriedades, demonstradas analiticamente. Em uma aplicação em dados reais, o estimador da média populacional via *URSS* também se mostrou mais eficiente que os estimadores da média via ACO e AAS.

1.1 Problema e Justificativa

Com base no que foi apresentado e a partir do interesse sobre o novo delineamento (*URSS*), este trabalho apresenta um estudo via simulação do desempenho do *URSS* na construção de gráficos de controle para a média, comparando-o em relação à AAS, ACO, *ERSS*, *MRSS*, *DQRSS*, *QDRSS*, *DERSS* e *EDRSS*. Além disso, este trabalho visa apontar a aplicabilidade de cada delineamento amostral, facilitando a escolha do operador e resultando em economia de custos.

1.1.1 Objetivo Geral

Analisar o desempenho do delineamento *URSS* no contexto do CEP para a construção de gráficos de controle para a média, por meio de um estudo de simulação.

1.1.2 Objetivos Específicos

- Comparar o desempenho do *URSS* quanto ao CMS de gráficos de controle para a média em relação à AAS e a outros delineamentos baseados em ranqueamento de amostras;
- Verificar o efeito do tamanho amostral e da eficácia do critério de ordenação nos diferentes graus de descontrole da média do processo para cartas de controle obtidas via *URSS*;
- Aplicação do *URSS* em dados reais;
- Propor um estimador da variância do estimador da média via *URSS* e estudar seu viés relativo via simulação;
- Avaliar a aplicabilidade do *URSS* como alternativa a outros delineamentos no contexto do CEP.

2 MATERIAL E MÉTODOS

Este capítulo apresenta os delineamentos ACO, *ERSS*, *MRSS* e *URSS*, descreve os correspondentes processos de seleção de amostras e apresenta os estimadores da média populacionais e a variância dos estimadores da média para cada um deles. Para os delineamentos *DQRSS*, *QDRSS*, *DERSS* e *EDRSS* é apresentado uma discussão dos passos para a seleção de amostras. Ainda, são apresentadas as definições e propriedades de estatísticas de ordem, fundamentais no contexto de amostragem baseada em conjuntos ordenados, nas situações em que não se tem erro de ordenação para a variável de interesse. Além disso, são descritos a forma de construção de gráficos de controle e demais métodos estatísticos utilizados neste estudo. Finaliza-se com a descrição do estudo por simulação e da aplicação do *URSS* em um conjunto de dados reais.

2.1 Delineamentos Amostrais por Conjuntos Ordenados

Em situações em que as unidades amostrais são de difícil acesso e/ou de alto custo para serem coletadas, mas é possível ordená-las, segundo a variável de interesse, sem de fato quantificá-la, seja por meio de julgamento pessoal ou por alguma variável concomitante, os delineamentos amostrais por conjuntos ordenados têm sido uma alternativa, pois apresentam alta eficiência e baixo custo ([BARRETO, 2001](#)). Nesse contexto, delineamentos como a Amostragem por Conjuntos Ordenados (ACO - ([MCINTYRE, 1952](#))), *Extreme Ranked Set Sampling* (*ERSS*- ([SAMAWI; AHMED; ABU-DAYYEH, 1996](#))) e *Median Ranked Set Sampling* (*MRSS*- ([MUTTLAK, 1997](#))) foram propostos, além do *Unified Ranked Set Sampling* (*URSS*- ([ZAMANZADE; AL-OMARI, 2015](#))), como mais uma alternativa à ACO. Nesta seção são apresentados cada um desses delineamentos, os respectivos processos para seleção de amostras e seus estimadores para a média. Além desses, são brevemente descritos também os delineamentos da classe *double-ranked set sampling*.

2.1.1 Amostragem por Conjuntos Ordenados

A amostragem por Conjuntos Ordenados foi proposta por [McIntyre \(1952\)](#), que constatou ser um método mais eficiente que a AAS quando as unidades amostrais são de difícil acesso ou de cara mensuração, sendo possível ordenar as unidades amostrais por julgamento pessoal ou algum outro procedimento de baixo custo, sem mensurar a variável de interesse. Para seleção de uma amostra de tamanho k via ACO, deve-se seguir os seguintes passos:

1. Selecionam-se, via AAS, k^2 unidades da população, dividindo-as ao acaso em k conjuntos de k unidades amostrais cada;
2. Em cada conjunto, ranqueiam-se as unidades amostrais segundo os valores da variável de interesse, mas com base apenas no critério de ordenação pré-estabelecido;
3. Do primeiro conjunto seleciona-se a unidade amostral ranqueada na primeira posição (aquele que se julga a menor, ou, caso a ordenação seja realizada de forma decrescente, aquela que se julga a maior) e mensura-se nela a variável de interesse. Do segundo conjunto seleciona-se a unidade amostral ranqueada na segunda posição e mensura-se a variável de interesse. Seguindo esse procedimento até o k -ésimo conjunto, tem-se uma amostra por conjuntos ordenados de k unidades amostrais;
4. Os passos 1, 2 e 3 podem ser repetidos n vezes, configurando n ciclos e produzindo assim uma amostra de tamanho nk .

Ao final do processo, a Amostra por Conjuntos Ordenados pode ser denotada por:

$$Y_{[i]j}; i = 1, 2, \dots, k; j = 1, 2, \dots, n, \quad (2.1)$$

em que $Y_{[i]j}$ refere-se à observação que foi ranqueada na i -ésima posição, no j -ésimo ciclo. Nesse caso, os $Y_{[i]j}$'s são variáveis aleatórias independentes mas não identicamente distribuídas, que seguem a distribuição da estatística de ordem i de uma amostra de tamanho k , sob ordenação perfeita. A Seção 2.2 aborda com mais detalhes a definição e propriedades das estatísticas de ordem.

A [Figura 1](#) apresenta um exemplo de ilustração da seleção de uma amostra de tamanho $k = 5$ via ACO, com um único ciclo e a ordenação das unidades amostrais feita com base nos valores de uma variável concomitante X , tal que $X_{(i)}$ é a i -ésima estatística de ordem da variável concomitante e $Y_{[i]}$ a i -ésima estatística de ordem julgada da variável de interesse. As colunas representam os 5 conjuntos de tamanho 5 (e as linhas as próprias unidades amostrais), necessárias para a seleção da amostra de tamanho 5.

Conjunto 1	Conjunto 2	Conjunto 3	Conjunto 4	Conjunto 5
$(X_{(1)}, Y_{[1]})$				
$(X_{(2)}, Y_{[2]})$				
$(X_{(3)}, Y_{[3]})$				
$(X_{(4)}, Y_{[4]})$				
$(X_{(5)}, Y_{[5]})$				

Figura 1 – Representação da seleção de uma amostra de tamanho $k = 5$ via ACO

Fonte: Os autores

A Figura 2 ilustra a seleção de uma amostra por conjuntos ordenados de tamanho $nk = 9$, utilizando-se de três ciclos ($n=3$), para tal fim. Em cada ciclo, os conjuntos estão representados nas linhas. A variável de interesse é a quantidade de líquido dentro de cada garrafa e o critério utilizado para ordenação baseia-se em inspeção visual. Considera-se que para a construção de cada ciclo, foram selecionadas ao acaso $k^2=9$ garrafas, divididas aleatoriamente em 3 conjuntos com 3 garrafas cada. Para cada ciclo, no primeiro conjunto seleciona-se a garrafa que se julga, com base na avaliação visual, ter mais líquido; no segundo a que se julga ter a segunda maior quantidade de líquido e no último conjunto seleciona-se a garrafa com menor quantidade de líquido. As unidades assinaladas em cada conjunto, para os três ciclos, irão compor a amostra final, sendo as únicas de fato mensuradas quanto à variável de interesse (quantidade de líquido).

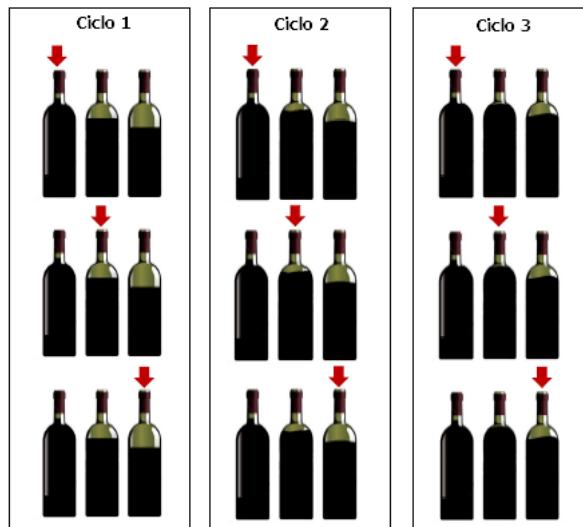


Figura 2 – Representação da seleção de uma amostra por conjuntos ordenados de tamanho 9, com 3 ciclos.

Fonte: Adaptado de [Silva e Rocha \(2013\)](#)

O estimador da média populacional proposto por [McIntyre \(1952\)](#), é a média da amostra por conjuntos ordenados, dada por:

$$\bar{Y}_{ACO} = \frac{1}{nk} \sum_{j=1}^n \sum_{i=1}^k Y_{[i]j}, \quad (2.2)$$

cuja variância é dada por:

$$Var(\bar{Y}_{ACO}) = \frac{\sigma^2}{nk} - \frac{1}{nk^2} \sum_{j=1}^n \sum_{i=1}^k (\mu_{[i]} - \mu)^2 \quad (2.3)$$

sendo μ e σ a média e o desvio padrão da população de interesse e $\mu_{[i]}$ corresponde à média da distribuição de $Y_{[i]j}$, $i = 1, 2, \dots, k$.

Diferentes publicações comprovaram o ganho em precisão do estimador da média via ACO em relação ao estimador da média via AAS. Dentre elas, [Takahasi e Wakimoto](#)

(1968) provaram matematicamente que \bar{Y}_{ACO} é não viesado e de menor variância em relação ao correspondente estimador obtido por meio da AAS. Dell e Clutter (1972) estudaram a estimação da média no caso de ordenação imperfeita e constataram que quanto menor o erro no ranqueamento, mais preciso o estimador produzido pela ACO se comparado ao estimador obtido via AAS. Além de ganho em precisão na estimação da média, Barreto (2001) destaca, para estimação de parâmetros populacionais, seja por estimação não-paramétrica, de máxima verossimilhança ou de mínimos quadrados que a ACO também se mostra mais vantajosa em relação à AAS.

2.1.2 Median Ranked Set Sampling

Como uma alternativa à ACO, e com o objetivo de propor estimadores mais precisos para a média de distribuições simétricas, Muttlak (1997) propôs o delineamento *Median Ranked Set Sampling (MRSS)*. Para obter uma amostra via *MRSS* de tamanho k , seguem-se os seguintes passos:

1. Selecionam-se, via AAS, k^2 unidades da população, dividindo-as em k conjuntos de k unidades amostrais cada;
2. Dentro de cada conjunto ranqueiam-se as unidades amostrais em relação à variável de interesse, através de julgamento pessoal ou dos valores de alguma variável concomitante;
3. Se k for ímpar, deve-se selecionar para medição a $((k+1)/2)$ -ésima unidade amostral (no caso, a mediana) de cada conjunto. Se k for par, deve-se selecionar a unidade amostral com o $\frac{k}{2}$ -ésimo menor valor julgado nos primeiros $k/2$ conjuntos e dos outros $k/2$ conjuntos selecionar a unidade amostral que se julgar com o $((k+2)/2)$ -ésimo menor valor;
4. Os passos 1, 2 e 3 podem ser repetidos n vezes, produzindo assim uma amostra de tamanho nk .

Seguindo os passos acima, obtém-se uma amostra via *MRSS* de k unidades amostrais (ou nk , considerando ciclo diferente de 1).

A Figura 3 ilustra a aplicação do *MRSS* para a seleção de amostras de tamanhos $k=4$ e 5 (com um único ciclo) e a ordenação das unidades amostrais feita com base nos valores de uma variável concomitante X. Cada coluna representa um conjunto de unidades amostrais.

Conjunto 1	Conjunto 2	Conjunto 3	Conjunto 4
$(X_{(1)}, Y_{[1]})$	$(X_{(1)}, Y_{[1]})$	$(X_{(1)}, Y_{[1]})$	$(X_{(1)}, Y_{[1]})$
$(X_{(2)}, Y_{[2]})$	$(X_{(2)}, Y_{[2]})$	$(X_{(2)}, Y_{[2]})$	$(X_{(2)}, Y_{[2]})$
$(X_{(3)}, Y_{[3]})$	$(X_{(3)}, Y_{[3]})$	$(X_{(3)}, Y_{[3]})$	$(X_{(3)}, Y_{[3]})$
$(X_{(4)}, Y_{[4]})$	$(X_{(4)}, Y_{[4]})$	$(X_{(4)}, Y_{[4]})$	$(X_{(4)}, Y_{[4]})$

(a) Tamanho de amostra ($k = 4$)

Conjunto 1	Conjunto 2	Conjunto 3	Conjunto 4	Conjunto 5
$(X_{(1)}, Y_{[1]})$				
$(X_{(2)}, Y_{[2]})$				
$(X_{(3)}, Y_{[3]})$				
$(X_{(4)}, Y_{[4]})$				
$(X_{(5)}, Y_{[5]})$				

(b) Tamanho de amostra ($k = 5$)Figura 3 – Representação da seleção de uma amostra de tamanhos $k = 4$ e 5 via *MRSS*

Seja $Y_{[i:m]j}$ o valor que se julga ser a mediana, no i -ésimo conjunto de tamanho k ímpar no j -ésimo ciclo. Também, denote por $Y_{[i:m]j}$ a $(k/2)$ -ésima estatística de ordem julgada do i -ésimo conjunto de tamanho k ($i = 1, 2, \dots, L = k/2$) e a $((k+2)/2)$ -ésima estatística de ordem julgada do i -ésimo conjunto de tamanho k ($i = L + 1, L + 2, \dots, k$) se o tamanho do conjunto é par. O estimador da média populacional via *MRSS*, proposto por [Muttlak \(1997\)](#) é dado por:

$$\bar{Y}_{MRSS} = \frac{1}{nk} \sum_{j=1}^n \sum_{i=1}^k Y_{[i:m]j}, \quad j = 1, 2, \dots, n, \quad (2.4)$$

com variância dada por:

$$Var(\bar{Y}_{MRSS}) = \frac{1}{nk^2} \sum_{j=1}^n \sum_{i=1}^k \sigma_{[i:m]j}^2, \quad (2.5)$$

tal que $\sigma_{[i:m]}^2 = E[Y_{[i:m]} - E(Y_{[i:m]})]^2$. Vale ressaltar que, para a estimação da média, o delineamento *MRSS* fornece um estimador não viciado nos casos em que a distribuição é simétrica, pois a mediana representaria a própria média da distribuição.

[Muttlak \(1998\)](#) apresenta a comparação do estimador da média via *MRSS* (\bar{Y}_{MRSS}) com o estimador da média via ACO (\bar{Y}_{ACO}), verificando a maior precisão de \bar{Y}_{MRSS} . Ainda, considerando estimadores do tipo regressão, foi comprovada a maior eficiência do *MRSS* quando as unidades amostrais são ordenadas através de uma variável concomitante, a menos que as variáveis de interesse e concomitante apresentem correlação maior que 0,90. [Muttlak e Al-Sabah \(2010\)](#) apresentaram um estudo comparando os delineamentos amostrais AAS, ACO, *ERSS* e *MRSS* no contexto de CEP, e comprovaram que o *MRSS*

superá os demais delineamentos no monitoramento da média de processos com base em gráficos de controle.

2.1.3 *Extreme Ranked Set Sampling*

O delineamento *Extreme Ranked Set Sampling* (*ERSS*) foi proposto por [Samawi, Ahmed e Abu-Dayyeh \(1996\)](#) com o intuito de reduzir os erros de ordenação, uma vez que requer apenas identificar valores extremos no momento da ordenação. Para obter uma amostra de tamanho k via *ERSS*, deve-se realizar as seguintes etapas:

1. Selecionam-se k^2 unidades da população via AAS, dividindo-as em k conjuntos de k unidades amostrais cada;
2. Ranqueiam-se as unidades amostrais em ordem crescente dentro de cada conjunto em relação à variável de interesse;
3. Se k for par, seleciona-se nos primeiros $k/2$ conjuntos, a unidade amostral que se julga ter o menor valor para a variável de interesse; para os demais $k/2$ conjuntos a unidade amostral que se julga ter o maior valor. Se k for ímpar seleciona-se dos primeiros $(k - 1)/2$ conjuntos, a unidade amostral que se julga ser a de menor valor ranqueada do conjunto, dos outros $(k - 1)/2$ conjuntos a unidade amostral que se julga ser a de maior valor ranqueada do conjunto, e do conjunto restante, a unidade amostral que se julga ser a mediana;
4. Os passos 1, 2 e 3 podem ser repetidos n vezes, formando uma amostra de tamanho nk .

Nota-se que para casos em que k for par, não é necessário ranquear todas as unidades amostrais em ordem crescente dentro de cada conjunto, uma vez que só é preciso identificar a menor e a maior unidade amostral. E se k for ímpar, só deve-se ter o trabalho de ranquear um único conjunto, aquele que se faz uso do valor da mediana.

A [Figura 4](#) ilustra a aplicação do *ERSS* para a seleção de amostras de tamanhos $k=4$ e 5 , ambos com um único ciclo e a ordenação das unidades amostrais realizada com base nos valores de uma variável concomitante X. Cada coluna representa um conjunto composto pelas unidades amostrais.

Conjunto 1	Conjunto 2	Conjunto 3	Conjunto 4
$(X_{(1)}, Y_{[1]})$	$(X_{(1)}, Y_{[1]})$	$(X_{(1)}, Y_{[1]})$	$(X_{(1)}, Y_{[1]})$
$(X_{(2)}, Y_{[2]})$	$(X_{(2)}, Y_{[2]})$	$(X_{(2)}, Y_{[2]})$	$(X_{(2)}, Y_{[2]})$
$(X_{(3)}, Y_{[3]})$	$(X_{(3)}, Y_{[3]})$	$(X_{(3)}, Y_{[3]})$	$(X_{(3)}, Y_{[3]})$
$(X_{(4)}, Y_{[4]})$	$(X_{(4)}, Y_{[4]})$	$(X_{(4)}, Y_{[4]})$	$(X_{(4)}, Y_{[4]})$

(a) Tamanho de amostra ($k = 4$)

Conjunto 1	Conjunto 2	Conjunto 3	Conjunto 4	Conjunto 5
$(X_{(1)}, Y_{[1]})$				
$(X_{(2)}, Y_{[2]})$				
$(X_{(3)}, Y_{[3]})$				
$(X_{(4)}, Y_{[4]})$				
$(X_{(5)}, Y_{[5]})$				

(b) Tamanho de amostra ($k = 5$)Figura 4 – Representação da seleção de uma amostra de tamanhos $k = 4$ e 5 via *ERSS*

Seja $Y_{[i:e]j}$ o menor valor julgado do i -ésimo conjunto ($i = 1, 2, \dots, L_1 = \frac{k}{2}$) ou o maior valor julgado do i -ésimo conjunto ($i = L_1 + 1, L_1 + 2, \dots, k$), se k for par. Ainda, seja $Y_{[i:e]j}$ o menor valor julgado do i -ésimo conjunto para $i = 1, 2, \dots, L_2 = \frac{k-1}{2}$, a mediana julgada do $(k+1)/2$ -ésimo conjunto e o maior valor julgado do i -ésimo conjunto para $i = L_2 + 2, L_2 + 3, \dots, k$, se k for ímpar. O estimador da média populacional via *ERSS* é dado por:

$$\bar{Y}_{ERSS} = \frac{1}{nk} \sum_{j=1}^n \sum_{i=1}^k Y_{[i:e]j}, \quad j = 1, 2, \dots, n, \quad (2.6)$$

sendo a variância de \bar{Y}_{ERSS} dada por:

$$Var(\bar{Y}_{ERSS}) = \frac{1}{nk^2} \sum_{j=1}^n \sum_{i=1}^k \sigma_{[i:e]}^2, \quad (2.7)$$

com $\sigma_{[i:e]}^2 = E[Y_{[i:e]} - E(Y_{[i:e]})]^2]$.

Assim como o *MRSS*, o delineamento *ERSS* fornece um estimador da média não viciado nos casos em que as distribuições são simétricas, pois utiliza dos valores extremos (o menor valor julgado de cada um dos primeiros $k/2$ conjuntos e o maior valor julgado de dos restantes $k/2$ conjuntos), além da mediana se k for ímpar.

O delineamento *ERSS* é recomendado para a construção de gráficos de controle de qualidade para a média, pois tem desempenho superior à *AAS* ([MUTTLAK; ALSABAH, 2010](#)), e é facilmente implementado na área de estudo, pois em algumas situações é necessário somente identificar a menor e a maior unidade amostral e medi-las, sem necessidade de ranquear as unidades amostrais ([JABEEN, 2012](#)).

2.1.4 Double-Ranked Set Sampling

Além dos delineamentos já citados, outros delineamentos que merecem destaque são os da classe *double-ranked set sampling*, chamados no decorrer do texto apenas por *doubles*. Recebem esse nome, pois o processo para seleção de amostra final é realizado em dois estágios de ordenação. Entre os delineamentos pertencentes a essa classe: *Double Quartile-Ranked Set Sampling* (*DQRSS* - ([JEMAIN; AL-OMARI, 2006](#))), *Quartile Double-Ranked Set Sampling* (*QDRSS* - ([AL-OMARI; AL-SALEH, 2009](#))), *Double Extreme-Ranked Set Sampling* (*DERSS* - ([SAMAWI, 2002](#))) e *Extreme Double-Ranked Set Sampling* (*EDRSS* - ([ABUJIYA; MUTTLAK, 2007](#))). A principal diferença é que, para esses delineamentos, selecionam-se, inicialmente, k^3 elementos da população de interesse para obter uma amostra de tamanho k . Num primeiro estágio de ordenação, as k^3 unidades amostrais originais são reduzidas para k^2 , e após isso, realiza-se outro estágio de ordenação para finalmente obter uma amostra de tamanho k . Já para os delineamentos apresentados anteriormente, inicia-se o procedimento de amostragem com uma amostra de k^2 elementos da população, obtendo-se uma amostra final de k elementos, após um único estágio de ordenação.

Os estágios de ordenação e consequente seleção de unidades amostrais para os *doubles*, são compostos seja por repetir o mesmo processo de ordenação e seleção duas vezes, ou mesclar dois diferentes processos de seleção. Por exemplo, para o *EDRSS*, em um primeiro momento, usa-se o processo de seleção de unidades amostrais da *ACO*, em seguida, o do *ERSS*. Já para o *DERSS* utiliza-se do processo de seleção de unidades amostrais do *ERSS* duas vezes. Para o *QDRSS*, utiliza-se do processo de seleção de unidades amostrais da *ACO*; em seguida, seleciona-se das unidades amostrais resultantes do primeiro processo, as unidades amostrais correspondentes aos quartis. Já para o *DQRSS*, utiliza-se do procedimento do quartil duas vezes para seleção de amostras. Esse parágrafo trata apenas da essência de cada delineamento, sendo que maiores detalhes e o procedimento completo para a aplicação deles, devem ser conferidos nos artigos originais.

[Al-Omari e Haq \(2012\)](#) realizaram um estudo de simulação para comparar o desempenho dos *DERSS*, *DQRSS*, *QDRSS* e *EDRSS* em relação à *AAS*, *ACO* e *ERSS* e constataram menor CMS para detectar uma alteração na média do processo, por meio de gráficos de controle para a média.

2.1.5 Unified Ranked Set Sampling

Proposto por [Zamanzade e Al-Omari \(2015\)](#) como uma alternativa à *ACO* para estimar a média e a variância populacional, o procedimento para seleção de amostras via *URSS* é descrito a seguir:

1. Para obter uma amostra de tamanho k , selecionam-se k^2 unidades amostrais via *AAS* da população de interesse, ranqueando-as em um único conjunto de acordo com

a variável de interesse, mas com base em uma variável concomitante, por julgamento pessoal ou algum outro procedimento livre de custos;

2. Seleciona-se a $[(i - 1)k + l]$ -ésima unidade amostral para mensuração da variável de interesse, para i variando de $1, \dots, k$. Se k é ímpar, $l = \frac{k+1}{2}$; se ambos k e i são pares, então $l = \frac{k}{2}$; mas se k é par e i é ímpar, então $l = \frac{k+2}{2}$.
3. Os passos 1 e 2 podem ser repetidos n vezes, se desejado, produzindo uma amostra de tamanho nk .

Nesse caso, a amostra via *URSS* pode ser denotada por $\{Y_{[(i-1)k+l]j}; i = 1, 2, \dots, k; j = 1, 2, \dots, n\}$, em que $Y_{[(i-1)k+l]j}$ se refere à estatística de ordem julgada $[(i - 1)k + l]$ de uma amostra de tamanho k . Diferentemente da ACO, em que as unidades amostrais são independentes, pois são provenientes de amostras distintas, no *URSS* elas não são independentes, uma vez que são oriundas dos elementos ordenados de uma mesma amostra, e implicando em covariância entre as unidades amostrais selecionadas. Para quaisquer valores de k e i , $Y_{[(i-1)k+l]j}$ segue a distribuição da $((i - 1)k + l)$ -ésima estatística de ordem de uma AAS de tamanho k^2 da população sob estudo, quando a ordenação é perfeita. Quando a ordenação é imperfeita, $Y_{[(i-1)k+l]j}$ segue a distribuição da $[(i - 1)k + l]$ -ésima estatística de ordem julgada para uma AAS de tamanho k^2 .

A título de ilustração, suponha que se queira obter uma amostra de tamanho $k = 3$ via *URSS* com um único ciclo, e que a ordenação seja feita com base nos valores de uma variável concomitante X. O processo inicia-se selecionando aleatoriamente via AAS uma amostra de $k^2 = 9$ unidades denotada por:

$$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), (X_4, Y_4), (X_5, Y_5), (X_6, Y_6), (X_7, Y_7), (X_8, Y_8), (X_9, Y_9).$$

As unidades amostrais são ranqueadas com base na variável de interesse Y, mas utilizando a variável concomitante X, produzindo:

$$\begin{aligned} & (X_{(1)}, Y_{[1]}), (X_{(2)}, Y_{[2]}), (X_{(3)}, Y_{[3]}), (X_{(4)}, Y_{[4]}), (X_{(5)}, Y_{[5]}), \\ & (X_{(6)}, Y_{[6]}), (X_{(7)}, Y_{[7]}), (X_{(8)}, Y_{[8]}), (X_{(9)}, Y_{[9]}), \end{aligned}$$

tal que $X_{((i-1)k+l)}$ é a i -ésima estatística de ordem da variável concomitante e $Y_{[(i-1)k+l]}$ a i -ésima estatística de ordem julgada da variável de interesse. Ao aplicar o delineamento apresentado para seleção da amostra via *URSS* as posições das unidades amostrais selecionadas para compor a amostra de tamanho $k = 3$ segundo o algoritmo apresentado, são 2, 5 e 8:

$$\begin{aligned} & (X_{(1)}, Y_{[1]}), \boxed{(X_{(2)}, Y_{[2]})}, (X_{(3)}, Y_{[3]}), (X_{(4)}, Y_{[4]}), \boxed{(X_{(5)}, Y_{[5]})}, \\ & (X_{(6)}, Y_{[6]}), (X_{(7)}, Y_{[7]}), \boxed{(X_{(8)}, Y_{[8]})}, (X_{(9)}, Y_{[9]}). \end{aligned}$$

Obtém-se assim uma amostra via *URSS* de tamanho $k=3$ composta pelas seguintes unidades amostrais:

$$(Y_{[2]}), (Y_{[5]}), (Y_{[8]}), \quad (2.8)$$

conforme [Zamanzade e Al-Omari \(2015\)](#).

Como neste estudo são consideradas amostras de tamanhos $k=3, 4, 5$ e 6 , aplicando o delineamento *URSS* as posições correspondentes às unidades amostrais selecionadas para esses tamanhos de k são apresentadas na [Tabela 1](#).

Tabela 1 – Representação das posições das unidades amostrais via *URSS*, a serem selecionadas de uma amostra de tamanho k^2 , após ordenação.

k	k^2	Posições das unidades amostrais selecionadas
3	9	2, 5 e 8
4	16	3, 6, 11 e 14
5	25	3, 8, 13, 18 e 23
6	36	4, 9, 16, 21, 28 e 33

Fonte: Os autores

O estimador da média populacional via *URSS* é não viesado se a ordenação for perfeita e se a distribuição da variável de interesse for simétrica, sendo definido por:

$$\bar{Y}_{URSS} = \frac{1}{nk} \sum_{j=1}^n \sum_{i=1}^k Y_{[(i-1)k+l]j}. \quad (2.9)$$

A variância de \bar{Y}_{URSS} é dada por:

$$Var(\bar{Y}_{URSS}) = \frac{1}{nk^2} \sum_{j=1}^n \sum_{i=1}^k Var(Y_{[(i-1)k+l]j}) + \frac{2}{nk^2} \sum_{j=1}^n \sum_{i < i'}^k Cov(Y_{[(i-1)k+l]j}, Y_{[(i'-1)k+l]j}). \quad (2.10)$$

Em um estudo de simulação em que foi assumido o modelo de ranqueamento imperfeito de [Dell e Clutter \(1972\)](#), [Zamanzade e Al-Omari \(2015\)](#) mostraram que a eficiência do estimador da média via *URSS*, supera a dos estimadores da média via ACO e AAS, considerando vários cenários: diferentes tamanhos de amostra ($k = 5$ e 10), quantidade de ciclos ($n = 10$ e 20) e valores de correlação entre a variável de interesse e a variável concomitante ($\rho = 0, 0.2, 0.4, 0.6, 0.8, 1$); e por último, sob diferentes distribuições de probabilidades, sendo elas: Normal (0,1), Uniforme (0,1), Logística (0,1), Exponencial (1) e Beta (0.2,1).

2.2 Estatísticas de Ordem

Como citado anteriormente, os delineamentos oriundos da ACO fazem uso da ordenação de amostras segundo a variável de interesse, com base em algum critério de

ordenação pré-estabelecido. Na situação em que a ordenação da variável de interesse ocorra sem erros, as variáveis ordenadas da amostra constituem estatísticas de ordem. A seguir, são apresentadas a definição de estatísticas de ordem e suas principais propriedades, como as distribuições marginais e conjunta, a esperança, a variância e a covariância.

2.2.1 Definição

Sejam Y_1, Y_2, \dots, Y_k variáveis aleatórias. Para obter as estatísticas de ordem, ranqueiam-se as k variáveis aleatórias Y'_i s, podendo denotá-las por $Y_{(1)k}, Y_{(2)k}, \dots, Y_{(k)k}$, de tal forma que $Y_{(1)k} \leq Y_{(2)k} \leq \dots \leq Y_{(k)k}$, em que o subscrito entre parênteses identifica a estatística de ordem i de uma amostra de tamanho k , com $i = 1, 2, \dots, k$. Desse modo, $Y_{(1)k}$ é a primeira estatística de ordem, que corresponde ao menor valor (caso o ranqueamento seja em ordem crescente) dos Y'_i s; $Y_{(2)k}$ é a segunda estatística de ordem que corresponde ao segundo menor valor dos Y'_i s, e assim sucessivamente até $Y_{(k)k}$, a k -ésima estatística de ordem, que corresponde ao maior valor dos Y'_i s ([CASELLA; BERGER, 2001](#)).

2.2.2 Distribuição Marginal

Seja Y_1, Y_2, \dots, Y_k uma amostra aleatória de uma variável Y com função distribuição acumulada $F(y)$ e função densidade de probabilidade $f(y)$, e $Y_{(1)k}, Y_{(2)k}, \dots, Y_{(k)k}$ as correspondentes estatísticas de ordem. A função distribuição acumulada da i -ésima estatística de ordem, de uma amostra de tamanho k , segundo [Balakrishnan e Rao \(1998\)](#) de $Y_{(i)k}$ ($i = 1, 2, \dots, k$) é dada por:

$$F_{i:k}(y) = P(Y_{(i)k} \leq y) = \sum_{r=i}^k \binom{k}{r} F(y)^r [1 - F(y)]^{k-r} = \int_0^{F(x)} \frac{k!}{(i-1)!(k-i)!} t^{i-1} (1-t)^{k-i} dt, \quad (2.11)$$

para $-\infty < y < \infty$. Casos particulares, as funções distribuições acumuladas para o mínimo e o máximo são dadas, respectivamente, por:

$$F_{1:k}(y) = 1 - [1 - F(y)]^k, \quad -\infty < y < \infty, \quad (2.12)$$

e

$$F_{k:k}(y) = [F(y)]^k, \quad -\infty < y < \infty. \quad (2.13)$$

A função densidade de probabilidade de $Y_{(i)k}$, ($i = 1, 2, \dots, k$), no caso contínuo é obtida derivando a equação (2.11) em y :

$$f_{i:k}(y) = \frac{k!}{(i-1)!(k-i)!} F(y)^{i-1} \{1 - F(y)\}^{k-i} f(y), -\infty < y < \infty. \quad (2.14)$$

Em particular:

$$f_{1:k}(y) = k \{1 - F(y)\}^{k-1} f(y), -\infty < y < \infty, \quad (2.15)$$

e

$$f_{k:k}(y) = kF(y)^{k-1} f(y), -\infty < y < \infty, \quad (2.16)$$

são as funções densidades de probabilidade para o mínimo e máximo respectivamente.

2.2.3 Distribuição conjunta

A função densidade de probabilidade conjunta para duas estatísticas de ordem $Y_{(i)k}$ e $Y_{(i')k}$ é dada por:

$$\begin{aligned} f_{i,i':k}(y_i, y_{i'}) &= \frac{k!}{(i-1)!(i'-i-1)!(k-i')!} F(y_i)^{i-1} \{F(y_{i'}) - F(y_i)\}^{i'-i-1} \\ &\quad \times \{1 - F(y_{i'})\}^{n-i'} f(y_i) f(y_{i'}), y_i < y_{i'}. \end{aligned} \quad (2.17)$$

2.2.4 Momentos e Momentos Produtos

A esperança da i -ésima estatística de ordem de uma amostra de tamanho k é dada por:

$$E(Y_{(i)k}) = \mu_{i:k} = \int_{-\infty}^{\infty} y f_{i:k}(y) dy, (i = 1, 2, \dots, k) \quad (2.18)$$

e a variância dada por:

$$\begin{aligned} Var(Y_{(i)k}) &= \sigma_{i:k}^2 = \int_{-\infty}^{\infty} [y - \mu_{i:k}]^2 f_{i:k}(y) dy \\ &= \int_{-\infty}^{\infty} y^2 f_{i:k}(y) dy - \left[\int_{-\infty}^{\infty} y f_{i:k}(y) dy \right]^2. \end{aligned} \quad (2.19)$$

A covariância entre $Y_{(i)k}$ ($i = 1, 2, \dots, k$) e $Y_{(i')k}$ ($i' = 1, 2, \dots, k$) para $i < i'$, pode ser determinada por:

$$\sigma_{i,i':k} = Cov(Y_{i:k}, Y_{i':k}) = \mu_{i,i':k} - (\mu_{i:k})(\mu_{i':k}), \quad (2.20)$$

sendo

$$\mu_{i,i':k} = \int \int_{x < y} xy f_{i,i':k}(x, y) dx dy \quad (2.21)$$

Esses resultados são necessários, particularmente, para calcular a variância dos estimadores da média dos delineamentos baseados em conjuntos ordenados quando se tem ordenação perfeita.

A maioria das equações apresentadas nesta seção foram para casos de variáveis aleatórias contínuas. Para os casos discretos, as mesmas propriedades são válidas, substituindo-se as integrais por somatórios.

2.3 Controle Estatístico de Processos

Um processo de fabricação necessita garantir a qualidade do produto final. Todo processo de produção apresenta algum nível de variabilidade, caso a variabilidade seja grande (classificada como tal pelo operador), pode vir a comprometer a qualidade do produto final. Para manter a qualidade dos produtos, busca-se diminuir a variabilidade dos itens produzidos de forma que a variabilidade resultante seja de causas aleatórias (inerentes ao processo). A variabilidade de um processo pode ser discriminada em dois tipos de causas: as causas aleatórias e as atribuíveis ([NOGUEIRA, 2011](#)). As causas aleatórias são inerentes ao processo e configuram sua variabilidade natural, enquanto as causas atribuíveis não são naturais do processo e devem ser diagnosticadas e corrigidas para garantir um maior nível de qualidade do produto. Busca-se então identificar e eliminar as causas atribuíveis de variação, configurando-se assim um processo sob controle, que opera somente com causas aleatórias de variação ([MONTGOMERY, 2011](#)). Ao longo da produção, o processo pode vir a sair de controle, devido a fatores como o desgaste de máquinas, erros de configuração e funcionários mal treinados, dentre outros. Para controlar o processo produtivo, identificar possíveis falhas na produção e reduzir sua variabilidade, o Controle Estatístico de Processos (CEP) desempenha um papel fundamental ([MONTGOMERY, 2011](#)), pois engloba um conjunto de procedimentos que permitem ao operador/supervisor conhecer melhor o processo, e, dentre outros, identificar mais rapidamente uma possível alteração em seus parâmetros, subsidiando a decisão de intervir no processo, que fica a cargo da pessoa responsável pela área.

Os procedimentos do CEP auxiliam no monitoramento de parâmetros do processo, atuando na detecção das possíveis causas de descontrole, fazendo uso de gráficos, além de outras ferramentas para a obtenção e manutenção de um padrão de qualidade. A principal ferramenta para realizar o monitoramento de processos é o gráfico de controle, que tem como objetivo monitorar a variabilidade do processo. Comprimento médio de sequência (CMS) e curva característica do processo configuram indicadores de desempenho

de gráficos de controle. Por sua vez, diagrama de causa e efeito (conhecido também por espinha de peixe) e gráfico de pareto são ferramentas úteis para identificar as possíveis causas de descontrole. Nesta seção descreve-se a construção e uso de gráficos de controle, no caso particular de gráficos de controle para a média, e a utilização do CMS como um parâmetro utilizado para avaliar o desempenho de gráficos de controle, sendo ambos focos deste trabalho.

2.3.1 Gráficos de Controle para a Média do Processo

Segundo [Abujiya, Riaz e Lee \(2013\)](#), a AAS é o delineamento mais tradicional na construção de gráficos de controle. Porém, os delineamentos apresentados na [seção 2.1](#), baseados em conjuntos ordenados, podem servir como alternativa para a construção de gráficos de controle, dado o ganho em desempenho em relação à AAS.

Os gráficos de controle são compostos por linhas horizontais, sendo elas: uma linha central (LC) e um par de limites (limite inferior de controle (LIC) e limite superior de controle (LSC), além de pontos que representam o valor de alguma estatística calculada em amostras que são coletadas regularmente do processo ([CHANDRA, 2001](#)).

A forma geral da construção do gráfico de controle para a média de um processo, considerando amostras aleatórias simples de tamanho k , é por meio da linha central (LC) e dos limites superior (LSC) e inferior (LIC) de controle, definidos por:

$$LIC = \mu_0 - A\sqrt{Var(\bar{Y}_{AAS})} = \mu_0 - A\sigma_{0\bar{Y}_{AAS}} = \mu_0 - A\frac{\sigma_{0Y_{AAS}}}{\sqrt{k}}, \quad (2.22)$$

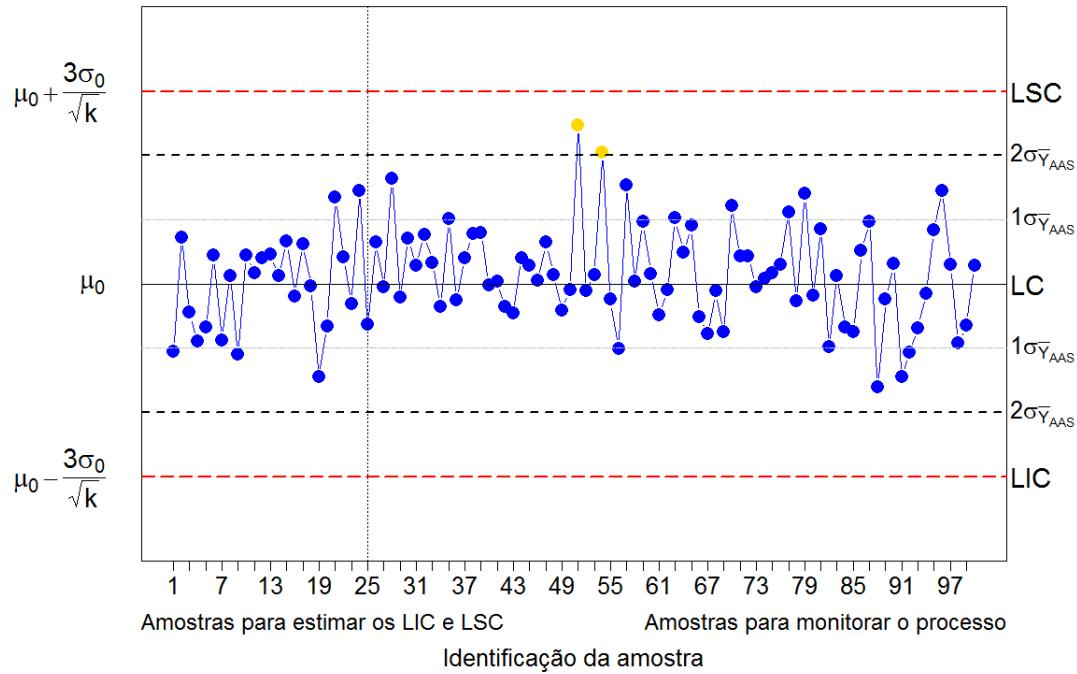
$$LC = \mu_0, \quad (2.23)$$

$$LIC = \mu_0 + A\sqrt{Var(\bar{Y}_{AAS})} = \mu_0 + A\sigma_{0\bar{Y}_{AAS}} = \mu_0 + A\frac{\sigma_{0Y_{AAS}}}{\sqrt{k}}, \quad (2.24)$$

sendo μ_0 e σ_0 a média e o desvio padrão do processo sob controle, \bar{Y}_{AAS} a média de uma amostra de k unidades amostrais extraída via AAS, A o valor correspondente à amplitude do gráfico de controle e $\sigma_{0\bar{Y}_{AAS}}$ o desvio padrão da média de amostras de k unidades amostrais extraída via AAS. Os limites de controle podem ser estimados com base em amostras extraídas do processo operando sob controle. Já na fase de monitoramento, amostras devem ser regularmente extraídas do processo, sendo usual classificar o processo como fora de controle caso a média de uma amostra de k unidades amostrais extrapole os limites de controle, sendo essa a regra mais usual. Essa regra é utilizada nos artigos de [Muttlak e Al-Sabah \(2010\)](#) e [Al-Omari e Haq \(2012\)](#) e também foi adotada neste estudo.

A seguir têm-se a ilustração de dois gráficos de controle para a média. A [Figura 5](#) ilustra um processo sob controle e a [Figura 6](#) um processo fora de controle, em que a média do processo é alterada de μ_0 para $\mu_1 > \mu_0$. Os pontos em azul representam pontos entre

os limites 2σ , pontos em amarelo, entre os limites 2 e 3σ , além dos pontos em vermelho, que representam um ponto além dos limites de controle. Na [Figura 5](#) é visto que os pontos se distribuem ao redor de μ_0 . Já na [Figura 6](#), é visto os pontos se distribuem acima de μ_0 , após a 25^a amostra, mais especificamente, em torno de μ_1 , ou seja, o gráfico de controle sugere que o processo teve uma mudança na média do processo.



[Figura 5](#) – Representação de um gráfico de controle para a média, para um processo sob controle.

Fonte: Os autores

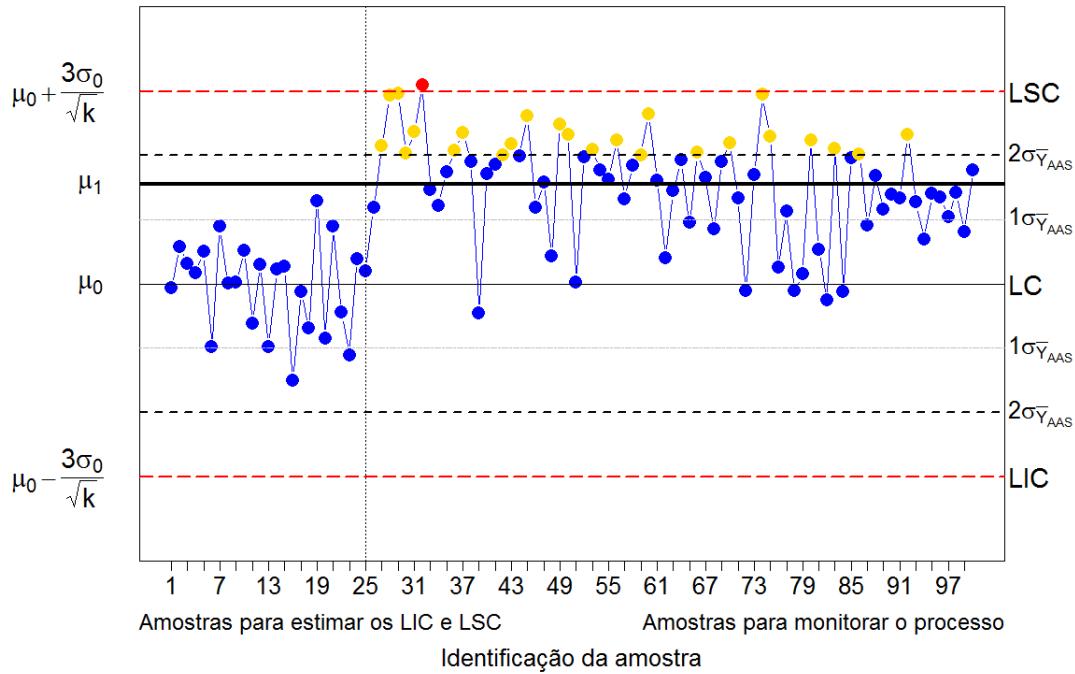


Figura 6 – Representação de um gráfico de controle para a média, considerando um processo fora de controle.

Fonte: Os autores

A literatura relaciona regras complementares para a detecção de um processo fora de controle (as chamadas regras sensibilizantes), como as regras da WESTERN ELECTRIC COMPANY (1956), que consideram possíveis indícios de descontrole do processo: um ponto fora da zona de controle; dois em três pontos consecutivos localizados além dos limites de alerta 2-sigma; quatro em cinco pontos consecutivos localizados a uma distância de um sigma ou mais da linha central ou oito pontos consecutivos localizados de um mesmo lado da linha central (MONTGOMERY, 2011). Ainda há outras regras adicionais, tais como: seis pontos em uma sequência sempre crescente ou decrescente; quinze pontos em sequência dentro dos limites um sigma; quatorze pontos em sequência alternado acima e abaixo da linha central (NELSON, 1984).

Na construção de gráficos de controle para a média, usualmente considera-se $A = 3$, o que implica em uma probabilidade de falso alarme (situação em que a média de uma amostra de tamanho k extrapola os limites de controle, mas o processo opera sob controle) igual a 0,0027 (MONTGOMERY, 2011). Outras regras podem ser consideradas para a determinação do valor de A . Utilizar algum valor maior que 3 para A expandiria os limites de controle, implicando em uma taxa menor de falsos alarmes, mas também dificultaria a detecção da alteração na média do processo. Usar um valor menor para A , por outro lado, estreitaria os limites de controle, o que possibilitaria identificar mais rapidamente uma alteração na média do processo, embora aumentasse a taxa de falsos alarmes. É importante identificar rapidamente uma alteração na média do processo, mas desde que

isso não aumente de forma demasiada a taxa de falsos alarmes, pois isso acarretaria na parada e consequente intervenção frequente no processo, resultando assim em aumento de custos. Deve-se então escolher cuidadosamente o valor de A que acomode adequadamente tais situações.

Propõe-se aqui a construção de gráficos de controle para a média via URSS mediante substituição de $Var(\bar{Y}_{AAS})$ por $Var(\bar{Y}_{URSS})$, sendo \bar{Y}_{URSS} a média de uma amostra de tamanho k obtida via URSS, e pela seleção regular de amostras via URSS e estimação da média usando \bar{Y}_{URSS} . Para os demais delineamentos, o procedimento anteriormente proposto em [Muttlak e Al-Sabah \(2010\)](#) e [Al-Omari e Haq \(2012\)](#) foi similar, mudando apenas a forma de seleção de amostras e o cálculo da variância da média.

2.3.2 Comprimento Médio da Sequência

Uma das formas de avaliar e comparar os desempenhos de gráficos de controle, é via Comprimento Médio da Sequência (CMS). O CMS representa o número médio de amostras antes que alguma delas produza um ponto que exceda os limites de controle. Caso as estatísticas a serem plotadas no gráfico de controle sejam não correlacionadas, o CMS fica determinado por:

$$CMS = \frac{1}{P(\text{um ponto que excede os limites de controle})}. \quad (2.25)$$

No caso em que o processo se encontra sob controle, tem-se:

$$CMS_0 = \frac{1}{\alpha}, \quad (2.26)$$

sendo α a probabilidade do erro tipo I, ou seja, a probabilidade de um ponto cair fora dos limites de controle caso o processo esteja sob controle, configurando um falso alarme. Se $A = 3$, α é igual a 0.0027. Caso o processo saia de controle, de tal forma que a média mude de μ_0 para μ_1 , o CMS pode ser expresso por:

$$CMS_1 = \frac{1}{1 - \beta(\mu_1)}, \quad (2.27)$$

em que $\beta(\mu_1)$ representa a probabilidade do erro tipo II, ou seja, a probabilidade de uma amostra produzir um ponto interno aos limites de controle se a média do processo alterar para $\mu_1 \neq \mu_0$.

2.4 Metodologia do Estudo

Essa seção apresenta a proposta de gráficos de controle para a média via *URSS*, e um estudo que compara o desempenho dos delineamentos citados na [seção 2.1](#) para o monitoramento da média via cartas controle. Além disso, trata do cálculo de $Var(\bar{Y}_{URSS})$, descrita na equação [\(2.10\)](#), nos casos de ordenação perfeita e imperfeita, bem como de sua estimativa.

2.4.1 Proposta para a Utilização do *URSS* para Gráficos de Controle

Dado o contexto de CEP, e apresentação do delineamento *URSS*, propõem-se utilizá-lo para aplicação em gráficos de controle para a média, uma vez que \bar{Y}_{URSS} , apresentado na equação [\(2.9\)](#), possui maior eficiência se comparado aos estimadores correspondentes obtidos via ACO e AAS ([ZAMANZADE; AL-OMARI, 2015](#)). A construção do gráfico de controle para a média de um processo via *URSS*, baseado em amostras de tamanho k , fica proposta por:

$$LIC = \mu_0 - A\sqrt{Var(\bar{Y}_{URSS})} = \mu_0 - A\sigma_{0\bar{Y}_{URSS}}, \quad (2.28)$$

$$LC = \mu_0, \quad (2.29)$$

$$LSC = \mu_0 + A\sqrt{Var(\bar{Y}_{URSS})} = \mu_0 + A\sigma_{0\bar{Y}_{URSS}}, \quad (2.30)$$

sendo μ_0 a média do processo sob controle, \bar{Y}_{URSS} a média de uma amostra de k unidades amostrais extraída via *URSS*, A o valor correspondente à amplitude do gráfico de controle e $\sigma_{0\bar{Y}_{URSS}}$ o desvio padrão da média do processo sob controle de uma amostra de k unidades amostrais extraída via *URSS*.

2.4.2 Estudo Via Simulação do Desempenho dos Gráficos de Controle para a Média Via *URSS*

Para o estudo por simulação, foram considerados diferentes tamanhos de amostra (k), níveis de correlação (ρ) entre a variável de interesse (Y) e a variável concomitante (X) e níveis de descontrole do processo (δ). Os valores de X e Y foram simulados de uma distribuição Normal bivariada, fixando-se $\mu_X = 0$, $\sigma_X = \sigma_Y = 1$ e, configurando-se assim:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim Normal \left(\begin{pmatrix} 0 \\ \mu_Y \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad (2.31)$$

em que μ_Y em um primeiro momento é o próprio μ_0 , configurando assim um processo operando sob controle (nesse caso com média 0). Em um segundo momento, são considerados

processos com níveis crescentes de descontrole, alterando a média para $\mu_Y = \mu_0 + \frac{(\delta\sigma_0)}{\sqrt{k}}$. Essa alteração na média do processo é controlada por:

$$\delta = |\mu_Y - \mu_0| \frac{\sqrt{k}}{\sigma_0}. \quad (2.32)$$

Para $\delta = 0$ tem-se o processo sob controle e, se algum ponto for detectado fora dos limites de controle, é considerado um falso alarme. Desse modo, quanto maior for o CMS para $\delta = 0$, menor a taxa de falsos alarmes. Por outro lado, se $\delta \neq 0$, o processo está fora de controle, e o CMS indicaria o número médio de amostras necessárias para a identificação da alteração na média do processo. Para o estudo por simulação, foram considerados diferentes valores de k , ρ , e δ , sendo eles:

- Tamanhos de amostras ($k = 3, 4, 5, 6$);
- Níveis de correlação entre a variável de interesse e a variável concomitante ($\rho = 0, 25; 0, 50; 0, 75; 0, 9; 1$);
- Níveis de descontrole do processo em relação à média: ($\delta = 0; 0, 1; 0, 2; 0, 3; 0, 4; 0, 8; 1, 2; 1, 6; 2, 2; 4, 3, 2$);

Para cada combinação de k , ρ e δ foram simuladas 1.000.000 de amostras via *URSS*, sendo o CMS, o indicador utilizado para comparar o desempenho dos gráficos de controle para a média via *URSS* aos produzidos pelos demais delineamentos baseados em conjuntos ordenados e à AAS, sendo calculado por:

$$CMS = \frac{1}{p}, \text{ sendo } p = \frac{\text{Número de pontos fora dos limites de controle}}{1.000.000}. \quad (2.33)$$

Foram escolhidos esses valores para os parâmetros da simulação para permitir a comparação dos resultados deste trabalho com os de [Muttlak e Al-Sabah \(2010\)](#), que os utilizaram em um estudo de simulação para avaliar o desempenho dos delineamentos ACO, *ERSS* e *MRSS* na construção de gráficos de controle para a média. Adicionalmente, os valores dos parâmetros utilizados contemplam situações práticas e recorrentes da aplicação dos delineamentos por conjuntos ordenados, como pequenos tamanhos amostrais, diversos níveis de correlação entre as variáveis de interesse e a concomitante, além de diferentes alterações na média do processo. Em particular, quando $\rho = 0$, a ordenação da variável de interesse é feita totalmente ao acaso, e os delineamentos que fazem uso de ordenação equivalem à AAS; já, quando $\rho = 1$ têm-se a ordenação perfeita, sendo $Var(\bar{Y}_{URSS})$, e os limites de controle calculados com base nas propriedades de estatísticas de ordem.

A comparação do *URSS* frente aos *doubles*, considera somente o caso de ordenação perfeita ($\rho = 1$), e amostras de tamanho 3, 4 e 5, que foram os únicos cenários considerados por [Al-Omari e Haq \(2012\)](#). Ainda, os valores da constante A para a determinação dos

limites de controle foram determinados de forma que CMS_0 fosse fixado em 370,508 (taxa de falsos alarmes igual a 0,0027) para todos os delineamentos baseados em conjuntos ordenados, que é um valor muito próximo do CMS_0 da AAS (370,398). Isso foi necessário pois os delineamentos *DERSS*, *DQRSS*, *QDRSS* e *EDRSS* apresentavam taxa de falsos alarmes muito superior às dos demais delineamentos. O CMS do *MRSS* via ordenação perfeita também foi calculado com metodologia similar ao apresentado via *URSS*, também por simulação, respeitando a forma da seleção de amostras e a variância do estimador da média obtido com base nas propriedades das estatísticas de ordem.

Ademais, o CMS via AAS foi calculado analiticamente com base nas propriedades da distribuição da média amostral.

2.4.3 Obtenção de $Var(\bar{Y}_{URSS})$

Para determinar os limites de controle apresentados nas equações (2.22) e (2.24) para o *URSS*, deve-se primeiramente determinar o valor de $Var(\bar{Y}_{URSS})$ para os diferentes cenários considerados no estudo por simulação. Foram utilizados dois diferentes métodos para encontrar o valor de $Var(\bar{Y}_{URSS})$, sendo o primeiro analítico (por meio de propriedades de estatísticas de ordem, no caso de ordenação perfeita) e o segundo via simulação de dados (no caso de ordenação imperfeita).

2.4.3.1 Obtenção de $Var(\bar{Y}_{URSS})$ sob Ordenação Perfeita

Para determinar $Var(\bar{Y}_{URSS})$, apresentada na equação (2.10), as variâncias e covariâncias das estatísticas de ordem foram calculadas utilizando as equações (2.19) e (2.20), com base na função densidade de probabilidade e função densidade de probabilidade conjunta, apresentadas nas equações (2.14) e (2.17) respectivamente. Todos esses cálculos foram implementados no [R Core Team \(2015\)](#), cujos códigos estão disponíveis no [Apêndice B](#) para obter os valores referentes as estatísticas de ordem e no [Apêndice D](#) para determinar $Var(\bar{Y}_{URSS})$ no caso perfeito. O valor de $Var(\bar{Y}_{URSS})$ foi obtido para amostras de tamanho $k = 3, 4, 5$ e 6 .

2.4.3.2 Obtenção de $Var(\bar{Y}_{URSS})$ sob Ordenação Imperfeita

Para o caso em que a ordenação da variável Y baseado nos valores de X ocorreu de forma imperfeita, obteve-se $Var(\bar{Y}_{URSS})$ por meio de simulação, cujos códigos estão disponíveis no [Apêndice C](#). O estudo consistiu em simular 1.000.000 de amostras de uma distribuição normal padrão bivariada ([MEYER, 1970](#)) para cada combinação dos valores de coeficiente de correlação entre a variável de interesse e a concomitante, e tamanhos de amostras. Após gerar as amostras via *URSS* para cada par de valores para k e ρ , ordenava-se a variável de interesse pelos valores da variável concomitante, selecionando-se as unidades amostrais segundo o delineamento *URSS*, de acordo com as unidades ranqueadas nas

posições apresentadas na [Tabela 1](#). Foram calculadas então as variâncias e covariâncias das variáveis ordenadas, necessárias para obter a variância da média estimada via *URSS*, e por fim, obter os limites de controle. Nesse caso, a variância da variável aleatória ordenada na posição $[(i - 1)k + l]$ foi estimada por:

$$\widehat{Var}(Y_{[(i-1)k+l]}) = \frac{\sum_{h=1}^{1.000.000} (Y_{[(i-1)k+l]}^h - \bar{Y}_{[(i-1)k+l]})^2}{1.000.000 - 1}, \quad (2.34)$$

em que,

$$\bar{Y}_{[(i-1)k+l]} = \frac{\sum_{h=1}^{1.000.000} Y_{[(i-1)k+l]}^h}{1.000.000}. \quad (2.35)$$

E seja $Y_{[(i-1)k+l]}^h$ o valor julgado de ordem $[(i - 1)k + l]$ na h -ésima simulação, cujos valores de $[(i - 1)k + l]$ correspondem às posições que definem uma amostra via *URSS*, conforme [Tabela 1](#). E por último, a covariância entre as unidades amostrais nas posições $[(i - 1)k + l]$ e $[(i' - 1)k + l]$:

$$\widehat{Cov}(Y_{[(i-1)k+l]}, Y_{[(i'-1)k+l]}) = \frac{\sum_{h=1}^{1.000.000} (Y_{[(i-1)k+l]}^h - \bar{Y}_{[(i-1)k+l]})(Y_{[(i'-1)k+l]}^h - \bar{Y}_{[(i'-1)k+l]})}{1.000.000 - 1}. \quad (2.36)$$

Uma vez que a variância e a covariância nas equações [\(2.34\)](#) e [\(2.36\)](#) foram obtidas, tais valores foram substituídos na equação [2.10](#), obtendo-se assim o valor de $Var(\bar{Y}_{URSS})$ para uma amostra de tamanho (k) e nível de correlação entre a variável de interesse e concomitante (ρ). Logo após, esses valores foram utilizados para obter os limites de controle, conforme apresentado na [Seção 2.4.1](#).

2.4.4 Proposta e Estudo do Viés de um Estimador de $Var(\bar{Y}_{URSS})$ Aplicado à Construção de Gráficos de Controle

Na maioria de situações práticas, não se sabe o valor verdadeiro dos parâmetros da distribuição da estatística de interesse, de tal forma que eles precisam ser estimados. Uma prática comum para a construção de gráficos de controle é usar os resultados de repetidas (B) amostras do processo operando sob controle para estimação dos limites ([TRIETSCH, 1999](#)). Em particular, no caso de gráficos de controle para a média via *URSS*, isso requer a estimação de $Var(\bar{Y}_{URSS})$, conforme equação [\(2.10\)](#), usados na construção dos limites de controle. Seja $Y_{[(i-1)k+l]j}$ a unidade amostral julgada $[(i - 1)k + l]$, do j -ésimo ciclo oriunda do *URSS*. O estimador proposto para $Var(\bar{Y}_{URSS})$ é dado por:

$$\widehat{Var}(\bar{Y}_{URSS}) = \frac{1}{nk^2} \sum_{j=1}^n \sum_{i=1}^k \widehat{Var}(Y_{[(i-1)k+l]j}) + \frac{2}{nk^2} \sum_{j=1}^n \sum_{i < i'}^k \widehat{Cov}(Y_{[(i-1)k+l]j}, Y_{[(i'-1)k+l]j}), \quad (2.37)$$

sendo

$$\widehat{Var}(Y_{[(i-1)k+l]}) = \sum_{b=1}^B \frac{(Y_{[(i-1)k+l]}^b - \bar{Y}_{[(i-1)k+l]})^2}{nk - 1}, \quad (2.38)$$

sendo

$$\bar{Y}_{[(i-1)k+l]} = \frac{\sum_{b=1}^B Y_{[(i-1)k+l]}^b}{B}, \quad (2.39)$$

e

$$\widehat{Cov}(Y_{[(i-1)k+l]}, Y_{[(i'-1)k+l]}) = \sum_{b=1}^B (Y_{[(i-1)k+l]}^b - \bar{Y}_{[(i-1)k+l]})(Y_{[(i'-1)k+l]}^b - \bar{Y}_{[(i'-1)k+l]}), \quad (2.40)$$

em que $Y_{[(i-1)k+l]}^b$ representa a unidade amostral julgada $[(i-1)k+l]$, da b -ésima quantidade de amostras. Importante ressaltar a diferença entre B e n . Enquanto n é utilizado para denotar o número de ciclos, utilizados em casos que a título de exemplo, para se obter uma amostra de tamanho $k = 10$, pode-se utilizar dois ciclos, resultando por cada ciclo uma amostra de tamanho $k = 5$, e no final uma amostra de tamanho $k = 10$. A notação B , por sua vez, é utilizada para denotar quantas amostras estão sendo utilizadas para estimar os parâmetros do processo operando sob controle, com base no monitoramento de amostras de tamanho k ; caso fossem retiradas 10 amostras de tamanho 3, teria-se nesse caso, o valor de $B = 10$ e $k = 3$.

Um estudo de simulação foi realizado para avaliar o viés do estimador proposto $\widehat{Var}(\bar{Y}_{URSS})$, conforme equação (2.37). Foram realizadas 50.000 simulações, cada uma delas compostas por B amostras ($B = 5, 10, 15, 20, 25, 30, 50, 100$) obtidas via URSS de uma distribuição normal bivariada padrão (MEYER, 1970), de tamanho $k = 3$ e correlação entre as variáveis de interesse e concomitante $\rho = 0, 5$. A variável de interesse foi ordenada com base na variável concomitante, para em um segundo momento, selecionar as unidades amostrais correspondentes para $k = 3$ via URSS (cujas posições podem ser encontradas na Tabela 1), para então estimar $\widehat{Var}(\bar{Y}_{URSS})$, conforme equação (2.37). Ao final, foi calculado o viés relativo do estimador em relação ao verdadeiro valor do parâmetro, com base nas 50.000 simulações. A expressão para o viés relativo, calculado para cada valor de B , é dado por:

$$\text{Viés}_{\text{relativo}}(\widehat{Var}(\bar{Y}_{URSS})) = \frac{\text{Viés}(\widehat{Var}(\bar{Y}_{URSS}))}{Var(\bar{Y}_{URSS})}, \quad (2.41)$$

sendo

$$\text{Viés}(\widehat{\text{Var}}(\bar{Y}_{URSS})) = E[\widehat{\text{Var}}(\bar{Y}_{URSS})] - \text{Var}(\bar{Y}_{URSS}), \quad (2.42)$$

sendo

$$\hat{E}[\widehat{\text{Var}}(\bar{Y}_{URSS})] = \frac{\sum_{h=1}^{50.000} \widehat{\text{Var}}(\bar{Y}_{URSS}^h)}{50.000}, \quad (2.43)$$

e o valor de $\text{Var}(\bar{Y}_{URSS})$ para $k = 3$ e $\rho = 0,5$ foi obtido pelo procedimento proposto na Seção 2.4.3, resultando em 0,280864483557273.

Os parâmetros ($k = 3$ e $\rho = 0.5$) foram escolhidos para o estudo por simulação, pois representam um dos cenários sujeitos a maior variação sob estudo, uma vez que se considera o menor tamanho de amostra e um nível de correlação moderado.

2.5 Materiais

Além de realizar um estudo por simulação para comparação dos delineamentos foi feita também a simulação da aplicação do delineamento *URSS*, seguida de uma comparação com a AAS e a ACO, na construção de gráficos de controle para a média a partir de uma base de dados real.

2.5.1 Conjunto de Dados

Para a ilustração da aplicação dos gráficos de controle para a média, foram considerados os delineamentos AAS, ACO e *URSS*, utilizando-se um conjunto de dados de construção civil disponibilizado por Yeh (1998). O conjunto de dados contém 1030 observações e 9 variáveis. Os dados foram coletados com o objetivo de avaliar a resistência do concreto de alto desempenho, cuja medição, segundo Aïtcin (2000), se dá pela pressão que a máquina utiliza para destruir um bloco de concreto (dependendo da tecnologia da máquina, não é necessário a destruição do bloco), registrada em megapascal (MPa). Todas as outras 8 variáveis são as quantidades dos insumos para a produção do concreto, sendo cimento, água, pedra e areia as principais, além de alguns componentes químicos, conhecidos por aditivos graúdos e miúdos. Entre todas as variáveis utilizadas para a produção de concreto, o cimento foi a variável mais correlacionada com a variável de interesse, sendo ela então escolhida para ser a variável concomitante, responsável por ordenar os valores da resistência do concreto.

Um estudo experimental de produção de blocos de concreto passa por várias fases. Em uma primeira, é necessário pesar as matérias primas, para em seguida misturá-las,

respeitando a ordem correta de inserção delas para a obtenção do concreto. Após isso, a mistura formada é despejada em um molde de formato cilíndrico de forma que fique adensado. Os moldes devem permanecer em repouso, isso é, sem nenhuma pressão, por pelo menos 12 horas, sendo em seguida, levados a um tanque úmido, onde passarão por um processo de cura, por quantos dias forem necessários, dependendo da necessidade do ensaio (3, 7, 28 dias - (KLOSS, 1991)). O bloco de concreto é então retirado e submetido a um teste de resistência, conforme a [Figura 7](#).

Para se mensurar a resistência do concreto, o processo de cura e consequente teste de resistência demanda certo tempo, sendo ainda, necessária a destruição do bloco de concreto, dependendo da máquina utilizada, para a mensuração, inviabilizando a utilização do item posteriormente. Sendo assim, uma vez que para obter amostras demanda-se de esforços extras (dentre eles tempo de cura e a própria mensuração), e dispõe-se de uma variável concomitante (razoavelmente) correlacionada com a resistência do concreto, os delineamentos baseados em conjuntos ordenados são alternativas à AAS e, em particular, o *URSS* pode ser utilizado para estimar a resistência média de concreto.



Figura 7 – Representação da mensuração da resistência do concreto.

Fonte: [Youtube \(2009\)](#)

2.5.2 Descrição da Aplicação do *URSS*, ACO e a AAS no Conjunto de Dados

Para o uso dos gráficos de controle para monitorar a resistência média do concreto foram considerados três diferentes delineamentos, AAS, ACO e o *URSS*. Considerou-se também diferentes tamanhos de amostra, $k = 3$ e 5 , além de diferentes deltas, $\delta = 0; 0,8$ e $1,2$, conforme descrito em equação (2.32). Para cada delineamento, inicialmente, foram selecionadas 25 amostras com reposição a partir dos dados originais (um tamanho de cada vez) para estimar os limites de controle, e, posteriormente, 75 amostras de mesmos

tamanhos para realizar o monitoramento da média do processo. Para $\delta = 0$, ou seja, na situação em que o processo opera sob controle, as 75 amostras posteriores foram selecionadas com reposição dos dados originais; para $\delta = 0,8$ ou $1,2$, foi simulado um descontrole ao processo, de forma que foi somado aos valores da resistência de concreto uma variável aleatória normal com média igual a $\frac{\delta\sigma_0}{\sqrt{k}}$ e desvio padrão igual a 2. O valor do desvio padrão foi escolhido cuidadosamente de tal forma que fosse suficientemente pequeno para caracterizar o descontrole, predominantemente, em virtude da alteração na locação (média) do processo, em detrimento à sua dispersão (variância).

A construção dos gráficos de controle para essa ilustração, ocorreu de forma que, cada amostra é representada pela sua média por um ponto no gráfico. Pontos azuis representam as amostras cujas médias cairam entre os limites 2 sigmas; pontos amarelos indicam pontos de atenção e estão entre os limites de 2 e 3 sigmas (LSC e LIC) e, por sua vez, pontos vermelhos representam amostras que cairam fora dos limites de controle, e são indícios de descontrole do processo. As linhas horizontais indicam os limites quando A têm valores iguais a 3, 2, 1, 0, -1, -2 e -3. Ainda, em particular para o URSS, utilizou-se o estimador descrito na seção 2.4.3 para obter a variância da média, e consequentemente, o cálculo das linhas horizontais dos gráficos. Para a AAS e a ACO, os limites foram estimados utilizando as respectivas expressões dos estimadores da variância da média de cada delineamento. Os códigos utilizados nessa etapa do estudo são apresentados no Apêndice H.

2.5.3 Recursos Computacionais

O software utilizado para realizar as simulações e a aplicação do delineamento é o R Core Team (2015), além do pacote MASS (VENABLES; RIPLEY, 2002) para gerar dados de uma distribuição normal bivariada.

3 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados do estudo por simulação, referentes ao *URSS* frente aos delineamentos e a aplicação do *URSS* em gráficos de controle para a média, além dos resultados do estudo do viés do estimador para a variância da média proposto do *URSS*.

3.1 Desempenho e Comparaçāo dos Gráficos de Controle entre os Delineamentos Apresentados

Na sequência são apresentados as tabelas para comparação dos desempenhos de gráficos de controle sob diferentes delineamentos, apresentando os respectivos valores do CMS's das simulações dos métodos citados na Seção 2.4.2. Para cada combinação de k , δ e ρ , foram realizadas 1.000.000 de simulações.

Nas Tabela 2 a 4, são apresentados os resultados para comparação dos CMS dos delineamentos AAS, ACO, *ERSS*, *MRSS*, *DERSS*, *DQRSS*, *QDRSS*, *EDRSS* e *URSS* sob ordenação perfeita. Os valores do CMS via AAS, *MRSS* e *URSS* foram calculados pelos autores e os valores dos demais delineamentos foram extraídos do estudo de simulação realizado por Al-Omari e Haq (2012). Para os delineamentos baseados em conjuntos ordenados, o valor de CMS_0 foi fixado em 370,508 para a construção destas tabelas.

É visto na Tabela 2, para $k=3$, que o *URSS* ganha de todos os delineamentos em todos os cenários, inclusive dos *doubles*, cujos delineamentos usam k^3 unidades e dois estágios de ordenação. A título de ilustração, para $\delta = 0,2$ o CMS gerado pelo *URSS* é de 234 enquanto que para o *EDRSS* é de 250 e de 266 para a ACO, que foram os melhores dentre os *doubles* e os que utilizam apenas um estágio de ordenação, respectivamente; ainda o CMS gerado via AAS, no mesmo cenário, resultou em 308,43. Para $\delta = 0,4$, ainda, o *URSS* produziu um CMS 26,4% menor, em relação à ACO, e 48,7% menor em relação à AAS. Ainda, para $\delta = 0,8$, nota-se que é necessário, em média, amostrar aproximadamente mais que o triplo de unidades via AAS ($CMS \approx 71$) do que se utilizar o *URSS* ($CMS = 21,25$). Para todos os outros cenários, embora não esmiuçados aqui, o ganho em desempenho do *URSS* frente aos demais delineamentos é constante.

Conforme Tabela 3, para $k=4$, novamente, os diversos CMS gerados pela *URSS* foram menores para todos os cenários, contra todos os delineamentos considerados (inclusive os *doubles*). Entre $0,1 \leq \delta \leq 0,4$, observa-se que se o delineamento escolhido for o *URSS*, é necessário de no mínimo 7 amostras em média a menos (CMS mais próximo do CMS via *URSS*, foi via *MRSS*, com $\delta = 0,1$) para a detecção de uma alteração na média do

Tabela 2 – Valores do CMS para gráficos de controle para a média, para amostras de tamanho $k=3$, utilizando AAS, ACO, ERSS, DERSS, EDRSS, DQRSS, QDRSS, MRSS e URSS, considerando ordenação perfeita

δ	AAS*	ACO	ERSS	MRSS*	DERSS	EDRSS	DQRSS	QDRSS	URSS*
0	370,40	370,51	370,51	370,51	370,51	370,51	370,51	370,51	370,51
0,1	352,93	333,89	340,25	339,67	337,72	338,75	333,00	332,67	325,63
0,2	308,43	266,03	272,18	265,11	256,67	250,00	251,89	254,07	234,03
0,3	253,14	196,93	197,20	186,22	175,62	172,89	170,94	179,53	157,23
0,4	200,08	139,43	137,99	128,12	117,83	115,51	117,19	119,56	102,60
0,8	71,55	35,43	35,35	29,52	26,11	25,61	26,04	26,42	21,25
1,2	27,82	11,54	11,43	9,22	7,77	7,63	7,82	7,83	6,41
1,6	12,38	4,76	4,75	3,80	3,17	3,13	3,18	3,20	2,76
2	6,30	2,50	2,49	2,06	1,75	1,74	1,76	1,76	1,61
2,4	3,65	1,61	1,61	1,40	1,26	1,25	1,26	1,26	1,20
3,2	1,73	1,09	1,09	1,04	1,02	1,02	1,02	1,02	1,01

Fonte: [Al-Omari e Haq \(2012\)](#).

* Os valores do CMS dos delineamentos AAS, MRSS e URSS, foram calculados pelos autores, com base nas propriedades da distribuição da média amostral (para a AAS) e via simulação (MRSS e URSS), sendo que a variância dos estimadores do MRSS e URSS foram obtidos com base nas propriedades de estatística de ordem.

Tabela 3 – Valores do CMS para gráficos de controle para a média, para amostras de tamanho $k=4$, utilizando AAS, ACO, ERSS, DERSS, EDRSS, DQRSS, QDRSS, MRSS e URSS, considerando ordenação perfeita

δ	AAS*	ACO	ERSS	MRSS*	DERSS	EDRSS	DQRSS	QDRSS	URSS*
0	370,40	370,51	370,51	370,51	370,51	370,51	370,51	370,51	370,51
0,1	352,93	328,08	341,30	318,07	323,63	334,56	329,60	327,01	310,56
0,2	308,43	249,81	266,81	232,45	246,73	251,26	242,72	249,19	210,30
0,3	253,14	174,89	192,64	156,42	168,55	173,16	167,81	169,78	126,90
0,4	200,08	119,36	135,85	100,29	111,45	116,69	111,63	115,01	77,86
0,8	71,55	27,78	33,69	21,42	24,15	25,77	24,26	25,26	13,89
1,2	27,82	8,54	10,70	6,38	7,19	7,59	7,20	7,45	4,09
1,6	12,38	3,55	4,41	2,73	2,99	3,08	2,99	3,06	1,89
2	6,30	1,94	2,33	1,59	1,69	1,71	1,69	1,70	1,25
2,4	3,65	1,35	1,53	1,19	1,23	1,23	1,23	1,23	1,06
3,2	1,73	1,03	1,07	1,01	1,02	1,02	1,01	1,02	1,00

Fonte: [Al-Omari e Haq \(2012\)](#).

* Os valores do CMS dos delineamentos AAS, MRSS e URSS, foram calculados pelos autores, com base nas propriedades da distribuição da média amostral (para a AAS) e via simulação (MRSS e URSS), sendo que a variância dos estimadores do MRSS e URSS foram obtidos com base nas propriedades de estatística de ordem.

processo, e no máximo, de 126 amostras a menos em média (caso mais distante do CMS via URSS, foi via AAS, com $\delta = 0,3$). Mais especificamente para $\delta = 0,3$, por exemplo, o CMS gerado via URSS (CMS = 126,90) foi aproximadamente 50% menor que o dos resultados obtidos via AAS (CMS = 253,14) e aproximadamente 34% menor que os obtidos

via *ERSS* ($CMS = 192.64$). Ainda, para $\delta = 1, 2$ o CMS produzido via *URSS* ($CMS=4,09$) é inferior em relação ao delineamento *MRSS* ($CMS=6,38$), que foi recomendado para aplicação no estudo de [Muttlak e Al-Sabah \(2010\)](#), e *DERSS* ($CMS=7,19$), que teve o melhor desempenho entre os doubles para o respectivo valor de delta.

Tabela 4 – Valores do CMS para gráficos de controle para a média, para amostras de tamanho $k=5$, utilizando AAS, ACO, *ERSS*, *DERSS*, *EDRSS*, *DQRSS*, *QDRSS*, *MRSS* e *URSS*, considerando ordenação perfeita

δ	AAS*	ACO	<i>ERSS</i>	<i>MRSS</i> *	<i>DERSS</i>	<i>EDRSS</i>	<i>DQRSS</i>	<i>QDRSS</i>	<i>URSS</i> *
0	370,40	370,51	370,51	370,51	370,51	370,51	370,51	370,51	370,51
0,1	352,93	331,68	333,00	329,60	335,01	317,76	259,54	291,12	299,58
0,2	308,43	244,98	254,77	223,41	243,61	226,09	128,52	160,90	181,06
0,3	253,14	165,54	173,73	136,91	159,49	146,09	60,58	83,67	104,59
0,4	200,08	107,88	117,44	85,20	101,00	93,88	30,64	45,22	60,14
0,8	71,55	22,53	26,59	15,56	19,47	18,18	4,02	6,37	9,55
1,2	27,82	6,73	8,13	4,55	5,55	5,17	1,45	1,99	2,86
1,6	12,38	2,83	3,38	2,04	2,35	2,21	1,05	1,18	1,46
2	6,30	1,63	1,87	1,31	1,42	1,36	1,00	1,02	1,10
2,4	3,65	1,21	1,32	1,08	1,12	1,09	1,00	1,00	1,01
3,2	1,73	1,01	1,03	1,00	1,00	1,00	1,00	1,00	1,00

Fonte: [Al-Omari e Haq \(2012\)](#).

* Os valores do CMS dos delineamentos AAS, *MRSS* e *URSS*, foram calculados pelos autores, com base nas propriedades da distribuição da média amostral (para a AAS) e via simulação (*MRSS* e *URSS*), sendo que a variância dos estimadores do *MRSS* e *URSS* foram obtidos com base nas propriedades de estatística de ordem.

Ainda na [Tabela 4](#), observa-se um melhor desempenho do *URSS* em relação a todos os outros delineamentos em 60 dos 80 cenários. Nos únicos cenários em que o *URSS* perde, é ao considerar os *doubles DQRSS* e *QDRSS*, indicando que para esse tamanho de amostra tais delineamentos acabaram detectando mais rapidamente o processo fora de controle do que o *URSS*. Mas em relação aos demais delineamentos, os valores dos CMS obtidos via *URSS* foram menores, contra os outros dois *doubles*, o *DERSS* e *EDRSS*. Por exemplo, observa-se que para $\delta = 0,4$, o *URSS* ($CMS=60,14$) detecta com menor número médio de unidades amostrais um descontrole que o *DERSS* ($CMS=101,00$) e o *EDRSS* ($CMS = 93,88$).

No geral, sob ordenação perfeita, o *URSS* se mostrou mais eficiente em 220 dos 240 (91,67%) cenários em relação a todos os delineamentos apresentados nas Tabelas [2](#) a [4](#). Ao compará-lo especificamente com a AAS, o *URSS* foi mais eficiente em todos os cenários estudados, detectando mais rapidamente um processo fora de controle (o *URSS* chegou a detectar mais rapidamente uma alteração na média do processo que a AAS em até 9 vezes). Em relação aos delineamentos ACO, *MRSS*, *ERSS*, o *URSS* também teve uma melhor eficiência, detectando, em todos os cenários, uma possível falha no processo

com menor número médio de unidades amostrais; comparando-o exclusivamente com os *doubles*, o *URSS* ganha ainda em 100 dos 120 (83,3%) casos. Ainda para todos cenários o *URSS* teve CMS inferior em relação aos CMS obtidos via *MRSS*, que foi recomendado no estudo realizados por [Muttlak e Al-Sabah \(2010\)](#) para construção de gráficos de controle para a média, pois apresentaram CMS menor em relação aos demais delineamentos por comparados pelos autores (*AAS*, *ACO* e *ERSS*), o que reduziria o número médio de amostras para detectar um possível descontrole.

Entre as Tabelas 2 a 4, foi visto os diferentes valores de CMS's para diferentes tamanhos de amostra (k). É visto que quanto maior o tamanho de amostra, menor o valor de CMS produzido pelos diferentes delineamentos baseados em conjuntos ordenados, detectando assim, mais rapidamente, uma alteração na média do processo. Por exemplo, para $\delta = 0,8$, os CMS's produzidos via *URSS* foram iguais a 21,25, 13,89 e 9,5 para $k = 3, 4$ e 5 respectivamente. Para $\delta = 1,6$, os CMS's produzidos via *ACO* foram iguais a 4,76, 3,55 e 2,83 para $k = 3, 4$ e 5 respectivamente. A mesma comparação é válida para os demais delineamentos baseados em conjuntos ordenados e deltas.

A importância de se detectar uma alteração na média do processo é mais evidente quando se tem um problema prático em questão. Suponha que uma amostra seja retirada da produção para realizar o monitoramento da média do processo a cada 1 hora. Tomando-se como exemplo a [Tabela 3](#) e $\delta = 1,2$, levaria, em média, 28 horas para detectar uma alteração na média do processo caso o delineamento escolhido fosse a *AAS* (pois CMS = 27,82), enquanto que levaria 6 horas em média via *MRSS*, e 4 horas via *URSS*. Ficando claro nessa situação, a escolha pelo *URSS*.

Ainda, interpretações e comparações adicionais podem ser realizadas para os resultados obtidos. Um intervalo de confiança para p , descrito na equação (2.33), pode ser obtido via Teorema do Limite Central, para a proporção de pontos além dos limites de controle. Partindo-se do resultado com CMS = 13,89 para uma amostra de tamanho $k=4$ obtida via *URSS* sob ordenação perfeita e $\delta = 0,8$, o verdadeiro valor do parâmetro está contido no intervalo entre 13,793 e 13,986 ao nível de 95% de confiança. [Abujiya e Lee \(2013\)](#) compararam os resultados de CMS produzidos via *ACO* contra à *AAS* para gráficos de controle para a média de três diferentes tipos, incluindo gráficos de controle *Shewhart*, gráfico da Soma Cumulativa (*CUSUM*) e Médias Móveis Exponencialmente Ponderadas (*EWMA*). Considerando-se um tamanho de amostra de amostra igual a 3, e que a ordenação da variável resposta foi feita sem erros, para um $\delta = 2$, o CMS produzido via *ACO* foi igual a 2,7 e 2,9 para o *CUSUM* e *EWMA*, respectivamente; contra CMS = 3,9 e 4 obtido via *AAS* para os gráficos do tipo *CUSUM* e *EWMA*, respectivamente; e um CMS = 1,61 obtido via *URSS* para o gráfico de controle tipo *Shewhart*. É visto que o CMS obtido via *URSS* para gráficos do tipo *Shewhart* foi menor, mesmo se comparar esses gráficos com os de *CUSUM* e *EWMA*, que foram elaborados com o objetivo de detectar uma alteração na média do processo mais rapidamente do que o de *Shewhart*. Demais

comparações devem ser realizadas, ficando a cargo do leitor.

As Tabelas 5 a 8 apresentam os resultados dos valores do CMS obtidos via *URSS* sob ordenação imperfeita, comparando com os resultados obtidos analiticamente via AAS e por simulação via ACO, cujos resultados foram extraídos do estudo realizado por [Muttlak e Al-Sabah \(2010\)](#).

Na [Tabela 5](#) observa-se que para amostras de tamanho $k = 3$ e processo sob controle, os diferentes CMS gerados via AAS são de maior magnitude que todos os CMS obtidos via ACO para todos os valores ρ , o que mostra que a AAS foi mais conservadora em relação à ACO, produzindo então uma menor taxa de falsos alarmes; já em relação ao *URSS*, a AAS foi levemente superior e inferior para diferentes valores de ρ . Em 98 dos 100 cenários considerados (desconsiderando $\delta = 0$), os CMS's obtidos via ACO e *URSS* foram menores em relação aos obtidos via AAS. O *URSS* teve um melhor desempenho, comparado com a ACO, em 46 dos 50 cenários. Por exemplo, para $\delta = 0,4$ e $\rho = 0,90$ o *URSS* ($CMS=126,44$) necessita, em média, de 17 unidades amostrais a menos que a ACO ($CMS=148,34$) para detectar um descontrole no processo.

Para amostras de tamanho $k = 4$, vide [Tabela 6](#), a taxa de falsos alarmes do *URSS* e ACO se mantiveram próximas da taxa de falso alarmes da AAS (370,40), indicando que os delineamentos não apresentaram taxas de falsos alarmes muito diferentes, e portanto, as comparações entre os demais níveis de alteração do processo podem ser realizadas. Para os cenários em que se considerou processos operando fora de controle, o CMS produzido via AAS foi maior que o CMS da ACO em todos os cenários; o *URSS* teve um CMS menor em 48 dos 50 cenários, perdendo somente para $\delta = 0,1$ e $0,2$, e $\rho = 0,25$, no entanto, para esse valor de ρ , os delineamentos baseados em conjuntos ordenados não são nem recomendados, pois a perda em eficiência é grande dado pequeno valor de $\rho = 0,25$. Comparando o *URSS* em relação à ACO, o *URSS* gera menor CMS em 43 dos 50 cenários. Observa-se também que para $\delta = 0,3$ e $\rho = 1$, em média, o *URSS* precisa de 38 unidades amostrais a menos que a ACO para detectar um possível descontrole do processo. Um dos únicos casos em que o *URSS* perde para a ACO, foi para $\rho = 0,25$ (situações que delineamentos baseados em conjuntos ordenados não são recomendados) e com graus de descontrole igual ou inferior a 0,3 .

A [Tabela 7](#) apresenta os resultados para $k=5$. Os delineamentos ACO e *URSS* ganham praticamente em todos os cenários da AAS. Comparando o *URSS* com a ACO, para $\rho = 0,75$ e $\delta \geq 0,4$ a ACO requer menor número médio de amostras para detectar o descontrole em relação ao *URSS*. Mas para todas as demais combinações de ρ e δ , os resultados obtidos via *URSS* foram melhores que os obtidos via ACO. Por exemplo, observa-se que para $\delta = 0,3$ e $\rho = 1$ o *URSS* chegou a detectar um possível descontrole do processo com 48 unidades amostrais, em média, a menos que a ACO.

A [Tabela 8](#) mostra os valores obtidos do CMS para $k=6$. Para todos dos cenários considerados, a ACO e o *URSS* foram mais eficientes que a AAS. Em 48 dos 50 cenários

Tabela 5 – CMS para gráficos de controle para a média baseados em amostras de tamanho $k = 3$, obtidos via simulação sob ordenação imperfeita.

δ	AAS*	$\rho=0,25$			$\rho=0,50$			$\rho=0,75$			$\rho=0,90$			$\rho=1$
		ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	
0	370,40	370,09	375,66	369,99	389,86	367,21	363,64	357,83	362,32	343,20	378,21			
0,1	352,93	354,25	346,02	342,18	355,11	344,07	329,82	338,68	326,37	326,80	333,00			
0,2	308,43	304,24	306,47	305,72	303,03	288,57	278,01	266,04	258,73	251,37	232,99			
0,3	253,14	249,55	245,46	238,74	241,49	217,42	211,69	197,23	179,86	175,48	154,01			
0,4	200,08	196,09	191,35	186,52	184,60	167,56	157,48	148,34	126,44	130,87	103,10			
0,8	71,55	69,58	68,11	62,87	59,69	51,17	43,59	41,51	31,15	34,09	21,61			
1,2	27,82	26,75	26,27	23,48	22,26	18,13	15,10	14,09	9,93	11,14	6,46			
1,6	12,38	11,83	11,58	10,24	9,60	7,73	6,37	5,92	4,16	4,65	2,76			
2	6,30	6,02	5,91	5,19	4,87	3,93	3,26	3,05	2,23	2,45	1,61			
2,4	3,65	3,49	3,43	3,04	2,86	2,36	2,01	1,90	1,49	1,60	1,20			
3,2	1,73	1,67	1,65	1,52	1,46	1,30	1,19	1,16	1,06	1,08	1,01			

Fonte: [Muttlak e Al-Sabah \(2010\)](#).

* Os valores do CMS dos delineamentos AAS e URSS, foram calculados pelos autores.

Nota: Valores AAS foram obtidos analiticamente, com base nas propriedades da distribuição da média amostral.

Tabela 6 – CMS para gráficos de controle para a média baseados em amostras de tamanho $k = 4$, obtidos via simulação sob ordenação imperfeita.

δ	AAS*	ACO	URSS*	$\rho=0,25$		$\rho=0,50$		$\rho=0,75$		$\rho=0,90$		$\rho=1$	
				ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*
0	370,40	368,89	353,48	369,53	362,06	365,84	375,94	359,82	372,44	343,37	378,50		
0,1	352,93	349,92	365,36	349,16	348,43	340,64	342,47	327,96	333,56	310,52	314,76		
0,2	308,43	304,57	313,38	299,23	293,51	281,79	281,93	262,34	245,64	237,41	204,67		
0,3	253,14	249,64	250,00	239,12	229,10	216,97	203,13	192,37	170,62	166,55	129,02		
0,4	200,08	196,38	193,65	184,02	177,78	161,02	148,28	135,40	110,07	112,80	77,56		
0,8	71,55	68,68	67,66	60,92	57,44	46,80	40,60	35,34	24,86	26,38	13,94		
1,2	27,82	26,48	26,16	22,62	21,31	16,22	13,29	11,49	7,67	8,17	4,09		
1,6	12,38	11,78	11,60	9,82	9,14	6,88	5,59	4,81	3,23	3,44	1,88		
2	6,30	5,96	5,85	4,98	4,67	3,52	2,89	2,53	1,82	1,90	1,25		
2,4	3,65	3,46	3,41	2,92	2,75	2,14	1,82	1,64	1,29	1,33	1,06		
3,2	1,73	1,66	1,64	1,48	1,42	1,23	1,14	1,09	1,02	1,03	1,00		

Fonte: [Muttlak e Al-Sabah \(2010\)](#).

* Os valores do CMS dos delineamentos AAS e URSS, foram calculados pelos autores.

Nota: Valores AAS foram obtidos analiticamente, com base nas propriedades da distribuição da média amostral.

Tabela 7 – CMS para gráficos de controle para a média baseados em amostras de tamanho $k = 5$, obtidos via simulação sob ordenação imperfeita.

δ	AAS*	$\rho=0,25$			$\rho=0,5$			$\rho=0,75$			$\rho=0,90$			$\rho=1$
		ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	
0	370,40	369,73	359,20	369,96	371,89	367,01	372,44	360,68	379,51	347,11	379,65			
0,1	352,93	354,64	341,41	347,28	346,50	340,64	345,18	327,95	324,78	300,66	299,85			
0,2	308,43	306,15	303,03	299,47	293,00	280,38	278,09	255,16	236,07	226,96	181,88			
0,3	253,14	249,98	246,43	237,32	230,63	212,80	197,63	182,35	155,13	153,63	102,83			
0,4	200,08	195,90	192,98	183,17	181,39	125,62	142,51	115,51	101,27	98,45	59,59			
0,8	71,55	69,63	66,90	59,63	57,01	30,98	37,25	26,71	21,06	21,02	9,61			
1,2	27,82	26,33	25,89	21,99	20,65	9,84	12,27	6,90	6,35	6,35	2,87			
1,6	12,38	11,65	11,45	9,54	8,95	4,12	5,11	3,47	2,72	2,72	1,46			
2	6,30	5,92	5,78	4,84	4,54	2,21	2,67	2,10	1,59	1,59	1,10			
2,4	3,65	3,44	3,38	2,85	2,68	1,49	1,71	1,48	1,19	1,19	1,01			
3,2	1,73	1,65	1,63	1,45	1,40	1,08	1,11	1,06	1,01	1,01	1,00			

Fonte: [Muttlak e Al-Sabah \(2010\)](#).

* Os valores do CMS dos delineamentos AAS e URSS, foram calculados pelos autores.

Nota: Valores AAS foram obtidos analiticamente, com base nas propriedades da distribuição da média amostral.

Tabela 8 – CMS para gráficos de controle para a média baseados em amostras de tamanho $k = 6$, obtidos via simulação sob ordenação imperfeita.

δ	AAS*	$\rho=0,25$			$\rho=0,5$			$\rho=0,75$			$\rho=0,90$			$\rho=1$
		ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	ACO	URSS*	
0	370,40	370,58	365,36	370,08	373,55	369,31	364,96	364,90	369,41	349,00	349,00	375,80	375,80	
0,1	352,93	351,55	339,56	350,35	344,35	341,89	337,04	328,51	318,07	304,54	304,54	287,36	287,36	
0,2	308,43	307,10	307,79	296,99	295,51	278,92	268,53	250,10	234,41	214,01	214,01	163,85	163,85	
0,3	253,14	251,51	246,97	236,85	233,97	209,33	194,29	174,20	146,82	137,56	137,56	89,08	89,08	
0,4	200,08	196,66	192,94	180,92	177,97	150,53	132,91	118,24	93,08	87,86	87,86	48,91	48,91	
0,8	71,55	68,51	67,48	58,86	55,86	41,77	35,09	27,84	18,73	17,28	17,28	7,23	7,23	
1,2	27,82	26,22	25,74	21,58	20,45	14,08	11,44	8,73	5,58	5,13	5,13	2,23	2,23	
1,6	12,38	11,58	11,46	9,33	8,74	5,92	4,80	3,63	2,43	2,26	2,26	1,26	1,26	
2	6,30	5,90	5,83	4,73	4,44	3,06	2,53	2,00	1,47	1,40	1,40	1,04	1,04	
2,4	3,65	3,42	3,38	2,79	2,63	1,91	1,64	1,38	1,14	1,11	1,11	1,00	1,00	
3,2	1,73	1,65	1,63	1,44	1,38	1,16	1,09	1,04	1,01	1,00	1,00	1,00	1,00	

Fonte: [Muttlak e Al-Sabah \(2010\)](#).

* Os valores do CMS dos delineamentos AAS e URSS, foram calculados pelos autores.

Nota: Valores AAS foram obtidos analiticamente, com base nas propriedades da distribuição da média amostral.

sob descontrole, os valores do CMS produzidos via *URSS* foram menores que os CMS's obtidos via ACO. Nos cenários em que o *URSS* apresentou um CMS maior que a ACO, observa-se valores muito próximos. Por exemplo, para $\delta = 0,2$ e $\rho = 0,25$ o valor do CMS para ACO foi de 307,10 e do *URSS* 307,79. E novamente, nota-se aqui um valor de ρ baixo, indicando ser um cenário pouco adequado para aplicação de delineamentos baseados em conjuntos ordenados, o que pode ser constatado pela proximidade dos resultados em relação à AAS.

Sob ordenação imperfeita, considerando todos os tamanhos de amostra, contatou-se que ao utilizar a *URSS*, teve se um ganho em eficiência em 197 dos 200 cenários considerados em relação ao delineamento AAS (diferentes valores para ρ e δ) para o processo sob descontrole; o ganho em eficiência da ACO em relação à AAS foi em 198 dos 200 cenários (vide [Tabela 5 a 8](#)). O *URSS*, em relação à ACO se mostrou mais eficiente em 179 dos 200 cenários. Ainda, quando os valores do CMS eram maiores que o da ACO, observou-se que esses valores eram bem próximos, nenhuma diferença discrepante. No geral, sob ordenação imperfeita, ao comparar os resultados via *URSS* obtidos com os valores do CMS produzidos via ACO, *MRSS* e *ERSS*, aqui não mostrados (os dois últimos), mas apresentados em [Muttlak e Al-Sabah \(2010\)](#), o *URSS* foi mais eficiente em 167 de 200 cenários (diferentes valores para ρ e δ), detectando um possível processo fora de controle com menor número médio de amostras na maior dos cenários. Além disso, pode ser visto entre as [Tabela 5 a 8](#) que quanto maior o tamanho de amostra (k), menor o CMS para a ACO e *URSS* para dado valor de delta e ρ .

3.2 Análise do Viés do Estimador Proposto para a Variância da Média via *URSS*

O viés relativo do estimador proposto para a variância da média, equação (2.37), considerando diferentes quantidades de amostras ($B = 5, 10, 15, 20, 25, 30, 50$ e 100) de tamanho $k = 3$, e correlação entre a variável de interesse e a variável concomitante igual a $0,5$, foi inferior a $0,01$ (em 2 dos 8 valores de B , inferior a $0,001$) para todos os valores de B considerados, produzindo indicativo de baixo viés associado ao estimador. Para cenários mais favoráveis, resultantes de maiores valores para b, k e ρ , espera-se que o viés relativo seja ainda menor.

3.3 Simulação de Aplicação em Dados Reais

Antes de apresentar os gráficos de controle para os dados de resistência, algumas medidas descritivas são apresentadas na [Tabela 9](#). A média de quantidade de cimento utilizada para produção de concreto foi de 281,2kg, com desvio padrão igual a 104,5Kg, mí-

nimo igual a 102kg e máximo igual a 540kg. Ainda, para a variável resistência do concreto, teve-se resistência mínima e máxima iguais a 2,3MPa e 82,6MPa, respectivamente, com resistência média igual a 35,8Mpa e desvio padrão igual a 16,7Mpa. Observa-se que dada a elevada variação nas quantidades de cimento e resistência do concreto, eles não configuram exatamente dados de um processo de produção, no contexto de controle de qualidade. No entanto, para efeito de aplicação, considerou-se a variação apresentada nos níveis de resistência como a variação natural de um processo industrial, apenas para fins de ilustração.

Tabela 9 – Medidas resumo para as variáveis quantidade de cimento e resistência do concreto

Medidas descritivas	Quantidade de cimento (Kg)	Resistência do concreto (MPa)
Mínimo	102, 0	2, 3
Média	281, 2	35, 8
Mediana	272, 9	34, 5
Máximo	540, 0	82, 6
Desvio padrão	104, 5	16, 7
Coeficiente de Correlação		0,50

Fonte: Os autores

Ainda, para investigar a distribuição da resistência do concreto e a correlação da mesma com a quantidade de cimento utilizada, dois gráficos são apresentados na [Figura 8](#). O gráfico da esquerda representa um histograma para a variável resistência do concreto, que tem um comportamento similar à distribuição normal (configurando uma distribuição aproximadamente simétrica, mas com uma leve assimetria à direita). Ainda, a curva vermelha representa a distribuição normal para a média e desvio padrão de concreto calculados na [Tabela 9](#), e a curva azul representa a densidade não paramétrica estimada via método de Kernel ([VENABLES; RIPLEY, 2002](#)). O gráfico de dispersão, localizado à direita, mostra a relação entre as variáveis quantidade de cimento e resistência do concreto. A reta alaranjada é baseada nas estimativas de mínimos quadrados de uma regressão linear simples ([WILKINSON; ROGERS, 1973](#)), enquanto que a curva cinza representa a curva de um modelo de regressão não paramétrico, ajustado com base em polinômios locais ([CLEVELAND; GROSSE; SHYU, 1992](#)).

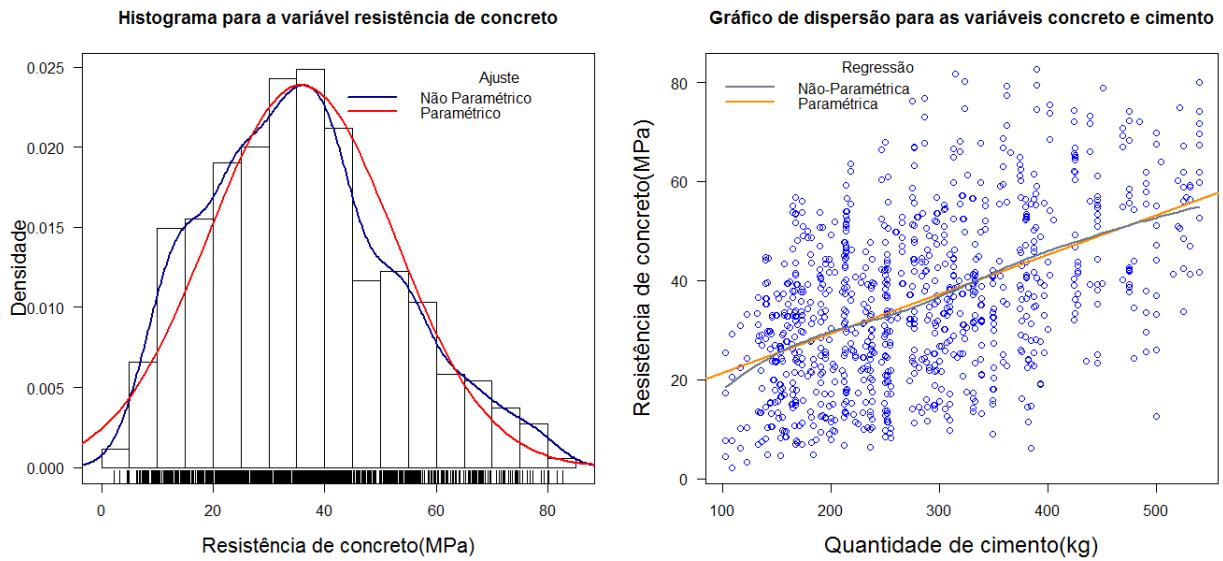


Figura 8 – Gráficos descritivos para a variável resistência do concreto
Fonte: Os autores

Na [Figura 8](#), é visto que a distribuição da variável de interesse se aproxima de uma normal, bem como a relação linear crescente com a variável concomitante (quantidade de cimento), viabilizando assim o uso do delineamento *URSS* para a estimativa da média de resistência de concreto, além da AAS e a ACO. Sendo assim, assume-se que a variância das resistências dos blocos de concreto é a variância natural do processo, e a aplicação de gráficos de controle para a média é apresentado na sequência:

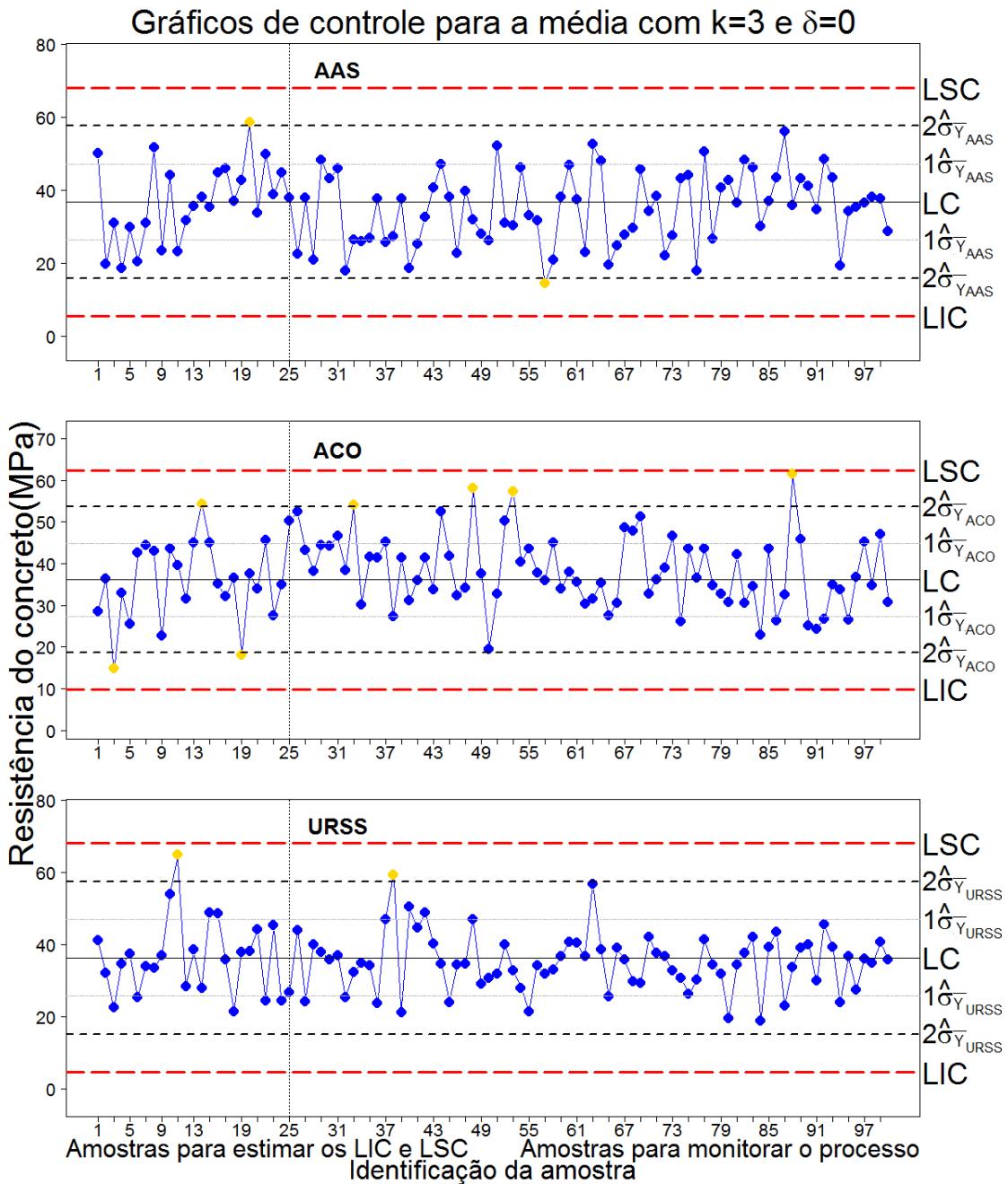


Figura 9 – Gráficos de controle para a variável resistência do concreto com $k = 3$ e processo sob controle ($\delta = 0$)

Fonte: Os autores

A Figura 9 apresenta gráficos de controle para a média para os delineamentos AAS, ACO e o URSS quando o processo opera sob controle. Observa-se um comportamento aleatório dos pontos ao redor da linha central, e todos os pontos entre os limites inferior e superior de controle, o que caracteriza e representa um processo operando sob controle; caso algum ponto extrapolasse esses limites, o mesmo seria considerado um falso alarme, pois sabe-se que o processo está sob controle. É visto ainda 2 pontos entre os limites 2 e 3σ para o gráfico gerado via AAS, outros 7 para a ACO e 2 para o URSS.

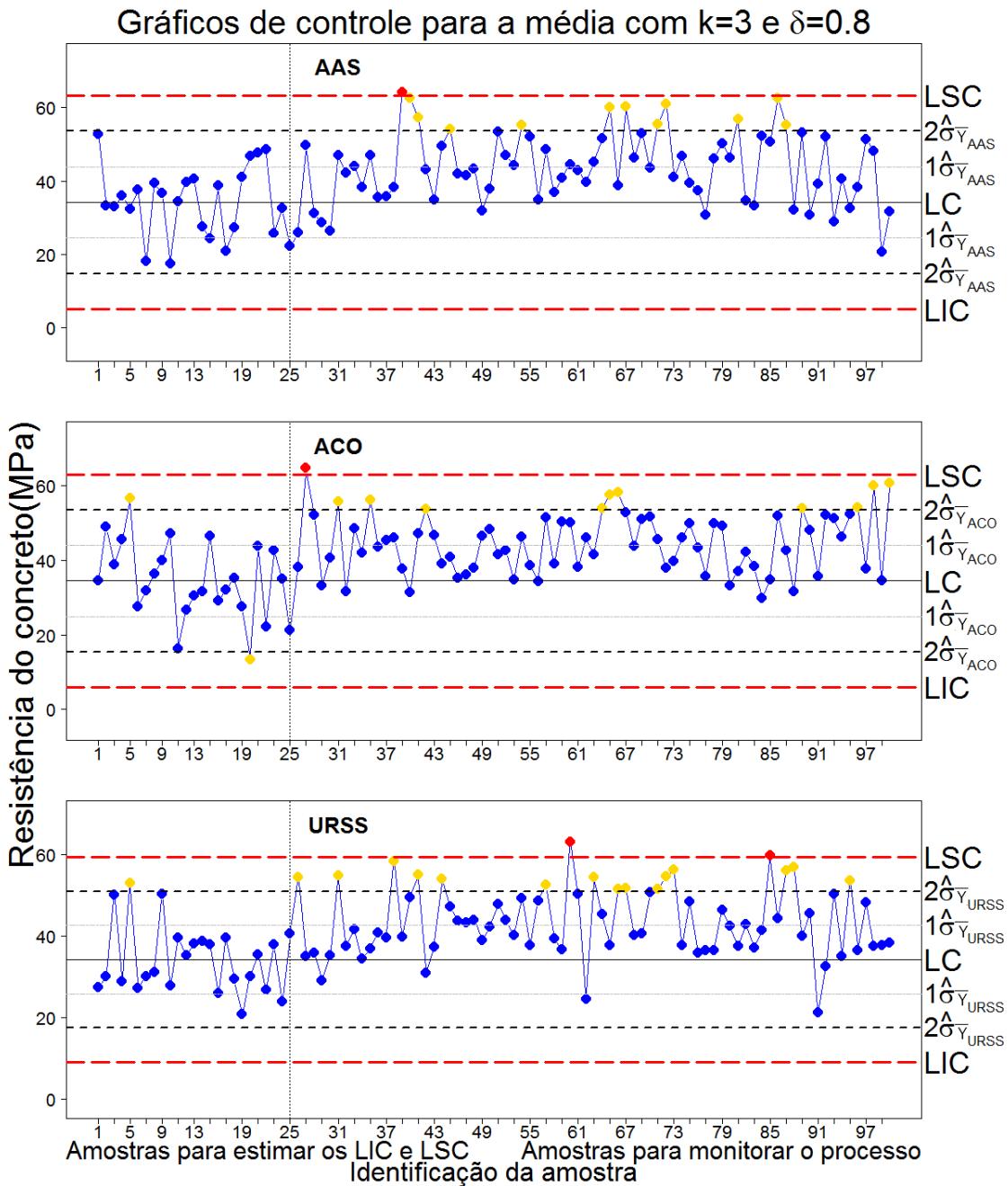


Figura 10 – Gráficos de controle para a variável resistência do concreto com $k = 3$ e $\delta = 1,2$

Fonte: Os autores

A Figura 10 representa o processo sob descontrole para $\delta = 0,8$ a partir do 26º ponto. Para todos os gráficos é notável que a partir do 26º ponto, ocorre uma alteração na média do processo, que passa de μ_0 para μ_1 , sendo, nesse caso, $\mu_1 = 35,8 + \frac{0,8*16,7}{\sqrt(3)} = 43,51$. Pode ser visto que a AAS e a ACO identificaram um ponto fora dos limites de controle, e para o URSS foram dois. Para a AAS, a identificação ocorreu no 12º ponto após alteração na média, enquanto para a ACO, no 2º e para o URSS, no 35º e no 60º respectivamente. Embora a ACO tenha sido o delineamento que produziu um ponto fora dos limites de

controle mais rapidamente, seguido por AAS e URSS, e que isso tenha uma importância na prática, vale ressaltar que as amostras (e consequentes pontos no gráfico) retiradas da linha de produção são independentes uma das outras, ou seja, o número de pontos extrapolando os limites de controle é mais importante do que a posição de suas ocorrências. Ainda, via URSS, teve-se o maior número de pontos identificados entre os limites 2 e 3σ , no total de 15, o que representa uma tendência do real descontrole do processo, enquanto que para a ACO e AAS foram 10 e 11 respectivamente.

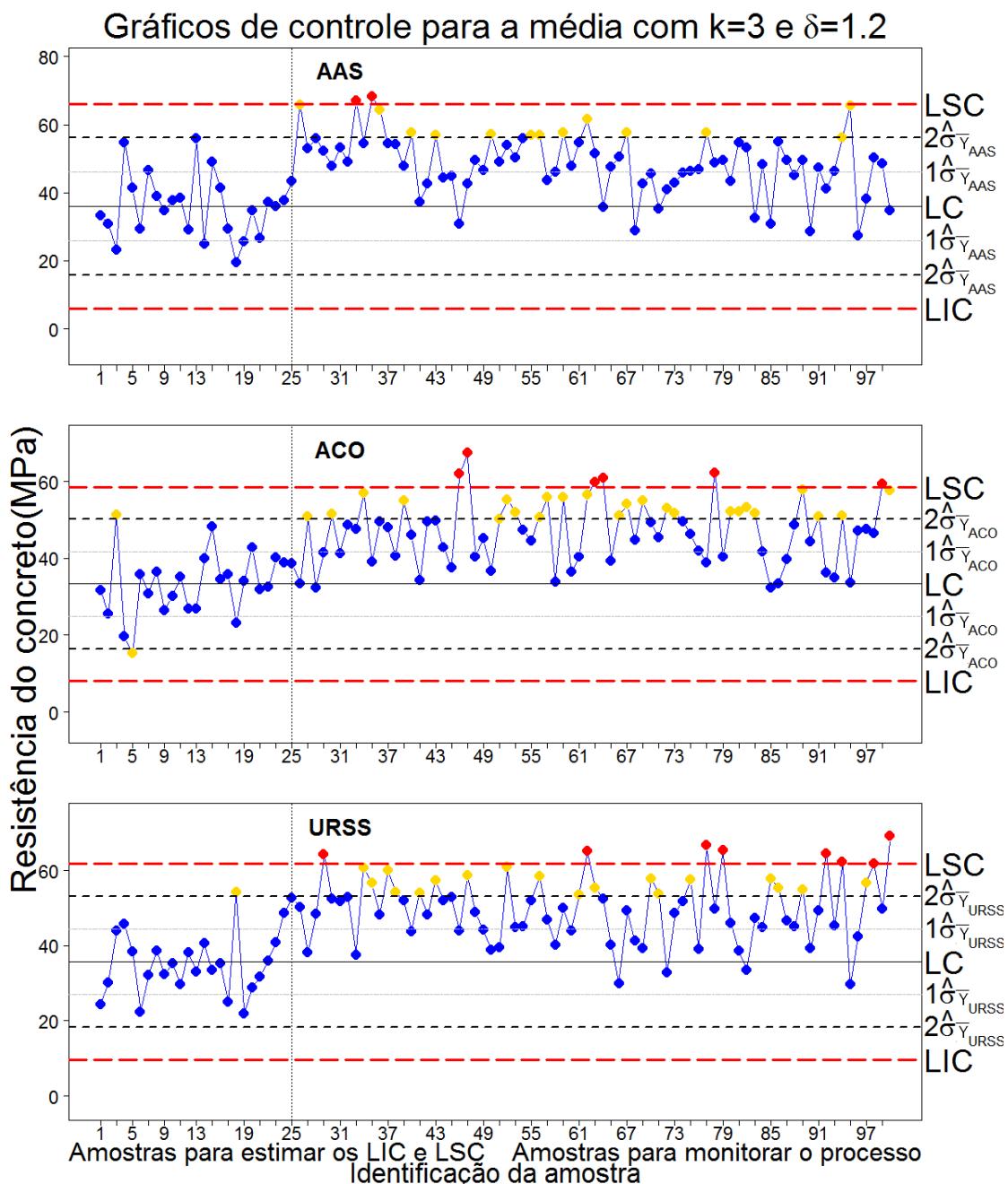


Figura 11 – Gráficos de controle para a variável resistência do concreto com $k = 3$ e $\delta = 1,2$

Fonte: Os autores

A Figura 11 representa novamente um processo em descontrole para tamanho de amostras igual a 3 e $\delta = 1, 2$, sendo o URSS, o delineamento que mais produziu pontos fora dos limites de controle, um total de 8, contra 6 da ACO e 2 da AAS. É visto também, um menor número de pontos entre os limites 2σ e 3σ da AAS (13), se comparado à ACO (24) e ao URSS(18).

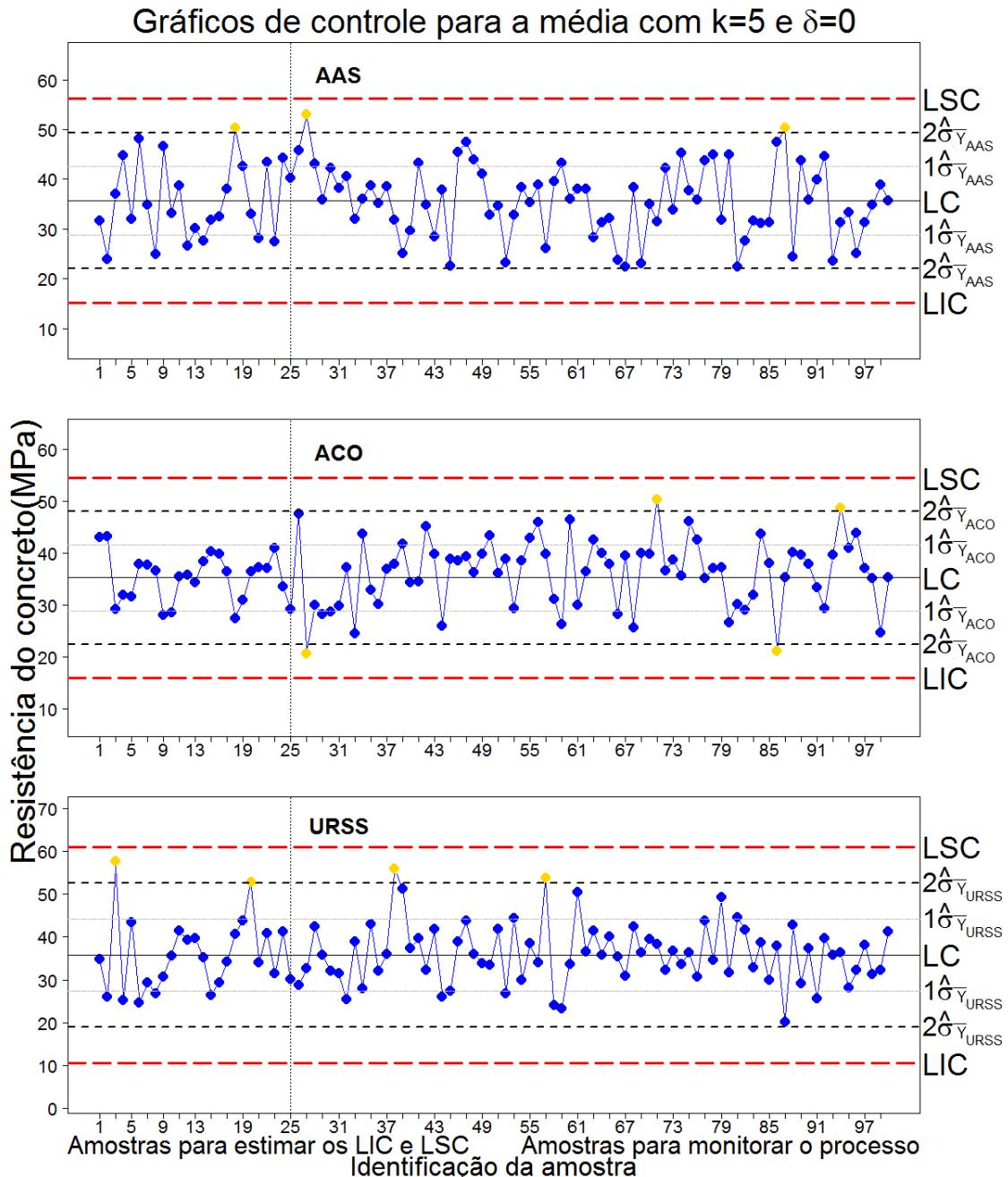


Figura 12 – Gráficos de controle para a variável resistência do concreto com $k = 5$ e processo sob controle ($\delta = 0$).

Fonte: Os autores

As Figuras 12, 13, e a 14 apresentam os gráficos de controle para a média, para um tamanho de amostra igual a 5. É visto que os limites de controle são mais estreitos

se comparados aos produzidos pelos gráficos de controle para $k = 3$. Ainda, para a Figura 12 é visto também um comportamento aleatório ao redor da linha central, mesmo comportamento observado para a Figura 9, e também que nenhum delineamento produziu pontos fora dos limites de controle. Esse resultado já era de se esperar, uma vez que o CMS para esse cenário foi próximo de 370 para os 3 delineamentos, conforme a Tabela 7.

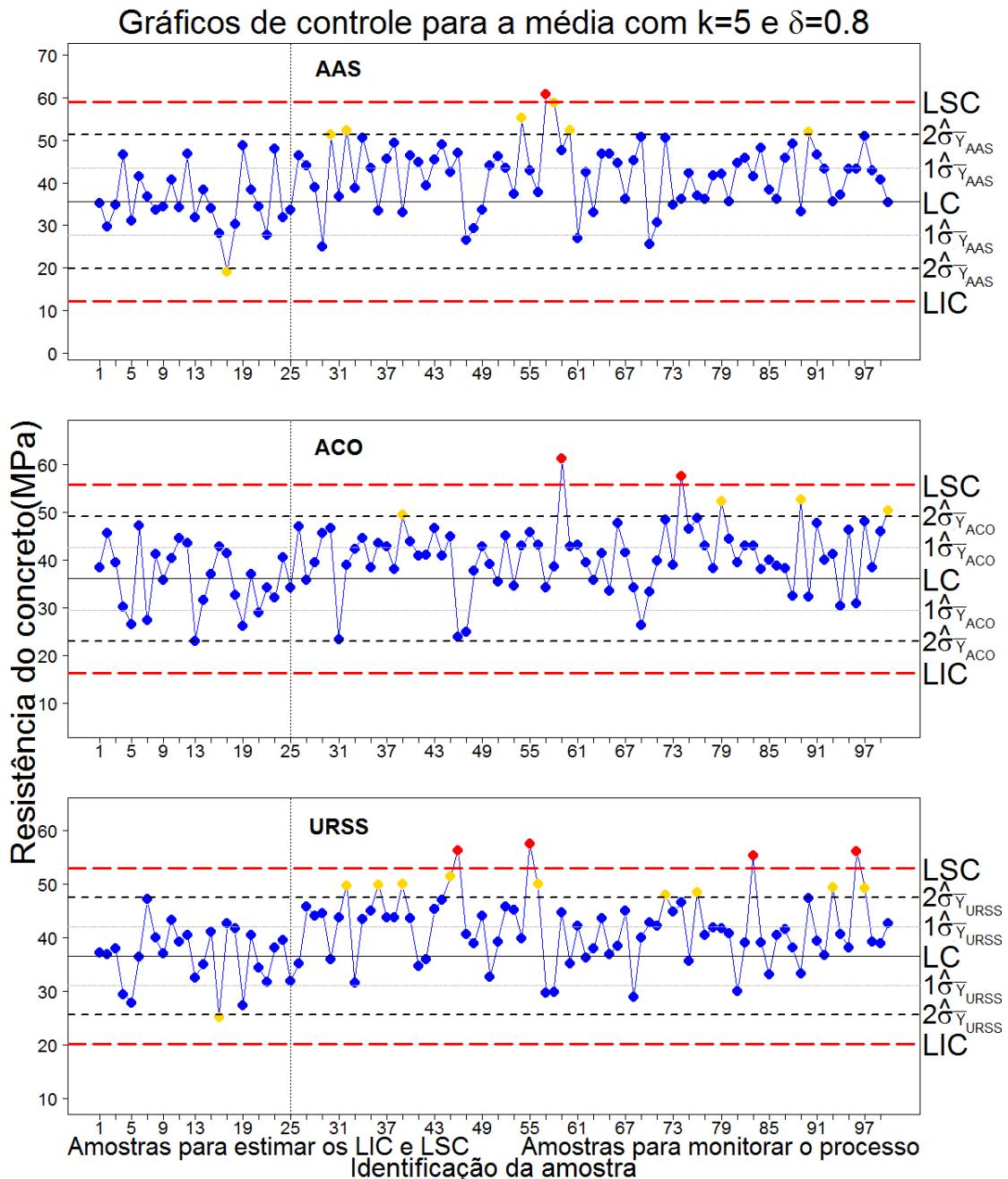


Figura 13 – Gráficos de controle para a variável resistência do concreto com $k = 5$ e $\delta = 0,8$

Fonte: Os autores

Na Figura 13 é visto que o delineamento URSS produziu um maior número de pontos (4) que extrapolando os limites de controle se comparado à ACO (2) e à AAS (1).

Para $k = 3$, a ACO detectou somente 1 ponto e o URSS 2, como consequência de limites de controle de maiores magnitudes se comparado ao caso em que $k = 5$. É nítido também que o URSS produziu um maior número de pontos entre os limites 2 e 3σ (9), contra 4 da ACO e 6 da AAS.

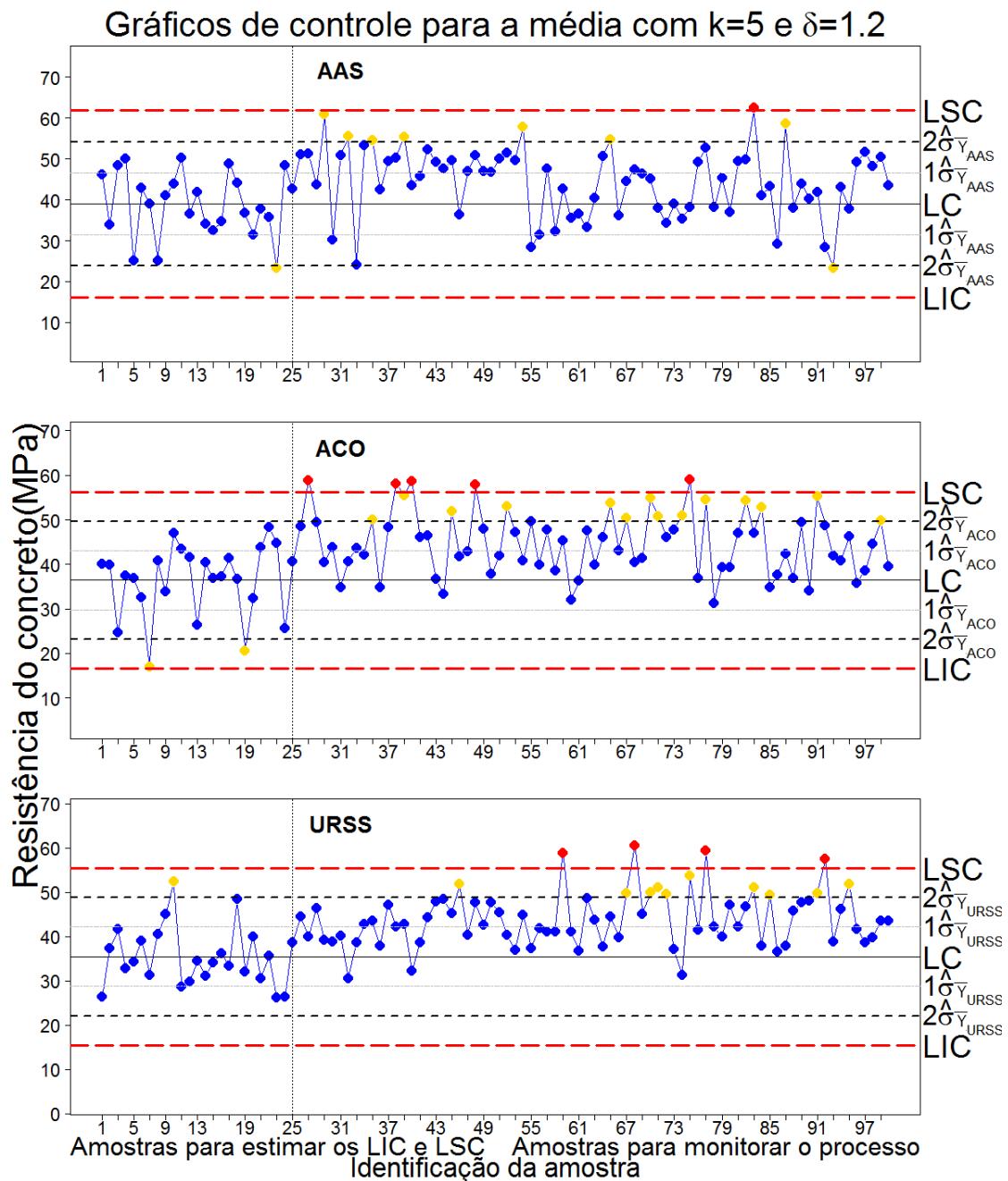


Figura 14 – Gráficos de controle para a variável resistência do concreto com $k = 5$ e $\delta = 1,2$

Fonte: Os autores

Na [Figura 14](#), foi considerada uma alteração na média igual a 1,2 sendo o delineamento que mais vezes produziu pontos fora dos limites de controle a ACO, 5 vezes ao todo, seguido pelo URSS com 4 e a AAS com apenas 1 amostra. É visto também que mesmo

após o incremento do δ , o delineamento da AAS continuou a produzir com frequência pontos abaixo da LC, com um total de 20, se comparado à ACO, com 8 pontos e somente 3 para o *URSS*.

4 Conclusões

Este trabalho teve como foco apresentar o delineamento *URSS*, suas principais propriedades, o estimador da média e a respectiva variância, além de compará-lo com os demais já consolidados na literatura, como a AAS e ACO, na construção de gráficos de controle para a média. O ganho em eficiência do *URSS*, em termos de um menor comprimento médio de sequência, para processos operando fora de controle, em relação aos delineamentos amostrais *ACO*, *ERSS*, *MRSS*, *EDRSS*, *DERSS*, *QDRSSe DQRSS* foi em 190 dos 210 (90,48%) cenários considerados sob ordenação perfeita, e em 167 dos 200 (83,5)% dos cenários sob ordenação imperfeita em relação aos delineamentos *ACO*, *ERSS* e *MRSS*, ambos considerando processos com algum grau de descontrole. O ganho em eficiência do *URSS* em relação à AAS foi em praticamente todos os cenários, perdendo apenas em alguns casos quando $\rho = 0,25$, cenário que não se recomenda o uso de delineamentos baseados em conjuntos ordenados. Ainda, para os delineamentos baseados em conjuntos ordenados, quanto maior o tamanho de amostra (k), menor o CMS, para dado valor de δ e ρ .

Ainda, na simulação da aplicação em dados reais, o *URSS* foi superior em 5 dos 6 cenários do processo operando sob descontrole. No único caso em que perdeu, foi por 1 amostra, confirmando a superioridade encontrada via simulação. Nota-se que os cenários em que foram considerados processos operando sob controle servem mais para averiguar se a taxa de falsos alarmes ocorre conforme o esperado (próximo a 1 em 370), não servindo no entanto para avaliar os delineamentos, em um primeiro momento. Ainda, foi proposto um estimador para a variância da média, para o qual se verificou ter vício desprezível.

Em termos de facilidade de aplicação (no caso que a ordenação remete em custos, e precisam ser reduzidos), dentre os delineamentos baseados em conjuntos ordenados, identifica-se o delineamento *ERSS*, como o de mais fácil aplicação, uma vez que requer apenas identificar as unidades amostrais que se julga ter o menor e maior valor. Em seguida, a *ACO*, pois sempre para dois conjuntos (independetemente do tamanho de k), é necessário somente observar as unidades amostrais com valores extremos, sem necessariamente ter que ranqueá-los. Após esses, viria o *MRSS*, pois para todos os conjuntos, seria necessário ordená-los a fim de se obter a unidade amostral que se julga a mediana. E em seguida, o *URSS*, pois é necessário ordenar todas as k^2 unidades uma única vez; no entanto, se o processo de ordenação for feito por alguma variável concomitante, por exemplo, a dificuldade de aplicação do *URSS* pode equivaler à dos demais delineamentos já citados. E por último, com maior complexidade de aplicação, os delineamentos *doubles*, pois requerem k^3 unidades amostrais e 2 estágios de ordenação; a comparação entre os diferentes delineamentos dessa classe não é discutida aqui.

Portanto, levando em consideração todo o material já apresentado, sugere-se o

uso do *URSS* como o delineamento a ser utilizado para o monitoramento do processo via gráfico de controle para a média, dado o seu ganho em eficiência em relação aos demais delineamentos baseados em conjuntos ordenados e a AAS. Importante ressaltar que o melhor delineamento é aquele que produziria uma combinação ótima de alguns aspectos: eficiência, tempo de execução, custo e aplicabilidade, dentre outros. A escolha do delineamento mais adequado para um determinado fim deve levar em conta todos esses aspectos, cabendo à pessoa responsável da área, escolher o de maior ganho.

Sugere-se também, para trabalhos futuros, avaliar o uso do *URSS* para variáveis com distribuições assimétricas no contexto de CEP. Outro estudo, seria o de identificar qual seria o tamanho de k necessário para AAS ter a mesma precisão que o delineamento *URSS* no mesmo contexto. Ainda, sugere-se utilizar as regras sensibilizantes para também comparar o *URSS* com os demais delineamentos apresentados; além de, elaborar gráficos de controle do tipo *CUSUM* e *EWMA* para a comparação do CMS obtido via *URSS* aos demais delineamentos.

Referências

- ABUJIYA, M.; MUTTLAK, H. Monitoring the process mean and variance using ranked set sampling and its modifications. *Journal of Statistical Theory and Applications*, v. 6, n. 4, p. 408–422, 2007.
- ABUJIYA, M. R.; LEE, M. H. The three statistical control charts using ranked set sampling. In: IEEE. *Modeling, Simulation and Applied Optimization (ICMSAO), 2013 5th International Conference on*. [S.I.], 2013. p. 1–6.
- ABUJIYA, M. R.; RIAZ, M.; LEE, M. H. Improving the performance of combined shewhart–cumulative sum control charts. *Quality and Reliability Engineering International*, Wiley Online Library, v. 29, n. 8, p. 1193–1206, 2013.
- AÏTCIN, P.-c. Concreto de alto desempenho. *São Paulo: Pini*, 2000.
- AL-HADHRAMI, S. A. Estimation of the population variance using ranked set sampling with auxiliary variable. *Int. J. Contemp. Math. Sciences*, v. 5, n. 52, p. 2567–2576, 2010.
- AL-OMARI, A. I.; AL-SALEH, M. F. Quartile double rankled set sampling for estimating the population mean. *Economic Quality Control*, v. 24, n. 2, p. 243–253, 2009.
- AL-OMARI, A. I.; BOUZA, C. N. Review of ranked set sampling: Modifications and applications. *Revista Investigación Operacional*, v. 3, p. 215–240, 2014.
- AL-OMARI, A. I.; HAQ, A. Improved quality control charts for monitoring the process mean, using double-ranked set sampling methods. *Journal of Applied Statistics*, Taylor & Francis, v. 39, n. 4, p. 745–763, 2012.
- BALAKRISHNAN, N.; RAO, C. R. *Order statistics: theory & methods*. [S.I.]: Elsevier Amsterdam, 1998.
- BARRETO, M. C. M. Planejamentos eficientes em pesquisa no meio ambiente usando amostragem em conjuntos ordenados. *Rev. Mat. Estat*, v. 19, p. 71–84, 2001.
- BOLFARINE, H.; BUSSAB, W. de O. *Elementos de amostragem*. [S.I.]: Edgard Blücher São Paulo, 2005. v. 1.
- BOWERMAN, B. L.; O'CONNELL, R. T. Forecasting and time series: An applied approach. 3rd. Duxbury press, 1993.
- CASELLA, G.; BERGER, R. L. *Statistical inference*. [S.I.]: Duxbury Press, 2001. v. 2.
- CHANDRA, M. J. *Statistical quality control*. [S.I.]: CRC Press, 2001.
- CHEN, Z.; BAI, Z.; SINHA, B. *Ranked set sampling: theory and applications*. [S.I.]: Springer Science & Business Media, 2004. v. 176.
- CLEVELAND, W. S.; GROSSE, E.; SHYU, W. M. Local regression models. *Statistical models in S*, Pacific Grove, p. 309–376, 1992.
- DELL, T.; CLUTTER, J. Ranked set sampling theory with order statistics background. *Biometrics*, JSTOR, p. 545–555, 1972.

- GARVIN, D. A. Competing on the eight dimensions of quality. *Harvard Business Review*, v. 65, n. 6, 1987.
- HATEFI, A.; JOZANI, M. J. Proportion estimation based on a partially rank ordered set sample with multiple concomitants in a breast cancer study. *arXiv preprint arXiv:1411.2573*, 2014.
- JABEEN, S. Ranked set sampling and its application in quality control. *Revista de Matemática e Estatística, São Paulo*, 2012.
- JEMAIN, A. A.; AL-OMARI, A. I. Double quartile ranked set samples. *Pakistan Journal of Statistics-all series-*, v. 22, n. 3, p. 217, 2006.
- JURAN, J. M. Juran na liderança pela qualidade: Um guia para executivos. 3^a edição. *Tradução de Mário Csillag. São Paulo: Editora Pioneira*, 1990.
- KLOSS, C. L. *Materiais para construção civil*. [S.l.]: CEFET-PR, 1991.
- MACEACHERN, S. N. et al. A new ranked set sample estimator of variance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 64, n. 2, p. 177–188, 2002.
- MCINTYRE, G. A method for unbiased selective sampling, using ranked sets. *Crop and Pasture Science*, CSIRO, v. 3, n. 4, p. 385–390, 1952.
- MEYER, P. L. Probabilidade: aplicações à estatística. In: *Probabilidade: aplicações à estatística*. [S.l.]: Livro Técnico, 1970.
- MONTGOMERY, D. C. *Introduction to Statistical Quality Control*. [S.l.]: John Wiley & Sons, 2011. v. 6.
- MUTTLAK, H. Median ranked set sampling. *Journal of Applied Statistical Science*, v. 6, n. 4, p. 245–255, 1997.
- MUTTLAK, H. Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators. *Environmetrics*, Wiley Online Library, v. 9, n. 3, p. 255–267, 1998.
- MUTTLAK, H.; AL-SABAHI, W. Statistical quality control based on ranked set sampling. *Journal of Applied Statistics*, Taylor & Francis, v. 30, n. 9, p. 1055–1078, 2010.
- NELSON, L. S. Column: Technical aids: The shewhart control chart–tests for special causes. *Journal of quality technology*, ASQ, v. 16, n. 4, 1984.
- NOGUEIRA, M. L. Avaliação do controle de qualidade de misturas asfálticas e análise do reflexo do nível de qualidade no desempenho dos revestimentos asfálticos. 2011.
- NUSSBAUM, B.; SINHA, B. Cost effective gasoline sampling using ranked set sampling. In: *Proceedings of the Section on Statistics and the Environment*. [S.l.: s.n.], 1997. p. 83–87.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015. Disponível em: <<https://www.R-project.org/>>.

- SAMAWI, H. M. On double extreme rank set sample with application to regression estimator. *Metron—International Journal of Statistics*, v. 60, p. 50–63, 2002.
- SAMAWI, H. M.; AHMED, M. S.; ABU-DAYYEH, W. Estimating the population mean using extreme ranked set sampling. *Biometrical Journal*, Wiley Online Library, v. 38, n. 5, p. 577–586, 1996.
- SAMOHYL, R. W. *Controle estatístico de qualidade*. [S.l.]: Elsevier, 2009.
- SHEWHART, W. A. Some applications of statistical methods to the analysis of physical and engineering data. *Bell System Technical Journal*, Wiley Online Library, v. 3, n. 1, p. 43–87, 1924.
- SILVA, L. J.; ROCHA, R. C. *Delineamento Amostral por Conjuntos Ordenados Aplicado ao Controle Estatístico de Qualidade*. 49 p. Monografia (Trabalho de Conclusão de Curso) — Setor de Ciências Exatas, Curitiba, Brasil, 2013.
- STOKES, S. L. Estimation of variance using judgment ordered ranked set samples. *Biometrics*, JSTOR, p. 35–42, 1980.
- TACONELI, C. A.; BARRETO, M. C. M. Avaliação de uma proposta de intervalos de confiança bootstrap em amostragem por conjuntos ordenados perfeitamente. *Revista de Matemática e Estatística, São Paulo*, v. 23, n. 3, p. 33–53, 2005.
- TAKAHASI, K.; WAKIMOTO, K. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, Springer, v. 20, n. 1, p. 1–31, 1968.
- TRIETSCH, D. *Statistical quality control: A loss minimization approach*. [S.l.]: World Scientific, 1999. v. 10.
- VENABLES, W. N.; RIPLEY, B. D. *Modern Applied Statistics with S*. Fourth. New York: Springer, 2002. ISBN 0-387-95457-0. Disponível em: <<http://www.stats.ox.ac.uk/pub/MASS4>>.
- WILKINSON, G.; ROGERS, C. Symbolic description of factorial models for analysis of variance. *Applied Statistics*, JSTOR, p. 392–399, 1973.
- YEH, I.-C. Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete research*, Elsevier, v. 28, n. 12, p. 1797–1808, 1998.
- YOUTUBE. *Compressao Forney*. 2009. Disponível em: <<https://www.youtube.com/watch?v=euzx9cxKwRA>>. Acesso em: 22/11/2015.
- ZAMANZADE, E.; AL-OMARI, A. I. Unified ranked set sampling. *arXiv preprint arXiv:1411.1351v2*, 2015.

Apêndices

APÊNDICE A – Seleção de uma amostra URSS

```
#####
#####
```

Instruções gerais para uso dos códigos:

```
#####
#####
```

É necessário observar os comandos estão na mesma linha para algumas comandos, fez-se necessário a quebra de linha, portanto, é necessário, deixar tudo na mesma linha, caso contrário, os comandos podem não funcionar corretamente

```
select <- function(aas){
  # Input:
  # Dados provenientes de uma AAS (tamanho k2)
  # Descrição
  # A função select retorna a posição dos elementos a serem
  # selecionados de uma AAS
  len <- length(aas)
  k <- sqrt(len)
  aas <- sort(aas)
  elementos <- rep(0,ncol=k)
  posicao <- rep(0,k)
  i <- 1
  if (len%%2 == 0){
    for(i in 1:k){
      if (i%%2 == 0){l <- k/2}
      else{l <- (k+2)/2 }
      posicao[i] <- (l+(i-1)*k)
      elementos[i] <- aas[posicao[i]]
    }
  }
  else{
    for(i in 1:k){
      posicao[i] <- ((k+1)/2)+(i-1)*k
      elementos[i] <- aas[posicao[i]]
    }
  }
}
```

```
    }
    return(posicao)
}
#Example
a <- rnorm(16)
select(a)
```

APÊNDICE B – Estatísticas de ordem

```

# Função densidade de probabilidade proveniente de uma amostra ordenada#####
FdpOrdem <- function(k2, i, x){
  # Input:
  # k2 é o tamanho da sua amostra k ao quadrado
  # i é a posição ordenada da sua amostra
  # x é a sua variável a ser integrada (usando a função Integrate)
  # Descrição:
  # Função densidade de probabilidade de uma amostra ordenada
  (factorial(k2)*(pnorm(x)^(i-1))*((1-pnorm(x))^(k2-i))*dnorm(x))#Continua
  /(factorial(i-1)*factorial(k2-i))
}

#fdpx (fdp + a multiplicação do x por FdpOrdem)
FdpOrdemX <- function(k2, i, x){
  # Input:
  # k2 é o tamanho da sua amostra k ao quadrado
  # i é a posição ordenada da sua amostra
  # x é a sua variável a ser integrada (usando a função Integrate)
  # Descrição:
  # Fdp de uma amostra ordenada multiplicada por "x", pela própria
  # variável, para ser usada no cálculo da Variância
  ((x*factorial(k2))*(pnorm(x)^(i-1))*((1-pnorm(x))^(k2-i))*dnorm(x))
  /(factorial(i-1)*factorial(k2-i))

}

#fdpx^2 (fdp + a multiplicação do x2 por FdpOrdem)
FdpOrdemX2 <- function(x,i,k2){
  # Input:
  # k2 é o tamanho da sua amostra k ao quadrado
  # i é a posição ordenada da sua amostra
  # x é a sua variável a ser integrada (usando a função Integrate)
  # Descrição:
  # Fdp de uma amostra ordenada multiplicada por "x2", pela própria
  # variável, para ser usada no cálculo da Variância

```

```

(((x^2)*factorial(k2))*(pnorm(x)^(i-1))*((1-pnorm(x))^(k2-i))*dnorm(x))
/(factorial(i-1)*factorial(k2-i))
}

# Calculando a Variancia
k2 <- 25
i <- 8
ex2 <- integrate(Fdp0rdemX2,lower=-Inf,upper=+Inf,k2=k2,i=i)$value
ex <- (integrate(Fdp0rdemX,lower=-Inf,upper=+Inf,k2=k2,i=i)$value)^2
var_i_8 <- ex2 - ex

k2 <- 25
i <- 3
ex2 <- integrate(Fdp0rdemX2,lower=-Inf,upper=+Inf,k2=k2,i=i)$value
ex <- (integrate(Fdp0rdemX,lower=-Inf,upper=+Inf,k2=k2,i=i)$value)^2
var_i_3 <- ex2 - ex

# Covariância #####
cov.ord <- function(k2, i, j, x, y){
  # Input:
  # k2 é o tamanho da sua amostra k ao quadrado
  # i é a posição do primeiro elemento amostral
  # j é a posição do segundo elemento amostral
  # x é a variável a ser integrada correspondente ao elemento i
  # y é a variável a ser integrada correspondente ao elemento j
  # Descrição:
  # Calcula a covariância entre dois elementos de uma amostra ordenada
  exy <- integrate(function(y) {
    sapply(y, function(y) {
      integrate(function(x) (x*y*factorial(k2)*(pnorm(x)^(i-1))*((pnorm(y)-pnorm(x))^(j-i-1))*((1-pnorm(y))^(k2-j))*dnorm(x)*dnorm(y))
      /(factorial(i-1)*factorial(j-i-1)*factorial(k2-j))
      , lower = -Inf, upper = y)$value
    })
  }, lower = -Inf, upper = +Inf)$value
  ex <- (integrate(Fdp0rdemX,lower=-Inf,upper=+Inf,k2=k2,i=i)$value)
  ey <- (integrate(Fdp0rdemX,lower=-Inf,upper=+Inf,k2=k2,i=j)$value)
  return(exy-(ex*ey))
}

```

```
# Example
correlaç&ao = cov.ord(k2=25,i=3,j=8,x,y)/(sqrt(var_i_8)*sqrt(var_i_3))
```

APÊNDICE C – Estimação da $Var(\bar{Y}_{URSS})$ para ordenação imperfeita

```

# Calculando a variância da média URSS para o caso imperfeito #####
aas <- c(9, 16, 25, 36)                      # Tamanho da amostra
pho <- c(0,0.25,0.50,0.75,0.9,1)    # Coeficiente de correlação
var <- 1                                         # Variância de x e y
l <- 1000000                                     # 1 milhão de amostras

require(MASS)
var_mean_imp <- function(aas, pho, l, var){
  # Input:
  # aas: tamanho da sua amostra k2
  # pho: coeficiente de correlação entre a variável de interesse
  # e a auxiliar
  # l : Quantidade de amostras
  # var: Variância de x e y
  # Descrição:
  # Calcula a variância da média URSS via simulação para os casos
  # de ordenação imperfeita

  la <- length(aas)
  lp <- length(pho)
  var_mean_imp_values <- matrix(0,nrow=la,ncol=lp,
                                 dimnames=list(sqrt(aas),pho))
  seed <- 0

  for (i in 1:la){                               #Para cada tamanho de amostra
    k2 <- aas[i]
    ktot <- k2*l
    for (p in 1:lp){                            #Para cada correlação
      seed <- seed + 1
      set.seed(seed)
      d1 <- mvrnorm(n = ktot,
                     mu = c(0,0),
                     Sigma = matrix(c(var,pho[p],pho[p],var)

```

```

            ,ncol=2,byrow=TRUE))
dx <- matrix(d1[,1], ncol=k2)
dy <- matrix(d1[,2], ncol=k2)
rm(d1)
# Ordem das observações de x
pos <- t(apply(dx,1,order))
rm(dx)
# Matriz vazia para receber as amostras ordenadas
dys <- matrix(rep(0,ktot),ncol=k2)
# Ordena uma matriz y por x
for (linha in 1:nrow(dy)){
  dys[linha,] <- dy[linha,pos[linha,]]
}
rm(dy);rm(pos)
# Matriz ordenada com as unidades amostrais necessárias
dys_amostras <- dys[,select(1:k2)]
rm(dys)
mcov <- cov(dys_amostras)
rm(dys_amostras)
vari <- diag(mcov)
mcov[lower.tri(mcov, diag=TRUE)] <- 0
var_mean_imp_values[i,p] <- (1/k2)*sum(vari)+(2/k2)
*sum(colSums(mcov))
# Mantenho o print, pois o código demora.
# E dessa forma, consegue ver se a função está trabalhando bem
print(var_mean_imp_values)
}
}
return(var_mean_imp_values)
}
require(MASS)
# Aumentar o tamanho da memória para rodar o código
memory.limit(size=10000) # (4GB de RAM não é suficiente)

aas <- c(9,16,25,36)          # Tamanho da amostra
pho <- c(0,0.25,0.50,0.75,0.90,1) # Coeficiente de correlação
var <- 1                      # Variância de x e y
l <- 1000000                  # 1 milhão de amostras
c <- var_mean_imp(aas,pho,l,var)

```

APÊNDICE D – Estimação de $Var(\bar{Y}_{URSS})$

para ordenação perfeita

```

# Função da Variância da média do URSS - ordenação perfeita #####
# É necessário entrar com as posições dos elementos dos quais
# você irá amostrar
var_urss2 <- function(urss){
  # Input:
  # As posições dos elementos de sua amostra
  # (Dica: usa a função select())
  # Descrição
  # Com base nas posições dos elementos , essa função retorna
  # a variância do estimador da média para o URSS
  k <- length(urss)
  k2 <- k^2
  cov <- matrix(0,nrow=k, ncol=k)
  for(i in 1:k){
    for(j in 1:k){
      if (j > i){
        # Covariância
        cov[i,j] <- cov.ord(k=k2, i=urss[i], j=urss[j], x,y)
      } else if(i == j){
        cov[i,j] <- (integrate(FdpOrdemX2,lower=-Inf
                               ,upper=+Inf,k=k2,
                               i=urss[i])$value)-(integrate(FdpOrdemX ,lower=-Inf
                               ,upper=+Inf,k=k2
                               ,i=urss[i])$value)^2 #Variância
      }
      else {
        cov[i,j] <- 0
      }
    }
  }
  var <- diag(cov)
  cov[lower.tri(cov, diag=TRUE)] <- 0
  return((1/k2)*sum(var)+(2/k2)*sum(colSums(cov)))
}

```

```
}
```

```
# Exemplo
a <- select(rnorm(9))
var_urss2(a)
```

APÊNDICE E – Obtendo CMS para ordenação imperfeita

```

# Comprimento médio de sequência - imperfeito #####
cms_imp <- function(aas, pho, var, l, var_da_media_matrix, delta){
  # Input:
  # aas: tamanho da sua amostra k2
  # pho: coeficiente de correlação entre a variável de interesse
  # e a auxiliar
  # l : Quantidade de amostras
  # var: Variância de x e y
  # var_da_media_matrix: Matriz da var_mean_imp() que guarda todos
  # os valores da variância da média (caso imperfeito)
  # Descrição:
  # Calcula o comprimento médio de sequência(CMS) no caso
  # de ordenação imperfeita utilizando limites 3-Sigma

  la <- length(aas)
  lp <- length(pho)
  ld <- length(delta)
  cms_imp <- list('n= 3' = 1) # Vai guardar o CMS
  for (i in 1:la){           # Para cada tamanho de amostra k2
    k2 <- aas[i]
    k <- select(1:k2)
    lk <- length(k)
    ktot <- k2*l
    cms_matrix <- matrix(0, nrow = ld, ncol = lp,
                          dimnames=list(delta,pho))
    # Para cada nível de correlação
    for (p in 1:lp){
      var_mean_u_imp <- var_da_media_matrix[i, p]
      # Limites de controle que depende apenas de k e pho
      lsc <- 3*sqrt(var_mean_u_imp)
      lic <- -3*sqrt(var_mean_u_imp)
      # Para cada nível de alteração da média
      for (d in 1:ld){
        cms_matrix[d, p] <- lsc
      }
    }
  }
}

```

```

set.seed(9087832*p*d)
d1 <- mvrnorm(n = ktot,
               mu = c(0,delta[d]/sqrt(lk)),
               Sigma = matrix(c(var,pho[p],pho[p],var),
                             ,ncol=2,byrow=TRUE))
dx <- matrix(d1[,1], ncol=k2)
dy <- matrix(d1[,2], ncol=k2)
rm(d1)
# Posição dos elementos da matriz x (var. auxiliar)
pos <- t(apply(dx,1,order))
rm(dx)
# Matriz vazia para receber as amostras ordenadas
dys <- matrix(rep(0,ktot),ncol=k2)
# Ordena a matriz y por x
for (linha in 1:nrow(dy)){
  dys[linha,] <- dy[linha,pos[linha,]]
}
rm(dy);rm(pos)
# Matriz ordenada com as observações necessárias
dys_amostras <- dys[,k]
rm(dys)
dys_amostras_mean <- rowMeans(dys_amostras)
rm(dys_amostras)
# Checa se a média de cada amostra urss está
# dentro dos limites de controles calculados
# quando o processo está sob controle
mrow <- as.numeric(dys_amostras_mean<lic | dys_amostras_mean>lsc)
s <- sum(mrow)/length(mrow)
cms_matrix[d,p] <- 1/s
gc()
}
}
cms_imp[[i]] <- cms_matrix
print(cms_imp)
}
names(cms_imp) <- paste('n=' ,sqrt(aas),sep=" ")
return(cms_imp)
}

```

```
aas <- c(9,16,25,36)                      # Tamanho da amostra
pho <- c(0,0.25,0.50,0.75,0.90,1) # Coeficiente de correlação
var <- 1                                # Variância de x e y
l <- 1000000                            # 1 milhão de amostras
delta <- c(0, 0.1, 0.2, 0.3, 0.4, 0.8, 1.2, 1.6, 2, 2.4, 3.2)
cms_imp(aas, pho, var, l, c, delta)
```

APÊNDICE F – Obtendo CMS para ordenação perfeita

```

# Comprimento médio da sequência - ordenação perfeita#####
aas <- c(9, 16, 25, 36)
delta <- c(0, 0.1, 0.2, 0.3, 0.4, 0.8, 1.2, 1.6, 2, 2.4, 3.2)

cms <- function(aas, delta){
  # Input:
  # aas: corresponde aos tamanhos da sua amostra aleatória simples (k2)
  # DELTA: corresponde aos níveis de alteração da média do processo
  # Descrição:
  # Calcula o comprimento médio de sequência(CMS) no caso
  # de ordenação perfeita utilizando limites 3-Sigma

  la <- length(aas)
  ls <- length(delta)
  # Matriz que receberá o CMS dos diferentes níveis de alteração
  # da média do processo (na linha) e o tamanho amostral (coluna)

  cms_matrix <- matrix(0, nrow = ls, ncol = la,
                        dimnames=list(c(delta),c(sqrt(aas)))))

  # Para cada tamanho de amostra
  for (a in 1:length(aas)){
    k          <- select(rnorm(aas[a]))
    lk         <- length(k)
    lk2        <- lk^2
    # calcula a variância da média que depende somente de k, i, j, x e y
    var_mean_u <- var_urss2(k)

    # Limites de controle quando o processo está sob controle
    lsc <- 3*sqrt(var_mean_u)
    lic <- -3*sqrt(var_mean_u)

    # Para cada nível de alteração na média do processo
  }
}

```

```

for (i in 1:ls){
  set.seed(i*809)
  # Gera os dados, com o correspondente delta
  dados <- rnorm((lk2*1000000), mean= delta[i]/sqrt(lk), sd= 1)
  # Os organiza em uma matriz
  m <- matrix(dados, ncol=lk2)
  # Ordena pela linha
  mrow <- t(apply(m,1,sort.int, method="quick"))
  # Seleciona apenas as colunas necessárias
  mrow <- mrow[,k]
  # Calcula a média de cada amostra urss
  mrow <- cbind(mrow, rowMeans(mrow))
  ncolu <- ncol(mrow)
  # Checa se a média de cada amostra urss está dentro dos limites
  # de controles calculados quando o processo está sob controle
  mrow <- cbind(mrow, mrow[,ncolu]<lic | mrow[,ncolu]>lsc)
  p <- sum(mrow[,ncolu+1])/dim(mrow)[1]
  cms_matrix[i,a] <- 1/p
  print(cms_matrix)
}
}

return(cms_matrix)
}
cms_perf <- cms(aas, delta)

```

APÊNDICE G – Estudo sobre o viés do estimador proposto para $Var(\bar{Y}_{URSS})$

```

sim_var_mean_imp <- function(aas, pho, n, var, reps){

  # Input:
  # aas : tamanho dos elementos da sua amostra k2
  # pho : coeficiente de correlação entre a variável
  # de interesse e a auxiliar
  # n   : Quantidade de ciclos
  # var : Variância de x e y
  # reps: quantidade de amostras

  # Descrição:
  # Estima a variância da média

  la <- length(aas)
  lp <- length(pho)
  ln <- length(n)
  var_mean_imp_values <- matrix(0,nrow=la,ncol=ln
                                  ,dimnames=list(sqrt(aas),n))
  semente <- 0
  #Para cada tamanho de amostra
  for (i in 1:la){
    k2 <- aas[i]
    for (j in 1:ln){           #Para cada ciclo
      var_mean_imp <- rep(0, reps)
      for (b in 1:reps){       #Número de repetições
        semente <- semente + 1
        set.seed(semente)
        ktot <- k2*n[j]
        d1 <- mvrnorm(n = ktot,
                       mu = c(0,0),
                       Sigma = matrix(c(var,pho,pho,var)
                                      ,ncol=2,byrow=TRUE))
        dx <- matrix(d1[,1], ncol=k2)
        dy <- matrix(d1[,2], ncol=k2)
        rm(d1)
      }
    }
  }
}

```

```

# Ordem das observações de x
pos <- t(apply(dx,1,order))
rm(dx)
dys <- matrix(rep(0,ktot),ncol=k2)
# Ordena uma matriz y por x
for (linha in 1:nrow(dy)){
  dys[linha,] <- dy[linha,pos[linha,]]
}
rm(dy);rm(pos)
# Matriz ordenada com as observações necessárias
dys_amostras <- dys[,select(1:k2)]
rm(dys)
mcov <- cov(dys_amostras)
rm(dys_amostras)
vari <- diag(mcov)
mcov[lower.tri(mcov, diag=TRUE)] <- 0
var_mean_imp[b] <- (1/k2)*sum(vari)+(2/k2)*sum(colSums(mcov))
}
var_mean_monte_carlo <- mean(var_mean_imp)
vies <- (var_mean_monte_carlo - 0.280864483557273)
vies_relativo <- (vies/0.280864483557273)
var_mean_imp_values[i,j] <- vies_relativo
}
}
return(var_mean_imp_values)
}
aas <- c(9)                                # Número de elementos por amostra
pho <- c(0.50)                             # Coeficiente de correlação
var <- 1                                    # Variância de x e y
reps <- 50000                                # Quantidade de repetições
n <- c(5,10,15,20,25,30,50,100)    # Quantidade de amostras
sim_var_mean <- sim_var_mean_imp(aas, pho, n, var, reps)

```

APÊNDICE H – Elaboração das cartas

controle em um conjunto de dados

```
# Processos de seleção de amostras do URSS e ACO ####

urss_process <- function(data, l, k2){
  # Input
  # data é o seu conjunto de dados (entrar somente com 2 colunas,
  # sendo a 2a coluna aquela que contém a sua variável de interesse
  # l é a quantidade de amostras
  # k2 é o tamanho da amostra
  # Output
  # l amostras de tamanho k2 via URSS

  dx <- matrix(data[,1], ncol=k2)
  dy <- matrix(data[,2], ncol=k2)

  # Posição dos elementos da matriz x (var. auxiliar)
  pos <- t(apply(dx,1,order))
  # Matriz vazia para receber as amostras ordenadas
  dys <- matrix(rep(0,l*k2),ncol=k2)
  # Ordena a matriz y por x
  for (linha in 1:nrow(dy)){
    dys[linha,] <- dy[linha,pos[linha,]]
  }
  # Matriz ordenada com as observações necessárias
  dys_amostras <- dys[,select(1:k2)]
  return(dys_amostras)
}

aco_process <- function(base, m, n){

  # Input
  # data é o seu conjunto de dados (entrar somente com 2 colunas,
  # sendo a 2a coluna aquela que contém a sua variável de interesse
```

```

# m é a quantidade de amostras
# n é o tamanho da amostra
# Output
# m amostras de tamanho n via ACO

# -----
# separando os dados em duas matrizes
# -----
x <- matrix(base[,1], n, m*n) # primeiro vetor
y <- matrix(base[,2], n, m*n) # segundo vetor

# indices da ordem da amostra x
i <- apply(x,2,order, decreasing=F)

# criando indices para ordenar os elementos de y com base em x
l <- i[1:nrow(base)]
c <- rep(1:(n*m),each=n)

y.order <- matrix(y[cbind(l,c)], n, n*m)

# -----
# selecao das diagonais
# -----
ll <- rep(sequence(nrow(x)), m) # linha selecionada
cc <- rep(sequence(ncol(x))) # coluna selecionada

# selecao do elementos na diagonal
amostra.sel <- matrix(y.order[cbind(ll,cc)], ncol = n, byrow = TRUE)
return(amostra.sel)
}

#Função para plotar o gráfico com legendas e tudo mais #####
plot_grafico <- function(k, delta, method, lc, lsc, lic, lsc1, lic1,
                           , lsc2, lic2, sl, dados, cor, cex){

# Input (Na ordem que aparece na entra da função) :
tamanho da amostra, delta, delineamento utilizado, lc, lsc, lic,
lsc(A=1), lic(A=1), lsc(A=2), lic(A=2), quantidade de amostras
utilizadas para monitorar o processo, dados(2 colunas, a 1a com

```

a variável concomitante, a 2a com a variável de interesse), vetor de cores para os respectivos pontos, tamanho dos labels.

Descrição:

Essa função é utilizada por gerar os gráficos.

```

plot(dados
      , type ="b"
      , col  = cor
      , main =""
      , xlab =""
      , ylab =""
      , ylim = c(lic-13, lsc+13)
      , las  = 1
      , pch  = 16
      , xaxt = "n"
      , cex  = cex
      , cex.axis = cex/1.25
)
# Axis
axis(1, at = seq(1, sl+75,2), cex.axis = cex/1.25)
#Divisão de amostras
abline(v=sl, lty = "dotted")
#Method
text(31,lsc+10,method, cex = cex, font = 2)
## Cima
#LSC
abline(h=lsc, col = "red", lty="longdash", lwd = cex*1.25)
mtext("LSC", at=lsc, side=4,line=0.2, adj=0, las=1
      , cex = cex/1.1)
#2sigma
abline(h=lsc2, col = "black", lty ="dashed", lwd = cex)
mtext(substitute(paste(2, hat(sigma)[bar(X)][method]))
      , list(method = method))
      , at=lsc2, side=4,line=0.2, adj=0, las=1, cex = cex/1.25)
#1sigma
abline(h=lsc1, col = "grey", lty ="twodash", lwd = 1.75)
mtext(substitute(paste(1, hat(sigma)[bar(X)][method])
      , list(method = method)))

```

```

    , at=lsc1, side=4, line=0.2, adj=0, las=1, ann=F, cex = cex/1.25)
#Central
abline(h=lc)
mtext("LC", at=lc, side=4, line=0.2, adj=0, las=1, cex = cex/1.1)
## Baixo
#LIC
abline(h=lic, col = "red", lty="longdash", lwd = cex*1.25)
mtext("LIC", at=lic, side=4, line=0.2, adj=0, las=1, outer=F, cex = cex/1.1)
#2sigma
abline(h=lic2, col = "black", lty ="dashed", lwd = cex)
mtext(substitute(paste(2, hat(sigma)[bar(X)][method])))
            , list(method=method))
    , at=lic2, side=4, line=0.2, adj=0, las=1, cex = cex/1.25)
#1sigma
abline(h=lic1, col = "grey", lty ="twodash", lwd = 1.75)
mtext(substitute(paste(1, hat(sigma)[bar(X)][method])))
            , list(method=method))
    , at=lic1, side=4, line=0.2, adj=0, las=1, cex = cex/1.25)
}

# Função que engloba tudo #####
graficos_controle <- function(data, sl, k, method, snl, A, desc, delta
                               , sig0, pert, cex){

# Input (Na ordem que aparece na entra da função):
dados; quantidade de amostras utilizadas para obter os limites de
controle; tamanho de amostra; delineamento; quantidade de amostras
utilizadas para monitorar o processo; valor de A (usualmente, é 3);
indicar (via T ou F), caso T, para outras snl amostras será
considerado descontrole para esse gráfico; valor do descontrole
(delta); desvio padrão do processo sob controle; desvio padrão de
perturbação a ser incorporado no rnorm(); cex).

# Descrição
Realiza o processo após a obtenção dos dados, até a de gerar os
gráficos.

if (method == "URSS" | method == "ACO"){
  k2 <- k^2
} else if (method == "AAS"){


```

```

k2 <- k
} else {stop("A função suporta apenas os delineamentos URSS, ACO e AAS")}

samples <- data[sample(1:nrow(data), size = sl*k2, replace = TRUE), ]

# nova média do processo, caso entre em descontrole
mu <- (delta*sig0)/sqrt(k)

if (method == "URSS"){
  # Faz todo o processo de ordenação e seleção de amostra
  amostras <- urss_process(samples, sl, k2)
  amostras_mean <- rowMeans(amostras)
  # Linha Central
  mean_mean <- mean(amostras)
  mcov <- cov(amostras)
  vari <- diag(mcov)
  mcov[lower.tri(mcov, diag=TRUE)] <- 0
  # Variância da média estimada URSS
  var_mean <- (1/k2)*sum(vari)+(2/k2)*sum(colSums(mcov))
  # Limites de controle
  lic <- mean_mean - A*sqrt(var_mean)
  lc <- mean_mean
  lsc <- mean_mean + A*sqrt(var_mean)
  # 1 sigma e 2 sigmas
  lic1 <- mean_mean - 1*sqrt(var_mean)
  lsc1 <- mean_mean + 1*sqrt(var_mean)
  lic2 <- mean_mean - 2*sqrt(var_mean)
  lsc2 <- mean_mean + 2*sqrt(var_mean)

  # Coletando novas amostras do processo
  if (desc){ # Sob descontrole
    # Desloca a média do processo com a perturbação
    data$concrete <- data$concrete + rnorm(nrow(data), mu, sd = pert)
    newdata <- data[sample(1:nrow(data), size = snl*k2
                           , replace = TRUE), ]
    # Amostra URSS
    amostras_new <- urss_process(newdata, snl, k2)
    mean_amostras_new <- rowMeans(amostras_new)
  }
}

```

```

else{ # Sob controle
  newdata <- data[sample(1:nrow(data), size = snl*k2
                        , replace = TRUE), ]
  # Amostra URSS
  amostras_new <- urss_process(newdata, snl, k2)
  mean_amostras_new <- rowMeans(amostras_new)
}
}

else if (method == "ACO"){
  amostras <- aco_process(samples, sl, k)
  amostras_mean <- rowMeans(amostras)
  # Linha Central
  mean_mean <- mean(amostras)
  # Variância
  variance <- var(as.numeric(amostras))
  parte1 <- variance/k
  parte2 <- sum((colMeans(amostras) - mean_mean)^2)/(k^2)
  # Variância da média estimada ACO
  var_mean <- parte1 - parte2
  # Limites de controle
  lic <- mean_mean - A*sqrt(var_mean)
  lc <- mean_mean
  lsc <- mean_mean + A*sqrt(var_mean)
  # 1 e 2 sigmas
  lic1 <- mean_mean - 1*sqrt(var_mean)
  lsc1 <- mean_mean + 1*sqrt(var_mean)
  lic2 <- mean_mean - 2*sqrt(var_mean)
  lsc2 <- mean_mean + 2*sqrt(var_mean)

  # Coletando novas amostras do processo com descontrole
  if (desc) { # Sob descontrole
    data$concrete <- data$concrete + rnorm(nrow(data), mean = mu, sd = pert)
    newdata <- data[sample(1:nrow(data), size = snl*k2
                           , replace = TRUE), ]
    # Amostra ACO
    amostras_new <- aco_process(newdata, snl, k)
    mean_amostras_new <- rowMeans(amostras_new)
  }
  else { # Sob controle
}
}

```

```

# Coletando novas amostras do processo sem descontrole
newdata <- data[sample(1:nrow(data), size = snl*k2
                      , replace = TRUE), ]

# Amostra ACO
amostras_new <- aco_process(newdata, snl, k)
mean_amostras_new <- rowMeans(amostras_new)
}

} else{ # AAS
  # Estimando os Limites de controle
  samples <- matrix(samples[,2], ncol = k2)
  amostras_mean <- rowMeans(samples)
  # Linha Central estimada
  mean_mean <- mean(samples)
  # Variância
  sd <- mean(apply(samples, 1, sd))
  if (k == 3){
    c4 <- 0.8862
  }else{ #k = 5
    c4 <- 0.94
  }
  # Variância da média estimada AAS
  var_mean <- (A*sd)/(c4*sqrt(k2))
  var_mean1 <- (1*sd)/(c4*sqrt(k2))
  var_mean2 <- (2*sd)/(c4*sqrt(k2))
  # Limites de controle
  lic <- mean_mean - var_mean
  lc <- mean_mean
  lsc <- mean_mean + var_mean
  # 1 e 2 sigmas
  lic1 <- mean_mean - var_mean1
  lsc1 <- mean_mean + var_mean1
  lic2 <- mean_mean - var_mean2
  lsc2 <- mean_mean + var_mean2

  # Coletando novas amostras do processo
  if (desc){ # Sob descontrole
    data$concrete <- data$concrete + rnorm(nrow(data), mean = mu, sd = pert)
    newdata <- data[sample(1:nrow(data), size = snl*k2

```

```

, replace = TRUE),2]
amostras_new <- matrix(newdata, ncol = k2)
# Amostra AAS
mean_amostras_new <- rowMeans(amostras_new)
}else{ # Sob controle
newdata <- data[sample(1:nrow(data), size = snl*k2
, replace = TRUE),2]
amostras_new <- matrix(newdata, ncol = k2)
# Amostra AAS
mean_amostras_new <- rowMeans(amostras_new) }
data_ploting <- as.data.frame(c(amostras_mean,mean_amostras_new))
data_ploting$color <- ifelse(as.numeric(
data_ploting<lic | data_ploting>lsc),"red",
ifelse(as.numeric(
data_ploting<lic2 | data_ploting>lsc2)
,"gold","blue"))

if(method=="AAS"){
print(var_mean)
}else{
print(A*sqrt(var_mean))
}
plot_grafico(k, delta, method, lc, lsc, lic, lsc1, lic1, lsc2, lic2
, sl, data_ploting[,1],data_ploting[,2], cex)
}

#####
Exemplo de gráficos utilizados na seção resultados
#####

# Inputs #####
k2 <- 25
sl <- 25
k <- 5
snl <- 75
A <- 3
desc <- T
delta <- 1.2
mu0 <- 35.8
sig0 <- 16.7

```

```

pert <- 2
cex <- 2.6
data <- matrix(rnorm(10000), ncol=2)
par(mfrow=c(3,1), oma=c(5,1,2,2),mai=c(0.5,1,0.5,1))
method <- c("AAS", "ACO", "URSS")
for (i in 1:3){
  graficos_controle(data = data, sl = sl, k = k, method = method[i]
                     ,snl = snl, A = A, desc = desc, delta = delta
                     , sig0 = sig0, pert = pert, cex = cex)
  if(i == 2){
    mtext("Resistência do concreto(MPa)", side=2, line=3, adj=1, cex = cex)
  }else if (i == 3){
    title(substitute(paste("Gráficos de controle para a média com
                           k=",k," e ",delta,"=",delta_value)
                     ,list(k = k, delta_value=delta)))
    ,outer=T, line=-1, cex.main=cex*1.5)
    mtext("Amostras para estimar os LIC e LSC", side=1, line=3, adj=0
          , cex = cex/1.35)
    mtext("Amostras para monitorar o processo", side=1, line=3, adj=1
          , cex = cex/1.35)
    mtext("Identificação da amostra", side=1, line=5, cex = cex/1.25)
  }
}

```

APÊNDICE I – Obtenção do CMS para MRSS

```

median_posicao <- function(x){
  # Input: Entrar com os elementos da AAS a serem amostrados
  # Output: Retorna a posição dos elementos a serem selecionados
  # de cada conjunto

  k <- sqrt(length(x))
  if (k%%2 == 0){
    vi <- k/2
    vs <- k/2 + 1
    cc <- c(rep(vi,k/2),rep(vs, k/2))}else{cc <- rep(median(1:k),k)}
  return(cc)
}
# Example
median_posicao(1:9)

# Variância de uma estatística de ordem - aplicado para MRSS
var_est_ord <- function(ordens){
  # Input: basta entra com o output da median_posicao
  # Output: valor da variância de uma dada amostra

  k <- length(ordens)
  var_i <- rep(0,k)
  for(i in 1:k)
  {
    var_i[i] <- integrate(FdpOrdemX2,lower=-Inf,upper=+Inf,k2=k
                           ,i=ordens[i])$value
    - (integrate(FdpOrdemX ,lower=-Inf,upper=+Inf,k2=k
                           ,i=ordens[i])$value)^2
  }
  return(sum(var_i)/(k^2))
}
# Example
var_est_ord(median_posicao(1:9))

```

```

# MRSS - ordenação perfeita para CMSo = 370.508#####
cms_370_508_ord_perf_mrss <- function(aas, delta, A, l){

  % Input:
  % aas: corresponde aos tamanhos da sua amostra aleatória simples (k2)
  % aas: corresponde aos níveis de alteração da média do processo
  % Descrição:
  % Calcula o comprimento médio de sequência(CMS) no caso de ordenação
  % perfeita utilizando limites 3-Sigma

  la <- length(aas)
  ls <- length(delta)
  cms_matrix <- matrix(0, nrow = ls, ncol = la, dimname=list(c(delta),
    ,c(sqrt(aas))))
  inc <- 0
  for (a in 1:la){
    k <- median_posicao(rnorm(aas[a]))
    lk <- length(k)
    lk2 <- lk^2
    var_mean_u <- var_est_ord(k)

    % Limites de controle quando o processo está sob controle
    lsc <- A[a]*sqrt(var_mean_u)
    lic <- -A[a]*sqrt(var_mean_u)
    % Para cada nível de alteração na média do processo
    for (i in 1:ls){
      inc <- inc + 1
      set.seed(inc+67456262)
      dados <- rnorm((lk2*l), mean= delta[i]/sqrt(lk), sd= 1)

      x <- matrix(dados, nrow = lk) % primeiro vetor
      x <- apply(x,2,sort.int, method="quick") % ordena pela coluna

      % -----
      % selecao das diagonais
      % -----
      aux <- (l*lk)
      ll <- 1:aux
      cc <- rep(k, l)
    }
  }
}

```

```
% selecao dos elementos na diagonal
amostra.sel <- matrix(x[cbind(cc,ll)], ncol=lk, byrow=TRUE)

mrow_mean <- rowMeans(amostra.sel)
mrow_mean_q <- as.numeric(mrow_mean<lic | mrow_mean>lsc)
p <- sum(mrow_mean_q)/length(mrow_mean_q)
cms_matrix[i,a] <- 1/p
gc()
}
print(cms_matrix)
}
return(cms_matrix)
}
```