

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

*COMPARAÇÃO DO MÉTODO DE ANÁLISE DE DADOS ESTATÍSTICO TRADICIONAL COM O MÉTODO
BOOTSTRAP*

Vagner Inácio de Oliveira
Professor Dr. Luis Aparecido Milan
Orientador

INTRODUÇÃO

A alta capacidade de métodos estatísticos na análise de dados, sejam eles quantitativos ou qualitativos, é inquestionável. Existem diversos métodos, muito poderosos, que podem ser utilizados na retirada de informações sobre uma determinada população ou amostra. Com o avanço computacional esses métodos foram sendo utilizados com mais frequência, devido à alta capacidade de processamento dos computadores que produz em um espaço muito curto de tempo o que demandaria dias de trabalho. Neste contexto, de utilização de métodos estatísticos, é que o presente trabalho se insere.

Uma indústria de alimentos aplicou o método estatístico, denominado “*Evolutionary Operation*” (EVOP), para a melhoria da qualidade do processo de produção de leite em pó, coletou dados antes e depois desta implementação e fez análises para verificar possíveis mudanças. O objetivo do presente trabalho é verificar se a análise feita com o método *Bootstrap* é similar a feita com o método *EVOP*, com os mesmos dados de antes e após a implementação do EVOP.

METODOLOGIA

Atualmente o mercado da indústria de alimentos se mostra exigente e competitivo o que intensifica o processo de melhoria contínua da qualidade e de desempenho do processo. Os sistemas de garantia de qualidade adquiriram um enfoque proativo, deixando de ser unicamente um controle de processo e passando a buscar a melhoria contínua e a solução antecipada de problemas. Diante deste quadro, o método “*Evolutionary Operation*” (EVOP) foi aplicado em um processo de produção de secagem de leite em pó, composto por evaporador tubular de película descendente e um secador atomizador de disco centrífugo, de uma indústria de alimentos. Este método utiliza técnicas estatísticas simples e eficazes, é desenvolvido diretamente na linha de produção sem promover interrupções do processo, sem demandar grande custo adicional e tempo de desenvolvimento de experimentos. A meta do programa EVOP foi minimizar a quantidade de leite em pó classificada como sedimento B, por meio da otimização das variáveis: teor de sólidos do concentrado (TS) e vácuo produzido na câmara de secagem. Outro objetivo do programa EVOP foi avaliar a influência do teor de sólidos e do vácuo sobre a vazão de leite fluido processado (*vazão*) e sobre o rendimento do processo de secagem (*rendimento*).

Para este trabalho serão feitas reamostragens pelo método bootstrap das amostras obtidas do processo de secagem do leite e feitos os cálculos necessários para a verificação de similaridade dos resultados finais com o método EVOP.

O MÉTODO EVOP (*Evolutionary Operation*)

Quase todos os processos industriais existentes são passíveis de melhoria na produtividade. Este potencial surge não apenas em função de falhas no planejamento original, mas também de mudanças que podem ocorrer durante a rotina de produção. A melhoria de um processo industrial introduz três limitações que não são relevantes para estudos em pequena escala. Primeiro, a maioria dos processos envolve múltiplas etapas ou estágios - mudanças em um estágio em particular devem ser introduzidas cuidadosamente para evitar romper com os procedimentos normais de operação em outra parte do processo. Segundo, devido a grandes quantidades envolvidas, deve-se evitar a fabricação de produtos fora da especificação. Finalmente, é importante que o produto permaneça consistente; embora o produto possa ser melhorado, propriedades diferentes do produto podem causar problemas para consumidores que estão satisfeitos com as características atuais. A violação de qualquer uma dessas limitações normalmente resulta em perdas econômicas desnecessárias.

Percebendo os problemas de experimentação em escala industrial, a imprecisão na determinação das condições ótimas em escala de laboratório e a aparente complexidade dos métodos de planejamento de experimentos, BOX (Box, 1957) propôs na década de 50 uma abordagem que contorna essas dificuldades e ainda encontra um caminho sistemático para atingir as condições ótimas do processo (MYERS & MONTGOMERY,). Assim sendo, a proposta desse método é o monitoramento e a melhoria contínua do processo em escala industrial alterando as condições operacionais atuais para condições ótimas.

Para BOX e DRAPER (1998) a EVOP produz informações de como melhorar as condições de operação atuais sem os riscos, pessoal especializado e as despesas geralmente associadas com experimentação em uma planta industrial. BOX & DRAPER (1998) afirmam que EVOP não é um método que substitui a investigação fundamental empregadas em estudos exploratórios em laboratórios, como por exemplo, a metodologia de superfície de resposta; sendo desenvolvida para evitar características indesejáveis de experimentos de processo em escala industrial. Segundo BOX, HUNTER & HUNTER (1978) para evitar mudanças apreciáveis nas características do produto são feitas somente pequenas mudanças nos níveis das variáveis do processo. Para determinar os efeitos dessas mudanças é necessário repetir cada experimento várias vezes e calcular a média das observações.

A Operação Evolutiva (EVOP) foi planejada para ser conduzida por operadores num processo de produção em escala industrial, sem, contudo impedir ou alterar a produção satisfatória de produtos simultaneamente a sua aplicação. Deste modo, as circunstâncias são muito diferentes daquelas de laboratório, em que tempo e dinheiro adicionais devem ser empregados para realização dos experimentos. Considerando a situação em que EVOP é utilizada, tem-se que manter a relação sinal-ruído baixa, assim, um grande número de experimentos é normalmente necessário para revelar os efeitos das mudanças; sendo imprescindível a

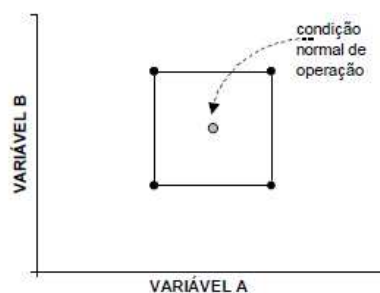
realização destes, que resultam em um custo adicional baixo (BOX, HUNTER & HUNTER, 1978).

Características do método EVOP

Para SCARRAH (1987), a utilização de algumas técnicas estatísticas simples pode ajudar a evitar perdas econômicas ocasionadas pela produção de produtos fora de especificação. Com a condução de vários ciclos em cada etapa, os pequenos efeitos virtualmente não detectados no produto durante as fases individuais do experimento podem ser avaliados utilizando técnicas estatísticas. Segundo DEMING & ÖBERG (2000), SAAD (1994); BOX, et al. (1978) outra vantagem da análise estatística é que as interações entre variáveis podem ser facilmente identificadas.

No processo de desenvolvimento de Operação Evolutiva (EVOP) BOX & DRAPER (1998) utiliza o planejamento de experimento fatorial 2^k para estudar efeitos de até três fatores simultaneamente. Esses fatores podem ter sido identificados para seleção de variáveis em um experimento conduzido para descobrir que fatores do processo influenciam significativamente a variável resposta. Esse tipo de planejamento fatorial consiste na combinação de todas as variáveis em seus dois níveis. Estes delineamentos freqüentemente são acrescidos de um ponto de referência, normalmente ao centro do delineamento, para formar um delineamento fatorial composto. Recomendamos selecionar as condições operacionais atuais para esse ponto de referência (DEMING & ÖBERG, 2000; SCARRAH, 1997; BOX, et al. 1978). SCARRAH (1997) descreve que as finalidades do ponto de referência são identificar os efeitos de curvatura associados com as variáveis e servir de base para a estimativa do custo do estudo comparando as respostas do processo obtidas do delineamento fatorial com aquelas das condições de operação normais.

Em um delineamento fatorial, um experimento exige a combinação de todas as variáveis do processo em seus dois níveis, geralmente distribuídos acima e abaixo do ponto de referência. Os tratamentos do fatorial correspondem aos vértices do esboço representado na Figura 1, no qual se têm quatro tratamentos para estudar a influência de duas variáveis.



Figural: Um delineamento fatorial 2^2 nas variáveis A e B.

De acordo com BOX & DRAPER (1998) e HAHN & DERSHOWITZ (1974) as principais características de processos favoráveis à aplicação do EVOP são: O processo deve ser repetitivo, ou o trabalho de produção deve durar um tempo razoavelmente longo para justificar os esforços de otimização; a EVOP deveria ser aplicada em áreas onde os potenciais benéficos de melhoria do produto são grandes, no entanto, esta não é uma limitação séria, isto porque raramente um processo está sendo operado

em condições ótimas; a terceira característica é a capacidade de perturbar facilmente as variáveis do processo; o processo deve estabilizar rapidamente depois de perturbado, já que as futuras condições são determinadas a partir da última execução e é desejável que a variável resposta tenha uma medida rápida.

HAHN & DERSHOWITZ (1974) destaca como vantagens da Operação Evolutiva (EVOP) a melhoria do produto, a melhoria da eficiência do processo, o aumento do desenvolvimento com o desempenho do processo por parte dos operadores. NAKAI (1981) e KRAMER (1965) afirmam que a técnica é vantajosa porque o método pode ser aplicado diretamente na linha de processamento e, portanto, pode ser apropriado para o controle de qualidade on-line, e não somente para pesquisa e desenvolvimento de produtos.

Com relação às desvantagens, HAHN & DERSHOWITZ (1974) aponta custos de tempo e dinheiro no treinamento de pessoal nas etapas de mudança dos níveis das variáveis e os custos associados com a manutenção e análise de registros e documentos. Esses custos podem ser modestos, mas como na maioria dos investimentos, procedem aos lucros e, considerando sua natureza especulativa, esses lucros infelizmente são sempre incertos.

Pode-se argumentar que por perturbar o processo de suas condições atuais, que podem ser ótimas, tolera-se a possibilidade de incorrer em algumas perdas. No entanto, esse argumento é considerado irreal para a maioria dos processos, visto que as condições ótimas na maioria dos casos não são as condições operacionais atuais. Essa, na realidade, é a razão para se utilizar o método EVOP.

O SISTEMA DE SECAGEM DO LEITE COM A EVOP

O processamento de leite em pó tem como pré-requisito óbvio a remoção de água do leite. Esta remoção de água é realizada em duas etapas, a primeira etapa é a concentração e a segunda é a secagem.

A tecnologia atual permite basicamente três abordagens para promover a concentração do leite: Evaporação; Osmose reversa e Ultra filtração. No presente trabalho verificou-se que as variáveis que realmente inferiam mudanças significativas no processo de secagem do leite são o de evaporação e o de ultra filtração.

Com isso foi implementado o método EVOP no processo de produção junto às variáveis concentração e vazão de leite processado.

MÉTODO BOOTSTRAP

Técnica introduzida por Efron no final da década de 70. A terminologia é basicamente uma técnica de reamostragem, que permite aproximar a distribuição de uma função das observações pela distribuição empírica dos dados baseada em uma amostra de tamanho finita. A amostragem é feita, com reposição, da distribuição da qual os dados são obtidos, se esta é conhecida (*bootstrap Paramétrico*) ou da amostra original (*bootstrap não-paramétrico*). Neste último utilizamos a hipótese de que as observações são obtidas da função de distribuição empírica $F(x)$, que designa uma massa de probabilidade igual $1/n$ para cada ponto amostral.

O *bootstrap* aborda o cálculo do intervalo de confiança de parâmetros e cálculos de valores-p, em circunstâncias em que outras técnicas não são aplicáveis. Esta técnica foi extrapolada para a solução de muitos problemas de difícil resolução através de técnicas de análise estatística tradicionais, baseadas na hipótese de um elevado número de amostras.

A técnica *bootstrap* tenta realizar o que seria desejável na prática, se tal fosse possível: repetir o experimento. As observações são escolhidas de forma aleatória e as estimativas recalculadas. A idéia básica da técnica *bootstrap* é que quando não temos toda a população de amostras (observações) podemos utilizar o conjunto amostrado $\mathbf{x} = (x_1, \dots, x_n)$ para análise. A técnica *bootstrap* trata a amostra original como se esta representasse exatamente toda a população (conjunto de experiências, realizações). Segundo Proença (1988), até agora o *bootstrap* chega aos mesmos resultados que o processo tradicional, baseado na máxima verossimilhança. A sua grande virtude consiste em apresentar solução para casos em que a dedução da precisão da estimativa, de seu viés e do Estimador de Mínimos Quadrados (EMQ) aparenta ser impossível ou mesmo demasiado complexa. É especialmente para estes casos que se utiliza o método *bootstrap*.

No presente trabalho este método será utilizado apenas para a verificação de similaridade dos resultados obtidos pelo método EVOP.

PROCEDIMENTO PARA A OBTENÇÃO DA AMOSTRA BOOTSTRAP

Seja uma amostra original e a estatística de interesse abaixo:

$$\mathbf{x} = (x_1, \dots, x_n) \quad \hat{\theta} = F(\mathbf{x})$$

1º) Geram-se as amostras *bootstrap* $x^{*(1)}, \dots, x^{*(B)}$ com reposição de \mathbf{x} .

2º) Calculam-se as estimativas da estatística de interesse:

$$\hat{\theta}_{(b)} = F[x_{(b)}], b = 1, \dots, B$$

3º) Calcula-se o erro padrão *bootstrap*, \widehat{S}_b , dado por:

$$\widehat{S}_b = \frac{1}{B-1} \left\{ \sum_{b=1}^B [\hat{\theta}_b - \hat{\theta}_{(*)}]^2 \right\}^{1/2} \quad \text{onde,} \quad \hat{\theta}_{(*)} = \frac{\sum_{b=1}^B \theta_{(b)}}{B}$$

O procedimento acima se aplica ao caso do *bootstrap* não-paramétrico. Para utilizar o *bootstrap* paramétrico, procede-se a mesma forma, com a única diferença de que cada amostra *bootstrap* é obtida da distribuição paramétrica que originou os dados que se têm em mãos, ao invés de reamostrar as observações disponíveis. Desde o aparecimento do *bootstrap*, vários autores vêm tentando estabelecer confirmação empírica ou teórica da sua validade. Devido ao fato desta técnica atuar como um método de aproximação de distribuições, todas as provas de consistência e precisão dos estimadores são resultados assintóticos, mas sua validade pode ser estendida para espaços amostrais finitos. As conclusões a que chegaram estes autores são que a aproximação *bootstrap* é válida para a maioria das estatísticas de interesse e que seus estimadores são consistentes.

Se $B \rightarrow \infty$, então as estimativas do erro-padrão, do enviesamento e do EMQ se igualam às estimativas de máxima verossimilhança (Efron, 1982). Para o cálculo

das estimativas *bootstrap* geralmente é suficiente um valor de B=100. Contudo, para se determinar a distribuição por amostragem com precisão deve considerar-se um valor para B substancialmente mais elevado. Segundo Efron (1982), geralmente B = 1000 proporciona bons resultados. E em ambos os casos, convém ensaiar diferentes valores para B até se verificar a convergência dos resultados.

Este trabalho pretende utilizar as amostras colhidas no processo de produção da indústria, antes e depois, e amostrá-los com o método *bootstrap* um certo número de vezes. Será utilizado o método *bootstrap* paramétrico, o qual utiliza a distribuição conhecida dos dados coletados e, o método não paramétrico, que utiliza a função de distribuição empírica. Com isso, serão feitas as análises necessárias e verificada a hipótese de similaridade com os resultados do método EVOP.

CONCLUSÃO

Após a aplicação do programa EVOP ao processo, verificou-se a redução da quantidade de leite em pó classificada como sedimento B de 4,23% para 0,25% da produção diária: No entanto, para as variáveis respostas vazão de leite fluido processado e rendimento do processo não foram verificados efeitos estatisticamente significativos. Além do ganho de produtividade, a utilização do método EVOP promoveu melhor entendimento do processo pelo pessoal de operação, permitindo uma evolução contínua do processo em estudo.

BIBLIOGRAFIA

- BOX, G. E. P. Evolutionary operation: a method for increasing industrial productivity. **Applied Statistics**, 6, 81-101. 1957.
- BOX, G.E.P.; DRAPER, N. R. **Empirical model-building and responses surfaces**. 1.ed. New York: John Wiley & Sons. 1987. 669p.
- BOX, G.E.P.; HUNTER, W.G.; HUNTER, J.S. **Statistics for experimenters - An Introduction to Design, Data Analysis and Model Building**. John Wiley & Sons. Inc. New York. 1978. 653p.
- Efron, B. (1979), **Bootstrap Methods: another look at Jackknife**, **Annals of Statistics**, Vol. 7, 1-26.
- FOX, M.M. EVOP; Tool for in-plant research. **Food Technology**. Vol. 22, p 293-300. 1968.
- MYERS, R. H. ; MONTGOMERY, D. C. **Response Surface Methodology - process and product optimization using designed experiments**. New York, Wiley Interscience. 1995. 700p.
- SCARRAH, W.P. Improve production efficiency via evolutionary operation. **Chemical Engineering**. 94(18)131-133. December 7, 1987.
- SOBRINHO, Paulo de Souza Costa. ; **Utilização da metodologia operação evolutiva (EVOP) para melhoria de processo na indústria de alimentos**. Tese (Doutorado em Ciência e Tecnologia de Alimentos)- Departamento de Tecnologia de Alimentos UFV. Viçosa :UFV, 2002.