



## ■ Uso de ferramentas estatísticas para otimização de vestimentas

### Introdução

Eventualmente, todos que trabalham em máquinas de papel e celulose já se depararam com as seguintes situações:

- A tela formadora teste atingiu um nível de teor seco maior do que a tela padrão da posição?
- O feltro *pick up* modificado reduziu o número de rompimentos de folha?
- As telas secadoras em experiência diminuiram a marcação da folha?
- O desenho da manta influencia na performance da máquina?

Essas são apenas algumas questões que precisamos responder diariamente para manter a eficiência da máquina e um nível de custos abaixo da meta estipulada. Este artigo abordará algumas técnicas simples, outras um pouco mais sofisticadas, para auxiliar na tomada de decisão sobre qual vestimenta é mais adequada para cada máquina e posição.

Para isso, vamos conversar principalmente sobre estatística. Sabemos que mesmo os profissionais com formação em exatas tendem a fugir de uma discussão sobre este tema rapidamente, principalmente pelo número de convenções de contorno, critérios de comparação, tamanho de amostra, normalidade, testes de hipótese e o temido p-valor, que entre outros podem inviabilizar qualquer teste.

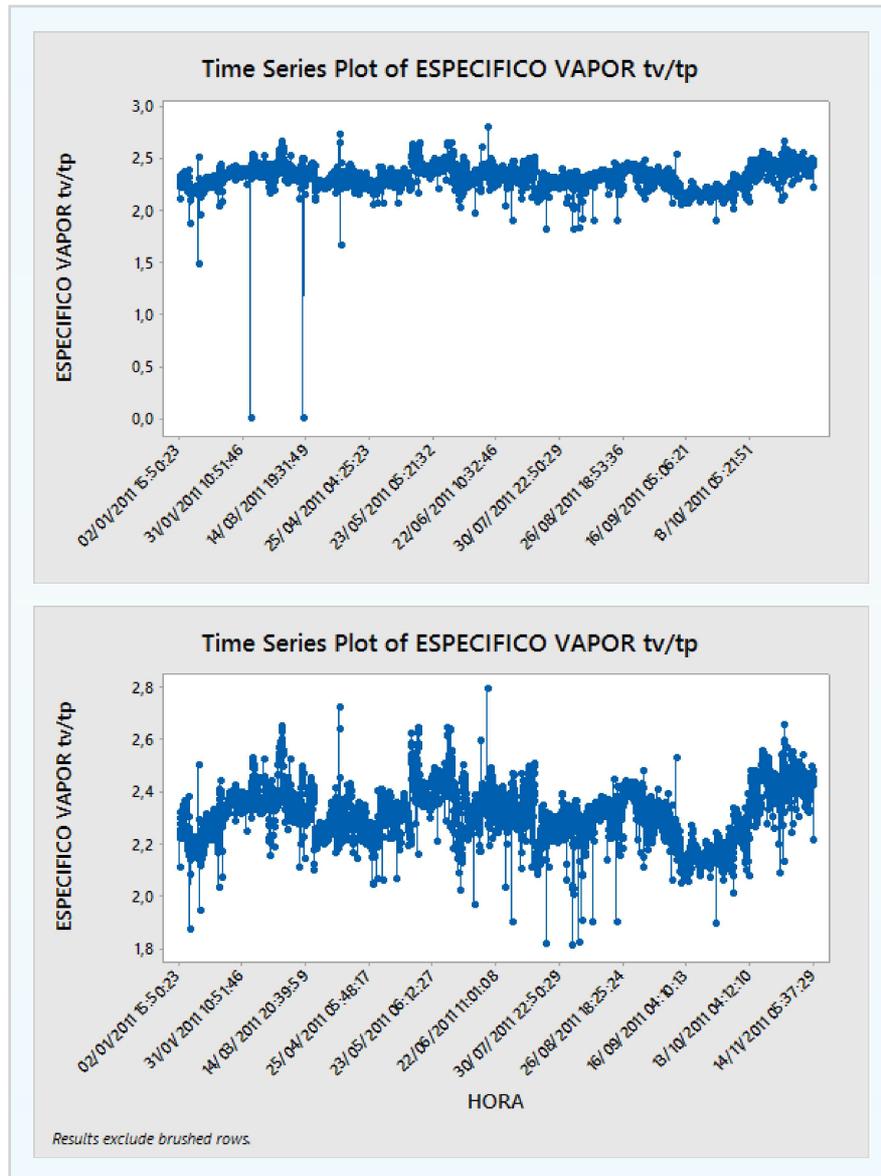
Convém salientar que não se trata de uma má vontade generalizada com a estatística. Nas universidades infelizmente não nos ensinam estatística suficiente para podermos analisar com conforto e com segurança os dados que chegam em nossas mãos, e esta dificuldade não está concentrada somente no setor de papel e celulose: praticamente todos os setores que precisam utilizar ferramentas estatísticas frequentemente apresentam a mesma dificuldade. Só para se ter uma ideia, um professor de Stanford chamado John P. Ioannidis publicou um artigo em 2005, intitulado: "*Why Most Published Research Findings Are False*" – Por que a maioria das descobertas científicas publicadas é falsa –, em uma tradução livre. Neste artigo, ele comenta os erros de interpretação comumente encontrados nos artigos, sendo o mais comum o da interpretação do p-valor.

O objetivo aqui é mostrar a quantidade de ferramentas que podem ser usadas para análise, conversar brevemente sobre o modo de funcionamento de cada uma e alertar para os riscos envolvidos em uma análise equivocada, principalmente quando expostos aos temidos falsos positivos, que podem nos assombrar por anos até que sejam efetivamente desmascarados.

Utilizarei, neste artigo, gráficos e resultados gerados no Minitab, apenas por uma questão de familiaridade. Afinal, existem outros softwares estatísticos disponíveis no mercado. Além disso, a grande maioria dos exemplos pode ser feita também no Excel, que possui um pacote estatístico razoavelmente completo.

### Gráficos de linha de tempo

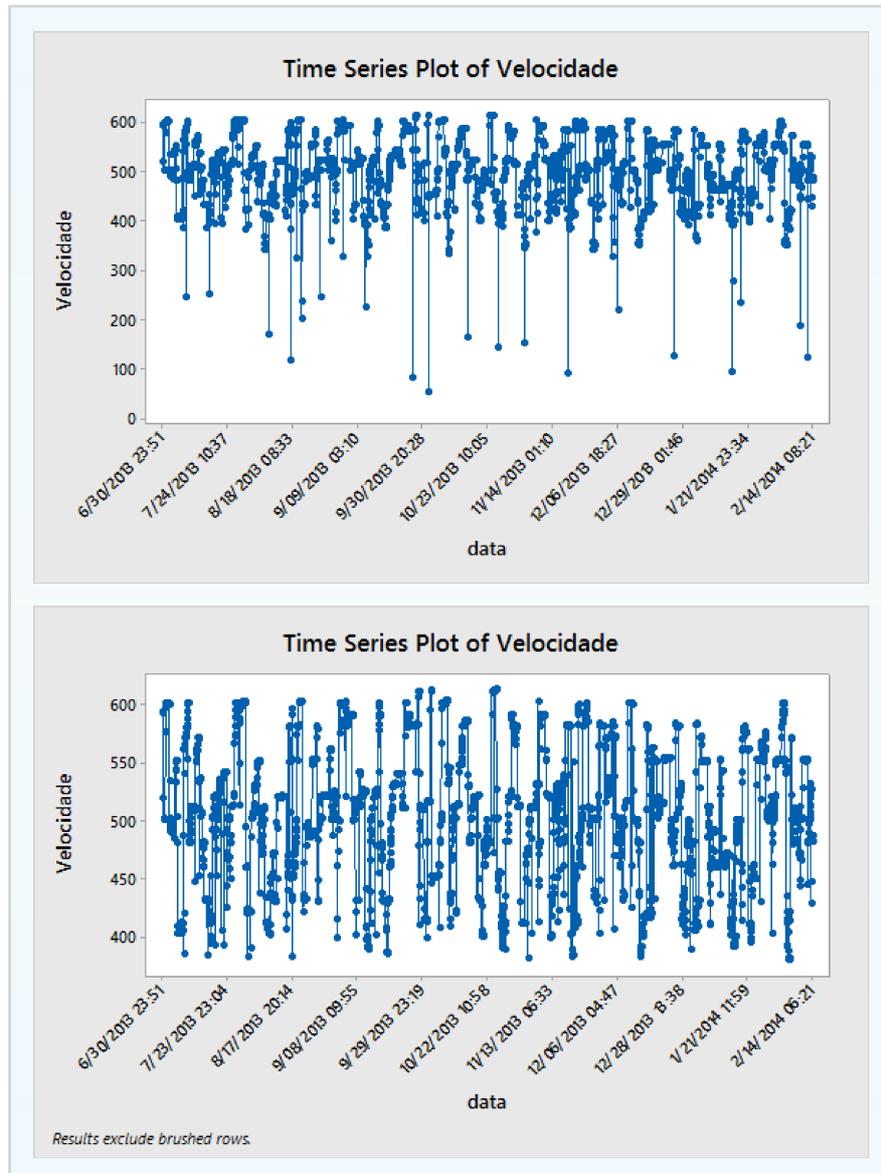
Praticamente toda a análise nas máquinas de papel e de celulose começam com um simples gráfico de variável versus tempo, como o que podemos ver a seguir:



Gráficos 1 e 2: Linha de tempo com e sem pontos extremos.

Neste gráfico estão os resultados de consumo específico de vapor por tempo. A primeira tarefa é tirar os famosos *outliers*, ou seja, os pontos que estão muito fora do que consideramos normal e que podem alterar a análise. Notem que, no gráfico da esquerda, a variação no consumo de vapor é pouco perceptível; já no da direita, sem os *outliers*, a variação se torna muito mais evidente.

Como eram poucos os pontos extremos, a análise da média, por exemplo, praticamente não muda, com 2,3071 tv/tp para o gráfico da esquerda, e 2,3094 tv/tp para o gráfico da direita. No entanto, nem sempre é isso o que ocorre, veja mais um exemplo:



Gráficos 3 e 4: Linha de tempo com e sem pontos extremos.

Neste caso, a média do gráfico da direita foi de 503,45 m/min, e o da esquerda 508,57 m/min. Parece pouco? Vamos assumir que a gramatura produzida seja de 220 g/m<sup>2</sup>, com uma largura de 5 m, e que a máquina opera 350 dias por ano. Arredondando o resultado, serão 2.838 t a mais, ou a menos, dependendo de qual gráfico você utilizou para coletar os dados e mostrar para seu gerente ou diretor que, seguramente, ficará aguardando ansioso por explicações.

A dica é: antes de começar qualquer trabalho, construir um gráfico com linha de tempo para visualizar seus dados. Assim, você garante que os pontos extremos que podem prejudicar os estudos foram devidamente retirados. Você pode estar se perguntando: então devo simplesmente ignorar os pontos extremos? A resposta é sim e não. Dependendo do que estiver avaliando, você pode retirar os pontos extremos sem nenhum peso na consciência, como no caso da velocidade de máquina que analisamos. Contudo, se estiver verificando, por exemplo, a carga de um rolo acionador, você estará muito interessado nos pontos extremos, pois provavelmente seus problemas de máquina ocorrem quando os atinge. Portanto, cuidado ao ignorar os *outliers*.

Ainda falando sobre gráficos de tempo, vamos verificar em seguida os efeitos da variação e como identificá-los graficamente. O exemplo trata sobre um teste físico do papel.

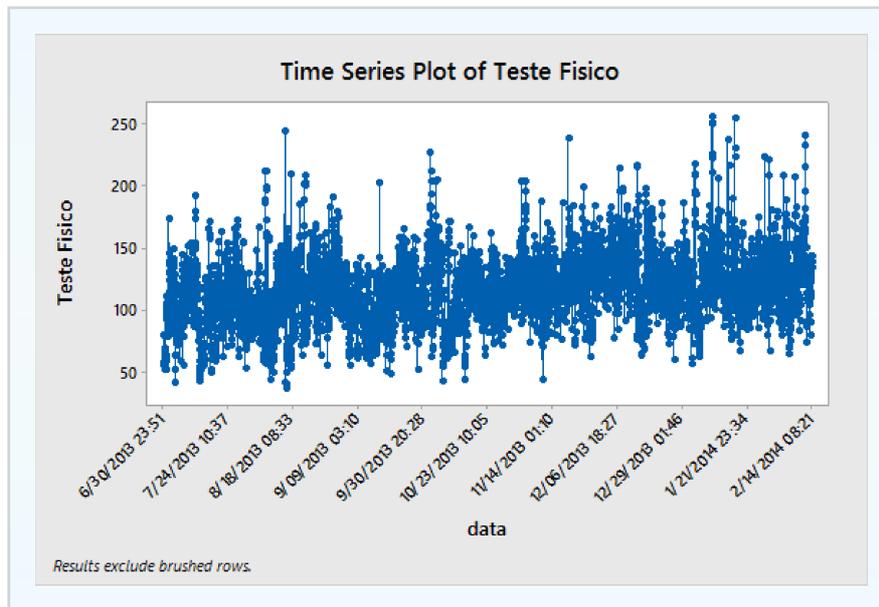
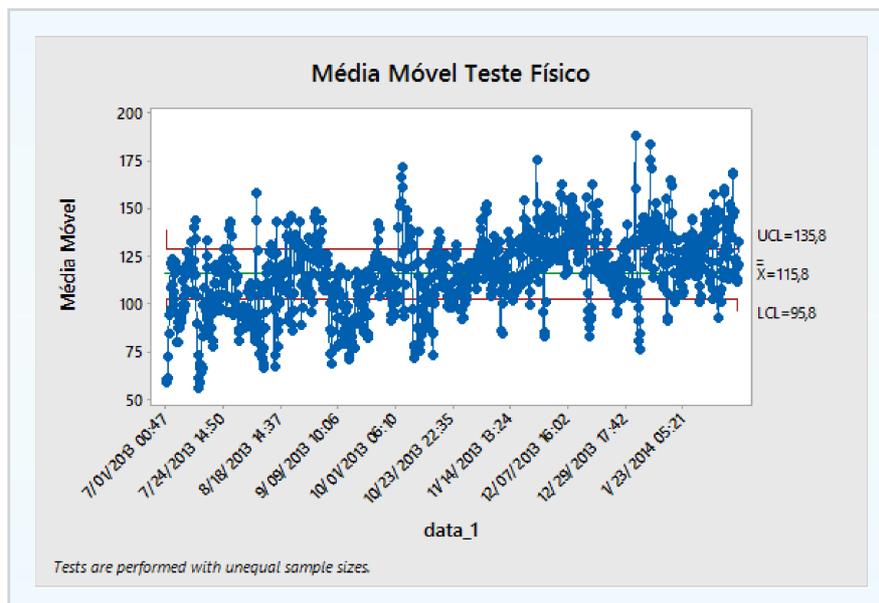
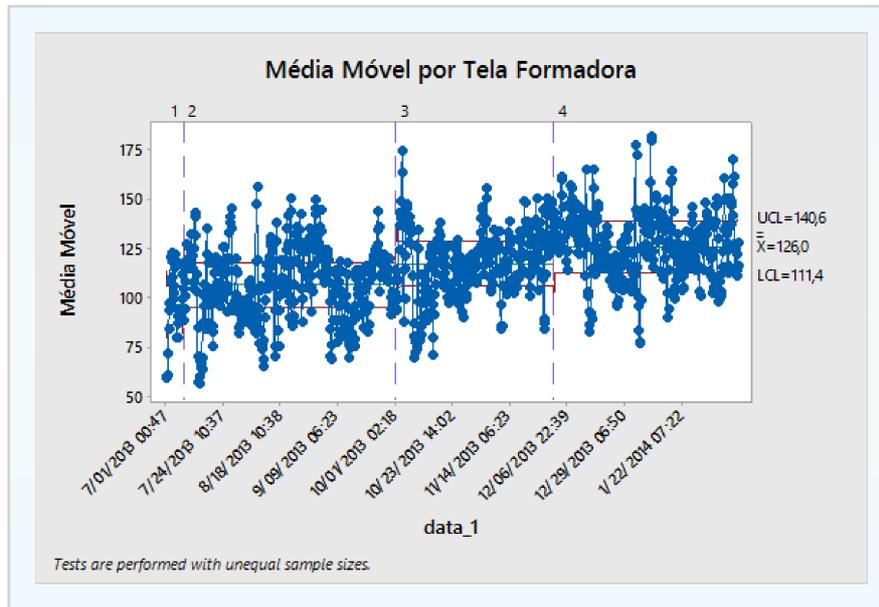


Gráfico 5: Linha de tempo para teste físico do papel.

Note que existe uma pequena tendência de aumento no valor, mas não é tão fácil assim identificar quando esta tendência começou. Por consequência, fica difícil notar este incremento em um gráfico de linha de tempo comum, principalmente se o visualizar dia a dia. Imagine se estamos bem no centro do período do gráfico 5, nada vai parecer errado, certo?

No mercado de ações, pequenas alterações nas tendências dos preços das ações precisam ser detectadas com rapidez. Por isso, um dos gráficos utilizados lá pode ser útil para nós também: é o gráfico de média móvel. Não entrarei na teoria de como ele é construído, o que importa aqui é saber que, neste tipo de acompanhamento, pequenas mudanças de trajetória nas curvas são melhor visualizadas do que em um gráfico comum. Veja os gráficos 6 e 7.





Gráficos 6 e 7: Média móvel com e sem categorização.

Note que com o gráfico de média móvel fica bem mais clara a tendência de aumento do teste físico, ou seja, o incremento no valor do teste poderia ter sido detectado antes. Se suspeitarmos que a tela formadora possa ter algum efeito neste comportamento, podemos inserir o momento de troca das vestimentas, como fiz no gráfico 7. Aparentemente não existe mudança de patamar com a substituição das telas. O mesmo pode ser feito para qualquer outra vestimenta ou elemento da máquina onde pare alguma suspeita.

Existe ainda mais um tipo de gráfico de tempo bastante útil quando estamos buscando indícios de variações no processo. É o gráfico de CUSUM ou gráfico de somas cumulativas. Mais uma vez, não entrarei na teoria por trás da ferramenta, o importante é saber que ele nasceu para ser uma ferramenta poderosa na detecção de mudanças. Vamos analisar o comportamento do mesmo teste físico com o gráfico de CUSUM.

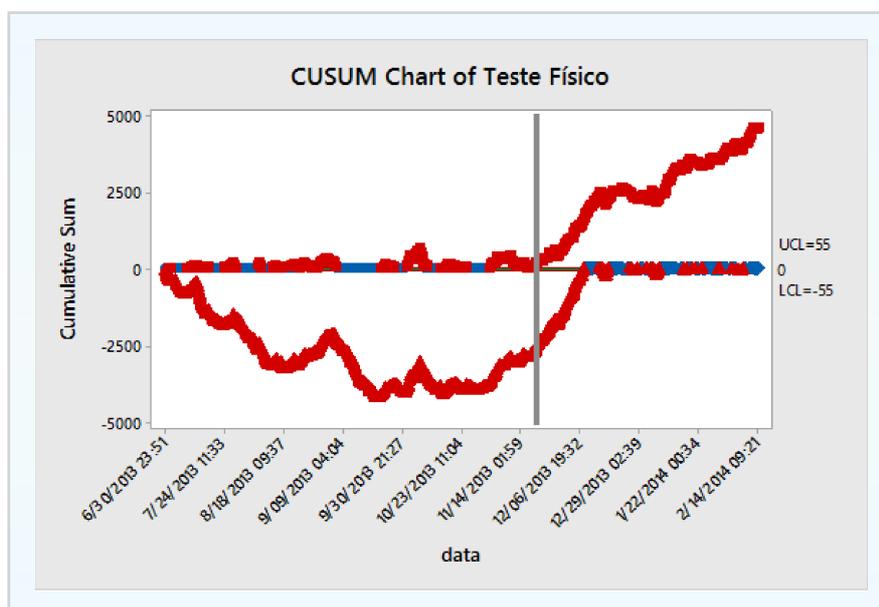


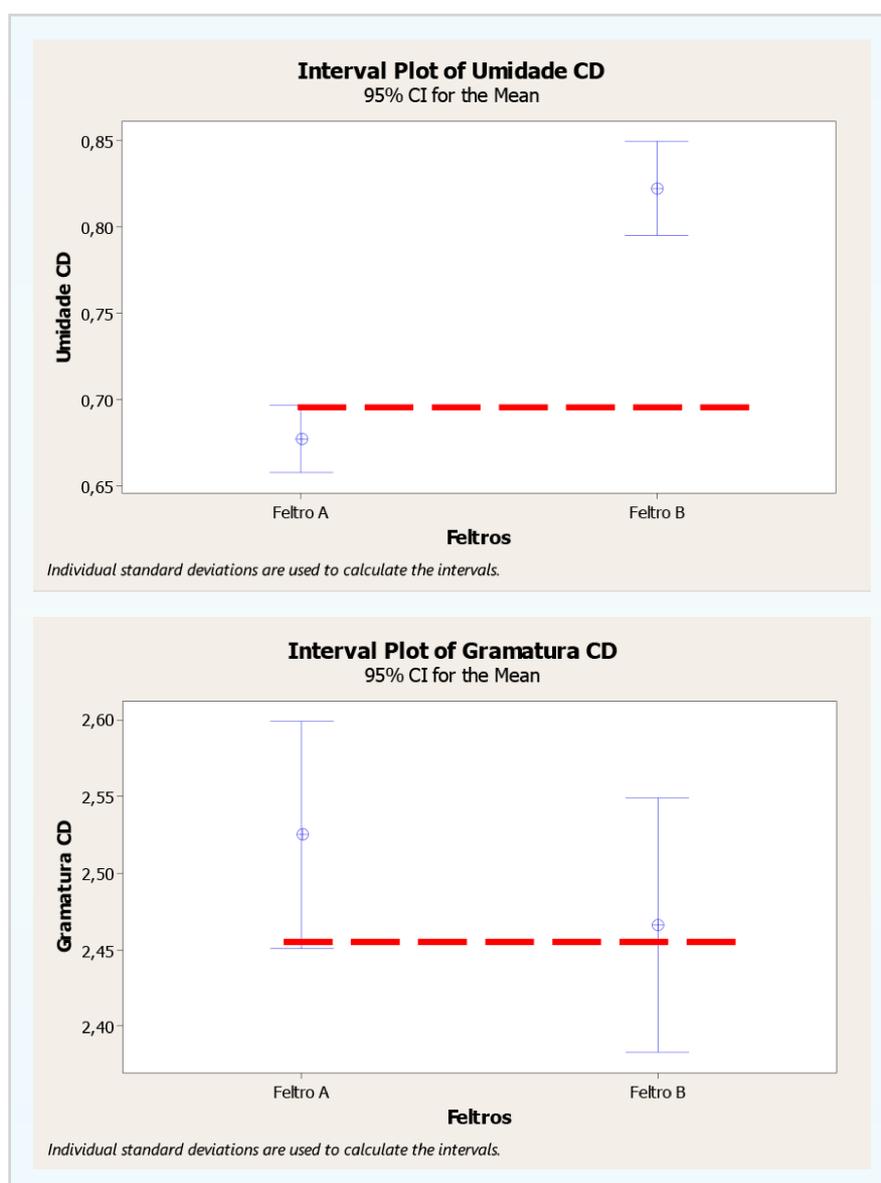
Gráfico 8: CUSUM.

Para analisar este gráfico basta saber o seguinte: a linha de baixo mostra a tendência do valor do teste diminuir, já a de cima, a de aumentar. Coloquei uma linha no ponto onde o valor deixa de ficar estável e começa a apresentar uma tendência de alta. Este é um bom ponto de partida para verificar o que estava diferente, ou foi modificado naquela data. Assim como o gráfico da média móvel, o CUSUM é muito bom em detectar mudanças antes que elas possam ser efetivamente percebidas usando um gráfico de linha do tempo comum.

### Estudos comparativos

Se entramos com uma vestimenta nova em uma máquina que promete um desempenho melhor, já sabemos como identificar as diferenças usando gráficos de linha de tempo. Agora será necessário saber se a vestimenta em teste é realmente superior à utilizada normalmente na posição, ou seja, se existe uma diferença estatística entre elas. Para isso, será necessário utilizar outro tipo de ferramenta.

Nos gráficos 9 e 10, podemos ver um estudo comparativo utilizando gráficos de intervalo, talvez este seja o meio mais rápido de verificar se existe ou não diferença entre as vestimentas analisadas.



Gráficos 9 e 10: Gráficos de intervalo para CD de umidade e gramatura.

No gráfico 9 temos os resultados do CD de umidade para cada feltro. Ao traçarmos uma linha do intervalo do feltro A até o feltro B, podemos ver que não se encontrarão em nenhum momento, pois o intervalo do feltro B está bem acima dos valores encontrados no A. Já no gráfico 10, a mesma linha traçada acaba interceptando o intervalo do feltro B. Estatisticamente, isso significa que, no caso do CD de umidade, o feltro A obteve valores menores, e além disso, os resultados obtidos têm diferença significativa. O mesmo não ocorreu com a comparação para o CD de gramatura, pois os intervalos se tocam em vários valores, ou seja, apesar das médias serem diferentes, os intervalos como um todo não são.

Você pode se perguntar: então devo escolher o feltro A, pois apresentou um valor de CD de umidade melhor? A resposta é não, até que você tenha certeza de que foi somente o feltro, e não outra variável da máquina que casualmente foi alterada durante a operação da vestimenta escolhida. Uma maneira de fazer isso é verificar o gráfico de linha de tempo: se houve uma mudança de patamar que coincidiu com a entrada do feltro, então o feltro realmente fez a diferença, caso contrário, o resultado obtido é tão somente obra do acaso.

Se precisar de mais um embasamento para verificar se realmente existem diferenças entre vários tipos de vestimentas, podemos incrementar um pouco a análise comparativa usando análise de variância de um fator, ou mais conhecida como ANOVA. Novamente não entrarei na teoria por trás da ferramenta, vamos focar nos resultados obtidos.

No gráfico 11 pode-se observar os resultados obtidos sobre a influência do feltro *pick up* na velocidade da máquina. Foram utilizados no período quatro feltros diferentes.

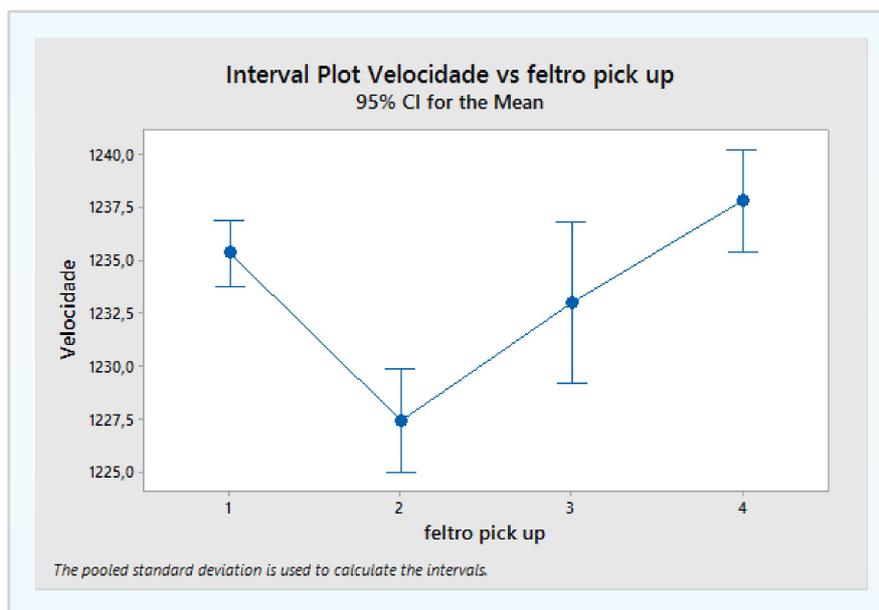


Gráfico 11: Gráfico de intervalo para velocidade.

Notem que utilizando somente o gráfico de intervalo, fica mais complicado avaliar quais feltros apresentaram resultados efetivamente distintos dos demais. Por isso, também foi realizado um teste dentro do ANOVA, que se chama Tukey. Basicamente, este teste agrupa entre semelhantes e não semelhantes, e o resultado da análise está demonstrado na tabela a seguir.

### Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

pick up	N	Mean	Grouping
4	377	1237,82	A
1	914	1235,34	A
3	152	1233,01	A B
2	368	1227,44	B

Means that do not share a letter are significantly different.

Tabela 1: Resultados teste Tukey.

Usando o resultado da tabela, agora sabemos que os feltros 1, 3 e 4 apresentaram resultados estatisticamente iguais e o mesmo vale para os feltros 2 e 3. Além disso, o feltro 3 apresentou resultados que podem ser considerados iguais ao grupo A e ao grupo B. Somente o feltro 2 não pertence ao grupo A.

Para dar um pouco mais de rigor à análise, convém assumir algumas hipóteses antes de fazer os testes. No caso dos feltros, montamos o teste da seguinte maneira:

H0, ou a hipótese nula: assume que todas as médias são iguais;

H1, ou hipótese alternativa: assume que pelo menos uma média é diferente;

Utilizarei um nível de significância,  $\alpha = 0,05$ . Este valor pode ser alterado, mas é o mais utilizado em qualquer tipo de estudo e tornou-se uma espécie de padrão devido ao uso.

O teste obteve um p-valor = 0,012, que é menor do que o  $\alpha$ , ou nível de significância que escolhi. Neste caso, a hipótese nula deve ser descartada e se instala a hipótese alternativa de que pelo menos uma média será diferente.

O p-valor nasceu para auxiliar na obtenção de conhecimentos sobre resultados de testes comparativos, já que os cientistas precisavam saber se os dados eram efetivamente diferentes, ou eram simples produtos da sorte. Convém lembrar que o p-valor pequeno por si só não garante o quanto você está certo a respeito de suas hipóteses: ele atua mais como um termômetro de surpresa, uma vez que inicialmente assumimos que não existe diferença entre as médias. Quanto mais baixo o p-valor, mais "surpresos" ficamos com os resultados obtidos.

O estatístico inglês Ronald Fisher talvez tenha sido o maior disseminador do p-valor no início do século 20. Foi ele que introduziu o conceito de que quanto menor o p-valor, mais surpresos devemos ficar com os resultados obtidos, e que deveríamos utilizá-lo para interpretar os dados praticamente em todos os casos. Alguns anos depois, Jerzy Neyman e Egon Pearson se depararam com algumas perguntas que não podiam ser respondidas confiando somente no valor baixo do p-valor. Uma das questões era saber se os resultados indicavam valores realmente confiáveis, ou se tratavam de falsos positivos. É impossível eliminar completamente os falsos positivos, visto que com isso corremos o risco de tomarmos decisões rapidamente e assumirmos riscos maiores do que o desejável.

Neyman e Pearson escreveram sobre este dilema: “Estamos inclinados a pensar que enquanto uma hipótese particular é construída, nenhum teste baseado na teoria da probabilidade pode, por si mesmo, fornecer alguma evidência de valor da verdade ou da falsidade da hipótese”. Resumindo, utilizem as ferramentas com responsabilidade e não vão com muita sede ao pote somente porque o p-valor encontrado foi baixo, ok?

## DoE (Design of Experiments) – Planejamentos de Experimentos

Até este momento, verificamos como visualizar os dados nas linhas de tempo e fazer um estudo comparativo entre variáveis. Porém, foram utilizados exemplos de apenas uma variável ou vestimenta. No caso anterior do feltro *pick up*, comparamos apenas o efeito de uma única vestimenta, mas e se quisermos saber, por exemplo, qual estilo de tela formadora e feltros propiciam uma melhor condição de produção, ou qual combinação de estilos de feltro tem maior impacto na qualidade da folha?

O DoE é bastante utilizado no meio acadêmico para planejar experimentos, onde selecionam-se alguns fatores que podem causar uma variação em uma determinada resposta. Na sequência é montada uma tabela com a ordem dos experimentos e das variáveis. O DoE também é amplamente utilizado na indústria, principalmente a alimentícia, onde eles determinam, por exemplo, as quantidades de sal, açúcar e gordura que causam maior sensação de sabor nas pessoas.

Esta ferramenta pode ser utilizada sempre que precisamos identificar quais fatores mais afetam as respostas, quais fatores agem em conjunto em determinada resposta, e também para determinar as condições ótimas dos fatores.

Como exemplo, vamos utilizar um caso onde queremos determinar a influência das telas inferior e superior, e dos feltros inferior e superior da 1ª prensa, na lisura do papel.

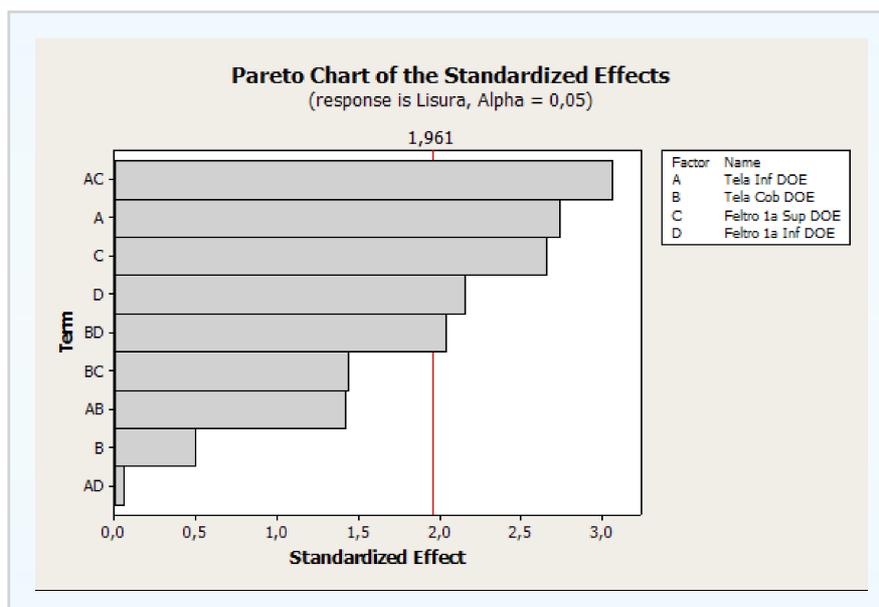


Gráfico 12: Pareto para vestimentas.

O gráfico 12 mostra um diagrama de Pareto, onde são listadas as combinações de vestimentas por influência na lisura do papel. Analisando o gráfico, percebemos que as combinações BC, AB, B e AD não têm influência. Baseado nos dados, isso equivale a dizer que, para a

lisura, a tela da cobertura pouco modifica os valores quando trocada por uma de outro estilo. Já a tela inferior aparentemente apresenta grande influência no resultado. Aqui sabemos somente que algumas combinações importam, e outras não, mas não conseguimos afirmar qual o comportamento da variável: se ela melhora ou piora, de acordo com as combinações. Para isso, a análise fornece outros gráficos interessantes, como o seguinte – um gráfico de interação.

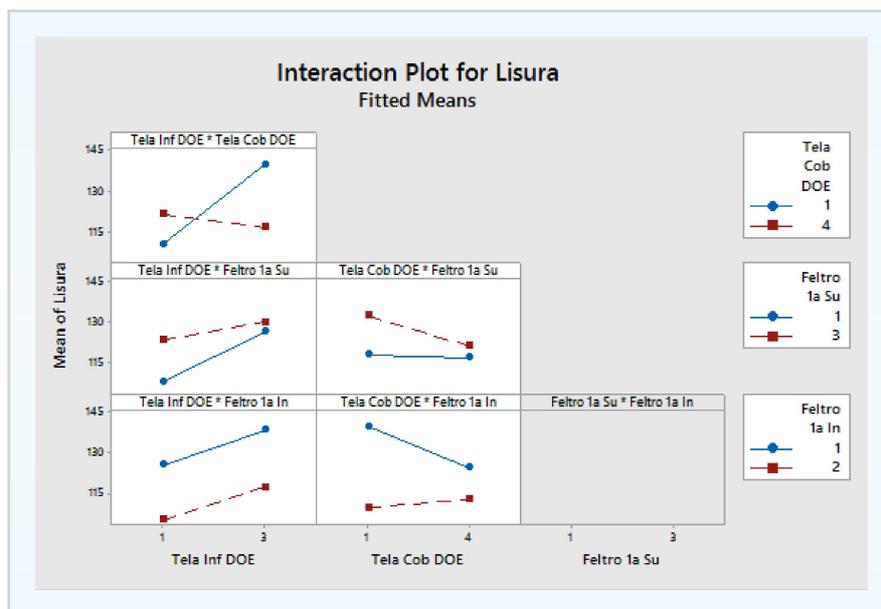


Gráfico 13: Gráfico de interação para lisura.

Com o gráfico de interação, conseguimos visualizar quais combinações apresentam resultados maiores, e quais apresentam resultados menores para a lisura da folha. A tela inferior 3 com o feltro da 1ª superior 3 apresentaram valores maiores de lisura.

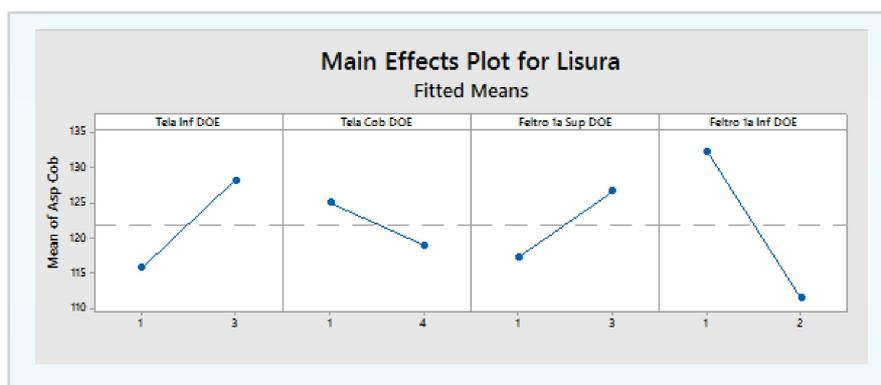


Gráfico 14: Efeitos principais para lisura por vestimenta.

No gráfico 14 são apresentados os efeitos principais de cada vestimenta na lisura da folha. Por exemplo, a tela inferior 1 apresentou resultados de lisura menores do que a 3.

Também é possível fazer previsões de comportamento das repostas e encontrar sugestão de combinações de vestimentas que atinjam determinado critério. No gráfico 15 foi inserido uma meta para lisura de 130, e o software Minitab calculou quais combinações de telas e feltros são mais interessantes de serem usadas para atingir a meta preestabelecida.



Gráfico 15: Otimização das vestimentas de acordo com a meta.

### Correlações e falsos positivos

A estatística e suas ferramentas nos fornecem várias opções de análise que são bastante robustas e versáteis. Temos sempre que tomar o cuidado para não distorcermos os dados. Alguns profissionais do meio dizem, em tom de brincadeira, que uma boa análise estatística é aquela em que os dados são torturados exaustivamente até que o resultado obtido seja aquele que o analista espera.

Este tipo de conclusão pode ocorrer facilmente nas análises de correlação, olhe este caso: minha avó diz com segurança que vai chover quando o dedinho do seu pé direito dói. Se fizermos uma tabela, inserindo as vezes em que ela afirma que seu dedinho está doendo, com as vezes em que efetivamente chove, teremos no final uma correlação bastante forte. Por isso, posso afirmar com certeza que a dor no dedo da minha avó causa chuva em quase 90% das vezes. Brincadeiras à parte, é este tipo de cuidado que temos que ter quando lidamos com correlações. Se faz necessário separar o que realmente tem uma relação verdadeira, do acaso ou da sorte, e também das relações de causa e efeito, como por exemplo a correlação da carga da calandra com a espessura da folha. Além disso, é preciso minimizar sempre o viés, aquele desejo incontrollável que todos temos de estarmos certos.

Os falsos positivos são o fantasma de qualquer análise, já que eles podem nos levar a tomar decisões equivocadas baseadas numa sensação falsa de certeza. Antes de mais nada, é preciso caracterizá-los. Usando o exemplo de tela formadora, sabemos que telas mais finas, ou seja, de menor espessura, apresentam alta probabilidade de ganhar teor seco no final da mesa. Os resultados para um teste com uma tela deste tipo podem ser: tela entra/teor seco sobe, tela entra/teor seco diminui, tela entra/teor seco se mantém igual, tela não entra/teor seco sobe. Se tudo der certo, o resultado esperado é o primeiro, no entanto vamos supor que a bomba de vácuo das caixas de sucção está com problemas e que o nível de vácuo caiu pela metade. O falso positivo aqui será a tela entra/teor seco cai ou se mantém igual, e assim perdemos a oportunidade de usar um produto que faz a diferença.

### Conclusão

Como vimos, existem inúmeras maneiras de se avaliar o desempenho das vestimentas nas máquinas de papel e celulose. Gostaria de reforçar os cuidados que temos que ter nas análises:

- Confie nos seus dados: se forem apontados manualmente, podem conter erros de lançamento; se forem direto do PI, assegure-se que o valor lido faz sentido;
- Retire os pontos extremos. Se não for possível, troque a média pela mediana, por exem-

plo. A segunda tem menos chance de ser influenciada por pontos extremos. Antes de removê-los, tente entender sua significância: um valor de velocidade da máquina igual a zero pode ser tirado caso esteja avaliando produção, mas não se precisar estudar a quantidade de rompimentos de folha;

- Converse exaustivamente com especialistas para que não analise relações de causa e efeito, perdendo tempo e chegando a falsas conclusões;
- Fique atento aos paradigmas, por exemplo: se o teor seco melhora na saída da prensa-gem, o passe diminui, e se muitas vezes isso não pôde ser comprovado, então investigue e se possível teste suas hipóteses antes de chegar a conclusões equivocadas.

Praticamente todos os engenheiros de serviço e vendedores da Albany International têm conhecimento e familiaridade com as análises apresentadas neste artigo. Fique à vontade para utilizar este conhecimento a seu favor no momento de otimizar suas máquinas.

### Bibliografia

- Statistics done wrong – Reinhart, Alex;
- O Sinal e o Ruído – Silver, Nate;
- O Andar do Bêbado – Mlodinow, Leonard;
- BayesTheorem – Morris, Dan;
- Practical Design of Experiments – Hardwick, Colin;
- Sal, Açúcar e Gordura – Moss, Michael;
- Em busca do Infinito – Stewart, Ian;
- Predictive Analytics – Siegel, Eric.

### Perfil do autor:

Volni Nunes de Moraes Junior é formado em Engenharia Química pela Universidade Federal de Santa Catarina UFSC, com MBA em Marketing pela FGV, Curso de Black BeltLeanSix Sigma pela M. I. Domenech Consultores e MBA em Inteligência de Mercado pelo IBRAMERC.

Iniciou suas atividades profissionais na Santa Maria Companhia de Papel e Celulose em 2000, e na Albany em 2002, como Engenheiro de Serviços de Formação. Atualmente exerce a função de Engenheiro de Serviços na linha de Prensagem, realiza auditorias em sistemas de vácuo para máquinas de papel e celulose e análise estatística de processos.