



Utilizando a metodologia

ARITMIA

na elaboração da previsão de vendas

ALEXANDRE GRACIOSO

Previsão de demanda é um assunto delicado dentro das empresas. Na minha experiência como consultor em empresas brasileiras e multinacionais, a prática mais comum para a elaboração da previsão de vendas é atualizar o volume vendido no ano anterior por um fator de crescimento convenientemente escolhido.

Infelizmente, o uso de ferramentas tornais de previsão é muito pouco adotado por empresas no Brasil. A opinião geral é que no Brasil é impossível fazer previsão. O ambiente econômico é muito instável, a inflação é, ou era, muito alta, um dia o governo abaixa impostos e no outro levanta e por aí vai. Embora esse clima de instabilidade seja inegável, atualmente existem métodos estatísticos que navegam com firmeza nesse mar de novidades que é a vida das empresas por aqui.

Além da descrença generalizada sobre a real validade de métodos estatísticos, as empresas brasileiras enfrentam uma segunda limitação na utilização dessas ferramentas: a falta de dados. A utilização de modelos estatísticos exige a disponibilidade de séries históricas e a maior parte das empresas simplesmente desconhece o seu passado. Às vezes é possível conseguir, a muito custo, o histórico para os últimos cinco anos de vendas; na maior parte dos casos, nem isso.

Portanto, empresas que queiram se utilizar dos métodos aqui descritos terão de suplantar várias barreiras. Em primeiro lugar, a empresa precisa de um profissional apto a utilizar tais modelos. Também são necessárias a criação e a colocação à disposição de bancos de dados que contenham o histórico de vendas dos produtos que a empresa vende. Finalmente, existe o aspecto cultural. É necessária uma mudança fundamental da cultura organizacional no sentido de se confiar mais nos métodos formais e menos no "achômetro".

Este artigo demonstra a utilização da metodologia Box-Jenkins, ou ARIMA, na previsão de demanda. Ao passo que uma discussão formal dos modelos matemáticos envolvidos é forçosamente objeto de um curso sobre previsão de demanda, o nosso objetivo é demonstrar a utilização dos modelos por meio de um caso prático: o consumo de bebidas alcoólicas nos EUA.

A segunda parte deste artigo detalha a série de dados utilizada. A terceira parte apresenta o modelo de previsão utilizado e compara os resultados modelos

bastante utilizados na prática. A seção de número quatro discute as limitações dos métodos utilizados e a de número cinco oferece sugestões para a construção dos bancos de dados necessários para a elaboração de modelos de previsão. A última parte resume os principais resultados e conclusões.

O consumo de bebidas alcoólicas nos EUA

Neste artigo, estudamos o consumo de bebidas alcoólicas nos EUA. A série utilizada fornece o consumo mensal nacional, em milhares de hectolitros, para todas as bebidas alcoólicas, de janeiro de 1968 a dezembro de 1994;

Essa série foi escolhida por apresentar um elevado grau de variação e sazonalidade, como pode ser visto na Figura 1. Embora seja verdade que o consumo absoluto de vendas tenha crescido durante quase todo o período em estudo, o consumo per capita diminuiu bastante e, em anos mais recentes, o consumo total apresentou declínio. Além disso, o mês de dezembro responde por um percentual elevado das vendas anuais de bebidas alcoólicas, se fazendo necessário um tratamento apropriado da sazonalidade. Na Figura 1, a sazonalidade pode ser identificada pelos picos que ocorrem todo final de ano.

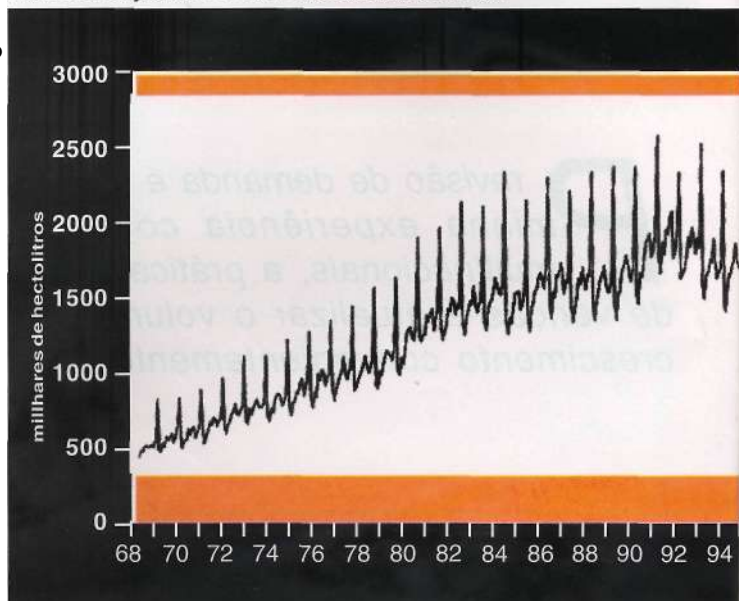
Consideremos por um momento todas as incertezas que existem ao se tentar prever essa série. Em primeiro lugar o comportamento dos consumidores muda ao longo do tempo. Além disso, o consumo é afetado pelos impostos incidentes sobre a ven-

da de bebidas alcoólicas. Um outro aspecto diz respeito ao consumo de bebidas importadas. Nesta série em particular não estamos preocupados com o impacto das importações sobre o consumo, mas caso estivéssemos trabalhando com o consumo de bebidas nacionais, somente esse aspecto deveria ser levado em conta. Em suma, existe uma enorme gama de variáveis que podem impactar o consumo de bebidas. Quais variáveis devemos incluir no nosso modelo e quais podemos desprezar? Como devemos tratar as variáveis que serão incluídas? Estas questões nos levam ao nosso próximo tópico, o modelo propriamente dito.

Prevendo a demanda por bebidas alcoólicas nos EUA

A metodologia ARIMA (Auto-regressivo-Integrado-Média-Móvel), também conhecida como Box-Jenkins, foi popularizada por dois professores da Universidade de Stanford, nos EUA. Box e Jenkins publicaram os primeiros estudos sobre a utilização desses métodos em meados da década de 70 e, a partir de então, eles vêm sendo cada vez mais utilizados.

Figura 1: Evolução do volume mensal de vendas de bebidas alcoólicas nos EUA de janeiro de 1964 a dezembro de 1995.



A metodologia ARIMA utiliza o passado da própria série que se quer prever a fim de prever o futuro. Essa é uma grande diferença em relação aos modelos de regressão e modelos econométricos que utilizam variáveis externas como o PIB, ou PIB per capita a fim de prever o futuro. A metodologia ARIMA está fundamentada na representação de Wald, segundo a qual qualquer processo pode ser representado por modelos matemáticos que contenham um número infinito de termos. Esta ideia é, obviamente, impossível de ser colocada em prática, mas levou à constatação de que um pequeno número de termos fornece aproximações muito boas do modelo com um número infinito de termos. Aliás, como regra geral, parcimônia é uma virtude em se tratando do número de termos em modelos de previsão.

Feitas as devidas apresentações, vamos ao que interessa. O modelo de previsão foi desenvolvido com base nos dados de janeiro de 1968 a dezembro de 1993. Os 24 meses compreendidos entre janeiro de 1994 a dezembro de 1995 foram utilizados para testar o desempenho do modelo.

Ao invés de trabalhar com o volume de vendas propriamente dito, optamos por utilizar o logaritmo natural do volume de vendas. A principal razão para tal, neste caso, é que o logaritmo minimiza problemas de escala que podem ocorrer ao se colocar no mesmo gráfico números de ordens de grandeza muito diferentes.

O modelo é composto por seis termos, cinco termos ARIMA e uma constante. Os cinco termos ARIMA captam a influência do passado da série enquanto a constante estima o consumo de bebidas alcoólicas no momento zero. Conforme já foi dito no início deste artigo, uma descrição mais aprofundada dos modelos envolvidos é objeto de um livro ou um curso na área. Portanto, vamos nos abster de explicar o processo pelo qual chegamos ao modelo aqui exposto, nos atendo somente aos resultados obtidos.

*A opinião
geral é que no
Brasil é
impossível fazer
previsão.*

A Tabela 1 contém o modelo estimado para a série histórica de vendas de bebidas. O modelo contém três termos ARIMA auto-regressivos, um termo auto-regressivo sazonal e um termo média-móvel. A representação do modelo é a seguinte: $Vendas(t) = c + AR(1)*Vendas(t-1) + AR(2)*Vendas(t-2) + AR(3)*Vendas(t-3) + SAR(12)*Vendas(t-12) + MA(12)*Vendas(t-12)$, onde t indica o período atual, c é uma constante, $t-1$, $t-2$, $t-3$ e $t-12$ indicam vendas ocorridas nos períodos anteriores, $AR(1)$, $AR(2)$ e $AR(3)$ são os coeficientes dos termos auto-regressivos normais, $SAR(12)$ é o coeficiente do termo auto-regressivo sa-

zonal e $MA(12)$ é o coeficiente do termo média-móvel.

Os três termos auto-regressivos captam os efeitos dos três meses anteriores ao mês atual. O termo $SAR(12)$ capta o efeito sazonal do modelo. Finalmente, o termo $MA(12)$ reforça o efeito do mesmo mês no ano anterior sobre o mês atual. Com exceção da constante, todos os termos são estatisticamente significativos, conforme indicado pelas estatísticas t maiores do que 2. O fato de que a constante não seja estatisticamente diferente de zero não deve causar desconforto, pois ela não possui interpretação econômica de interesse.

A Tabela 2 informa a performance do modelo na explicação do passado. Não há dúvidas de que esse modelo explica o passado muito bem. O R-quadrado de 0,99 indica que os termos escolhidos explicam 99% da variação no consumo de bebidas. A estatística-F testa a hipótese de todos os coeficientes serem zero ao mesmo tempo. Em outras palavras, ela testa a hipótese do modelo inteiro estar furado. Um valor elevado como o obtido é sinal claro de que o modelo representa adequadamente a realidade.

O próximo passo é utilizar o modelo para realizar previsões. É fato que uma boa descrição do passado não con-

duz automaticamente a boas previsões. Para avaliar a performance do modelo escolhido, devemos comparar as previsões com valores reais. É para isso que separamos as observações de janeiro de 1994 a dezembro de 1995.

Tabela 1:
Modelo ARIMA para o volume mensal de vendas de bebidas alcoólicas nos EUA

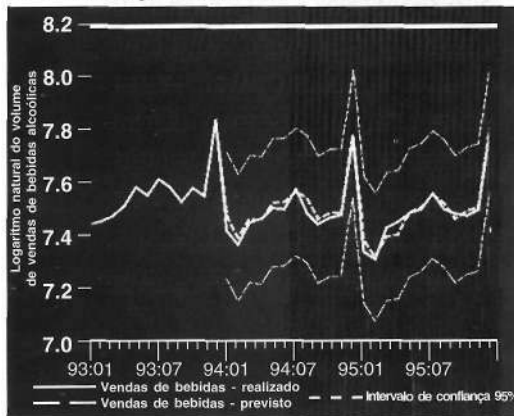
Variável	Coefficiente	Erro-padrão	Estatística t	Prob.
C	21,92009	16,21425	1,351903	0,1775
AR(1)	0,283766	0,045668	6,213721	0,0000
AR(2)	0,244570	0,046675	5,239808	0,0000
AR(3)	0,446727	0,046129	9,684396	0,0000
SAR(12)	0,998050	0,001708	584,2309	0,0000
MA(12)	-0,885800	0,000138	-6440,241	0,0000

Tabela 2: Medidas de avaliação do modelo

R-quadrado	0,994263
Estatística F	10086,34

A Figura 2 fornece as previsões do modelo para o mês seguinte ao atual. Esse tipo de previsão é chamado de previsão de curto prazo, pois somente prevê um período à frente. A linha contínua é o realizado e a linha tracejada em negrito a previsão de curto prazo do modelo. As duas linhas tracejadas acima e abaixo representam o intervalo de confiança de 95% para a previsão.

Figura 2: Previsão de curto prazo versus realizado de janeiro de 1994 a dezembro de 1995

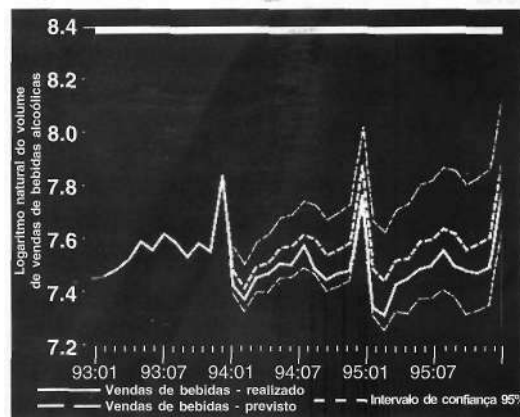


Fica evidente que a previsão é muito boa. Algumas vezes erramos para mais, outras para menos, mas de forma geral estivemos muito próximos ao realizado durante todo o período.

Muitas vezes estamos interessados no longo prazo e desejamos prever vários períodos à frente. Isso também é possível e a Figura 3 demonstra os resultados. Algumas diferenças em relação à previsão de curto prazo merecem ser comentadas. Em primeiro lugar, a previsão é menos precisa. Isso faz sentido, pois na previsão de curto prazo o modelo é atualizado com o realizado de cada mês e utiliza a nova informação para prever o mês seguinte. Na previsão de longo prazo, por outro lado, o modelo não utiliza informação nova. Imagine que, no dia 1º de janeiro de 1994, você tenha previsto o consumo mensal de bebidas alcoólicas para os próximos 24 meses. Essa é a previsão de longo prazo.

Além disso, o intervalo de confiança da previsão fica maior conforme nos afastamos do último mês realizado. Esse

Figura 3: Previsão de longo prazo versus realizado de janeiro de 1994 a dezembro de 1995



fenômeno é decorrente das fórmulas utilizadas na estimação dos parâmetros e reflete a incerteza associada a períodos mais distantes no futuro.

Também fica evidente um certo viés na previsão, que está consistentemente acima do realizado. Esse viés pode ser explicado pela diminuição de consumo em níveis absolutos que não era conhecida em dezembro de 1993. Em defesa dessa previsão de longo prazo, vale lembrar que o realizado está consistentemente dentro do intervalo de confiança construído para a previsão.

Nosso próximo passo é comparar os resultados obtidos a partir do modelo ARIMA com dois modelos muito utilizados na prática. O primeiro modelo ajusta

A metodologia ARIMA (Auto-regressivo-Integrado-Média-Móvel), também conhecida como Box-Jenkins, foi popularizada por dois professores da Universidade de Stanford, nos EUA.

o realizado no período anterior e o segundo modelo é o da regressão linear utilizando o tempo como variável independente.

Vamos começar pelo ajuste dos dados do período anterior. Existe algum fundamento para esse procedimento? Na verdade sim, esse procedimento é o equivalente ao modelo AR(1), ou auto-regressivo de ordem 1. Matematicamente, a representação deste modelo é a seguinte: $Vendas(t) = c + AR(1) * Vendas(t-1)$, onde t representa o período atual, $t-1$ o período imediatamente anterior, c é uma

constante e AR(1) é o coeficiente de ajuste das vendas no período anterior.

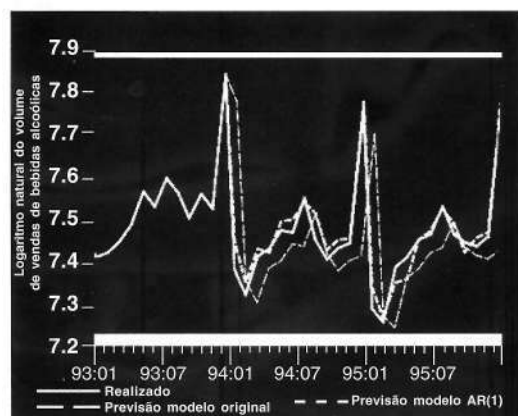
Porém, nem sempre o processo real segue esse modelo. Além disso, mesmo que o processo siga um modelo auto-regressivo de ordem 1, somente por acidente o analista chegaria ao coeficiente correto do ponto de vista matemático.

E quanto à regressão linear com o tempo como variável independente? Existe algum fundamento estatístico para esse modelo? Novamente, a resposta é sim. Ao supor que o processo siga esse modelo, o analista está assumindo a premissa que a cada período a mudança na variável dependente é constante. Em outras palavras, no período $t=1$ o volume de vendas foi 10 unidades, no período $t=2$ o volume de vendas foi 20 unidades, em $t=3$ foi 30 unidades etc. A representação deste modelo é a seguinte: $Vendas(t) = c + b1 * t$, onde t representa o período atual, c é uma constante e $b1$ é o incremento nas vendas que ocorre a cada período.

A desvantagem deste procedimento é que muitas vezes o processo se comporta de acordo com modelos estocásticos. Nesses casos, a modelagem por meio de modelos determinísticos conduzirá o analista inevitavelmente a previsões errôneas.

A Figura 4 compara o realizado para o período de janeiro de 1994 a dezembro de 1995 com a previsão de curto prazo

Figura 4: Previsão de curto prazo do modelo original e previsão de curto prazo do modelo AR(1) versus realizado de janeiro de 1994 a dezembro de 1995

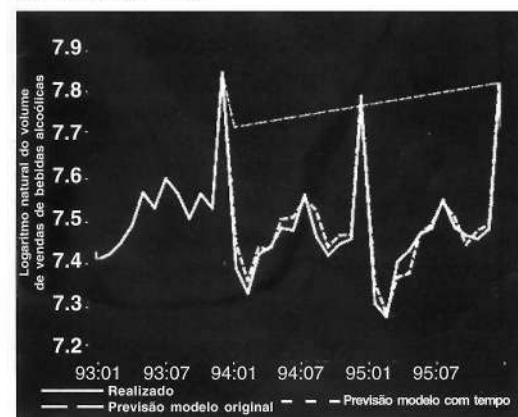


do modelo original e a previsão de curto prazo do modelo que ajusta o volume de vendas ocorrido no período anterior.

A linha contínua representa o realizado mensal. A linha tracejada em negrito representa a previsão de curto prazo do modelo original e a linha tracejada normal representa a previsão de curto prazo do modelo AR(1). É evidente, a partir da figura, que o modelo original gera melhores previsões que o modelo AR(1), que são menos precisas e apresentam grande viés: ora estão sempre acima do realizado e ora sempre abaixo do realizado.

A Figura 5 compara o realizado com a previsão do modelo original e a previsão do modelo que considera o tempo como

Figura 5: Previsão de curto prazo do modelo original e previsão de curto prazo do modelo com tempo versus realizado de janeiro de 1994 a dezembro de 1995



variável independente. Mais uma vez o modelo original se mostra mais satisfatório. O outro modelo é claramente inadequado para a modelagem deste processo.

Limitações do modelo exposto

Infelizmente, nem tudo são flores e a metodologia ARIMA apresenta limitações que o analista precisa ter em mente. Em primeiro lugar, os modelos ARIMA não permitem simulações. Em outras palavras, não é possível utilizar um modelo

ARIMA para responder perguntas do tipo "o que aconteceria se eu aumentasse o meu preço em 10%?". Da mesma forma, o modelo ARIMA aqui exposto não poderia estimar o efeito da mudança do comportamento do consumidor na demanda por bebidas alcoólicas nos EUA.

Uma importante consequência do que foi dito acima diz respeito aos picos e vales. Modelos ARIMA não prevêem em que momento uma série ascendente irá começar a cair ou em que momento uma série descendente irá começar a subir. Modelos ARIMA extrapolam o passado da série alicerçados na ideia de que qualquer processo real pode ser modelado com alguns poucos termos que se aproximem da representação de Wald.

Essas limitações não invalidam a metodologia ARIMA, no entanto. Embora os modelos ARIMA sejam incapazes de prever o momento exato da mudança, eles "aprendem" com a mudança e as previsões seguintes irão levar as novas circunstâncias em consideração. Em outras palavras, embora o modelo não saiba prever que o comportamento do consumidor irá mudar, a partir do momento em que isso ocorre, essas novas circunstâncias serão consideradas.

O leitor pode estar se perguntando se não seria mais interessante desenvolver um modelo que con-

sidere explicitamente o comportamento do consumidor como variável independente a fim de se prever os efeitos de mudanças no comportamento. Em algumas situações isso é verdade, mas inevitavelmente a empresa irá enfrentar a dificuldade de quantificar o comportamento do consumidor corretamente e medir esse comportamento ao longo do tempo. O mesmo é verdade para qualquer outra variável independente que se utilize. Uma das grandes vantagens de se utilizar a metodologia ARIMA é que na maioria dos casos não é necessário utilizar mais variáveis independentes, bastando o passado da própria variável que se quer prever.

Uma das grandes vantagens de se utilizar a metodologia ARIMA é que na maioria dos casos não é necessário utilizar mais variáveis independentes, bastando o passado da própria variável que se quer prever.

Informações necessárias para previsões

A fim de se implantar os modelos aqui discutidos, a empresa necessita, obrigatoriamente, de séries históricas longas e confiáveis. Esse pré-requisito é absolutamente indispensável. No entanto, minha experiência como consultor de firmas nacionais e multinacionais de vários tamanhos mostra que, de forma geral, as empresas não dispõem das séries históricas necessárias para a elaboração de previsões. Na maioria das vezes a empresa possui um

Vamos começar pelo ajuste dos dados do período anterior... esse procedimento é o equivalente ao modelo AR(1), ou auto-regressivo de ordem 1.

pequeno histórico de alguns anos, quando muito de cinco anos. Além disso, existe a questão de quais dados devem ser guardados. Todos os dias a empresa gera uma quantidade enorme de informações: quais são úteis para previsão?

De forma geral, é mais indicado manter os dados no nível de detalhe mais baixo possível. Dessa forma, se a empresa possui dez linhas de produtos e cada linha possui dez produtos diferentes, o banco de dados deve manter as informações para cada um dos cem produtos que a empresa comercializa. Somente dessa forma poderão ser geradas previsões por produto, ao passo que previsões por linha de produtos também são possíveis, bastando somente a consolidação dos números já existentes no banco de dados.

Uma palavra de cautela sobre bases de dados em Cobol. As bases em Cobol apagam o passado automaticamente quando são atualizadas e qualquer tipo de pesquisa exige grande esforço por parte do pessoal de informática. Se esse for o caso da sua empresa, monte uma base de dados em separado para fins de previsão. Quem sabe um pequeno banco de dados em Access seja o suficiente.

Uma outra questão é a natureza dos dados que a empresa deve guardar. Somente quantidade é o suficiente, ou será necessário guardar também os dados de

preço, investimento em propaganda etc? Os modelos mostrados aqui utilizam somente quantidade para prever o futuro e em muitos casos isso será o suficiente. Modelos mais complexos, que consideram várias variáveis, podem ser montados desde que os dados estejam disponíveis.

Por um lado, é melhor ter mais dados disponíveis do que menos, porém existem custos associados à manutenção dessas bases que devem ser considerados. Em vista disso, eu começaria guardando somente quantidade vendida e preço de venda, informações que devem ser introduzidas no sistema de faturamento a cada venda de qualquer maneira.

Conclusões

Previsão de demanda é uma atividade essencial de qualquer empresa. O dimensionamento de toda empresa depende dessa previsão. Apesar disso, muitas empresas ainda se contentam com previsões elaboradas com métodos, quando muito, aproximados.

Este artigo demonstrou uma metodologia disponível a qualquer empresa que se dê ao trabalho de montar um banco de dados e contratar um profissional capacitado: ARIMA.

Essa metodologia é capaz de produzir previsões bastante próximas ao que vai ocorrer no futuro utilizando somente

Embora os modelos ARIMA sejam incapazes de prever o momento exato da mudança, eles "aprendem" com a mudança e as previsões seguintes irão levar as novas circunstâncias em consideração.

E quanto à regressão linear com o tempo como variável independente? Ao supor que o processo siga esse modelo, o analista está assumindo a premissa que a cada período a mudança na variável dependente é constante.

o passado da própria série, o que a torna uma alternativa interessante em relação aos modelos econométricos habituais em que várias variáveis são necessárias.

Os modelos ARIMA possuem duas limitações que devem ser levadas em consideração pelo analista. Em primeiro lugar, não é possível simular cenários. Por exemplo, não é possível simular o que aconteceria nas vendas da empresa se o preço aumentasse em 10%. Além disso, os modelos ARIMA não prevêem com absoluta precisão o momento no qual uma série ascendente irá se tornar descendente e vice-versa. Para esse tipo de simulação são necessários outros tipos de modelos.

No entanto, essas limitações não invalidam a metodologia ARIMA. A forma como os modelos são montados garante que os mesmos "aprendam" com o passado. Por exemplo, digamos que estamos interessados em prever importações de camisetas. É claro que a taxa de câmbio afeta a quantidade de camisetas importadas. Se houver uma mudança brusca na taxa, em um primeiro momento as previsões não serão confiáveis, mas, assim que atualizarmos o modelo com os novos dados, o modelo irá se recalibrar.

Alexandre Gracioso - professor da ESPM