



Bootcamp Data Analytics 2022

Challenge Probabilidade e Amostragem



Desafio 1: Entendendo o consumidor.

Dificuldade: Média.

Suponha que você trabalha em uma empresa de produtos de beleza populares e, com a pandemia, você suspeita que o perfil de seus clientes mudou. O time de marketing, pediu a sua ajuda e gostaria de entender se 1. isso de fato aconteceu e 2. quais as diferenças do novo perfil de clientes para o antigo, pois assim poderiam personalizar melhor os seus anúncios. Para isso, eles te enviaram a base de dados ([desafio_01_consumidor.csv](#)). Como você poderia ajudá-los?

Para responder esse desafio:

A base de dados: [desafio_01_consumidor.csv](#) contém informações de seus clientes (anonimizados) no período pré e pós pandemia. A coluna "pré-pandemia" indica se o dado é pré-pandemia ou não, sendo 1, caso positivo ou 0 caso negativo.

Dica / Passo a Passo:

1. Entenda a distribuição de variáveis como de idade / renda / região / produtos mais comprados antes e depois da pandemia
 - a. Identifique as distribuições
 - b. Plote histogramas / boxplots
2. Você notou diferenças nítidas pré e pós pandemia? Como você poderia provar estatisticamente que essas diferenças são significativas ? (pergunta extra)
3. Suponha agora que os dados recebidos são somente uma amostra da sua população total de clientes. Suponha que a sua "população" de clientes verdadeira é de 100000. Qual o tamanho da amostra necessária para chegarmos a conclusões com 95% de confiança e 2% de margem de erro? A base informada é suficiente? E com uma margem de erro de 10%?

Desafio 2: Determinando o Estoque de Segurança.

Dificuldade: Alta.

Suponha agora que você trabalha em uma empresa de produtos perecíveis, como Danone ou Nestlé. Para a empresa, é muito importante que a sua política de estoque seja eficiente. É necessário manter um estoque de segurança em todos os produtos (acima da previsão de vendas), pois, caso contrário, as pessoas podem não encontrar os produtos no supermercado. Mas, ao mesmo tempo, se o estoque estiver em excesso, isso pode representar um grande custo. Os produtos podem não ser comprados, e conseqüentemente, jogados fora, o que na indústria se chama "desova".

Para achar o equilíbrio perfeito, o gerente de vendas pediu a ajuda a você, do time de análise de dados, e ele gostaria de saber qual o estoque de segurança ideal.

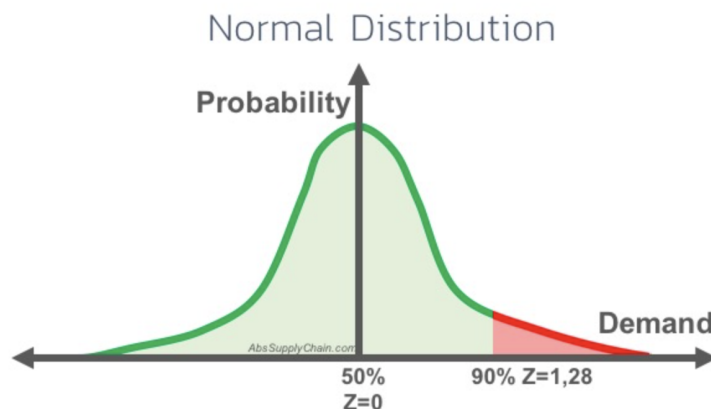
Para responder esse desafio, antes vamos abordar alguns conceitos importantes na indústria:

- **Estoque de segurança:** é uma quantidade de estoques que se tem, mas que não se deseja usar. É como um seguro do carro: é bom ter um, mas você não quer precisar usá-lo. Isto porque ele é calculado apenas para suprir uma segurança em casos de variações inesperadas, algo que você não pode prever ou controlar: problemas no fornecedor, atrasos na entrega, uma demanda que não foi prevista...
- **Demanda:** se a demanda é bem estável e conhecida com antecedência, então temos pouca variabilidade a cada mês e não precisamos nos proteger muito contra essas variações (pois sabemos que elas não ocorrem); por outro lado, se seu produto tem uma variabilidade nas vendas muito grande, então precisaremos de estoque de segurança maior. Isto tudo é medido matematicamente através do desvio padrão da demanda, que neste caso é calculado como o desvio padrão da previsão da demanda. Um bom sistema de previsões é capaz de oferecer este número, ou ele pode ser estimado de maneiras mais simples, mas menos precisas.
- **Lead time (tempo de entrega) do produto:** se o tempo de entrega é elevado e sua variabilidade é alta (se uma entrega é feita em 5 dias, outra em 8 dias, outra em 2 dias), então é preciso ter uma segurança frente à este tempo média de entrega de 5 dias, pois algumas vezes ela chega a demorar 8 dias. Mas não queremos nos planejar sempre para receber apenas depois de 8

dias, pois isto acarretaria custos muito altos, então o estoque de segurança utiliza a estatística para auxiliar nessa tarefa.

- **Nível de serviço desejado:** nem todos os produtos merecem a mesma atenção e o mesmo cuidado; alguns produtos são críticos, mais importantes ou mais atrativos, e por isso merecem estar **sempre** presentes, enquanto em outros produtos podemos nos dar ao luxo de não tê-lo em estoque sempre. Matematicamente, isto é modelado através do nível de serviço desejado: quanto maior o nível de serviço (um número percentual de 0 a 100), maior será o estoque de segurança pois queremos mais garantias que o produto estará **sempre** disponível. O nível de serviço depende de cada setor: palitos de fósforo num supermercado não devem ter nível de serviço muito alto, enquanto antibióticos numa farmácia hospitalar devem ter nível de serviço altíssimo. O nível de serviço indica o quanto queremos estar seguros frente às variabilidades que ocorrem, em outras palavras, frente aos desvios padrões da demanda e do *lead time*.
- **A Fórmula Mãe do Estoque de segurança:**

A maioria das indústrias calcula o estoque de segurança tomando como base a distribuição normal. Assumindo que suas vendas seguem essa distribuição, ela irá determinar qual a probabilidade de se vender certo produto.



Por exemplo, se você vende na média 1000 unidades de produto por semana, a distribuição mostra que você tem uma alta probabilidade de vender em torno de 1000 todas as semanas, e uma baixa probabilidade de esse número ser 20 ou 200000 de produtos vendidos.

Assim, você pode utilizar a distribuição normal, para entender qual o seu nível de serviço desejado. Ou seja, se você deseja um nível de serviço de

90%, a curva normal irá te fornecer o multiplicador desejado (variações da média) para você alcançar 90% da demanda.

Portanto, matematicamente, podemos determinar o estoque de segurança utilizando os conceitos abordados em estatística, como a tabela da distribuição normal, utilizando a fórmula abaixo.

$$\text{Estoque de Segurança} = Z * \sqrt{\sigma_d^2 t + \sigma_t^2 d^2}$$

sendo :

z = o valor da tabela da distribuição normal, que indica a quantos desvios padrões ao redor da média temos que tomar para cobrir a proporção da área sob a curva normal /nível de serviço desejado.

d = demanda média passada ou estimada

t = lead time médio.

σ_d = desvio padrão / variação da demanda

σ_t = desvio padrão / variação do lead time

Assim se o tempo de entrega tiver pouca variabilidade a fórmula pode ser aproximada por :

$$\text{Estoque de Segurança} = Z * \sqrt{\sigma_d^2 t}$$

e a intuição da fórmula pode ser vista como : O quanto de variação da média da demanda eu estou disposto a aceitar além do natural (z-score do nível de serviço) * desvio médio da demanda que já ocorre naturalmente * tempo de entrega.

Matematicamente o valor obtido pela fórmula deve ser **sempre** arredondado para cima, para garantir que estamos cobertos contra variações indesejadas.

Dados esses conceitos vamos ao passo a passo do desafio.

O gerente de vendas te enviou o arquivo [desafio_02_estoque.csv](#) contendo os dados de vendas de iogurte e o lead time de 2015 a 2022.

1. Analise os dados das vendas e do lead time. Eles seguem uma distribuição normal? Que características você identificou para chegar a essa conclusão?
2. Assuma que:
 - a. Os dados seguem uma distribuição normal.
 - b. O time de vendas utilize para o produto em questão o nível de serviço de 99,87%
 - c. O produto não tem sazonalidade (todos os meses vende valores próximos) e a demanda média passada é uma boa estimativa para a projetada, de acordo com o time de planejamento de demanda.

Calcule, com os dados fornecidos, o estoque de segurança desejado.