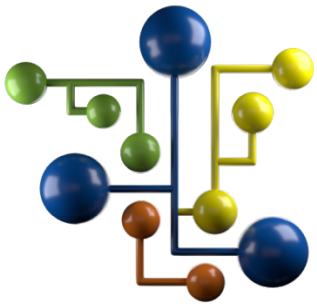


Data Science Academy

www.datascienceacademy.com.br

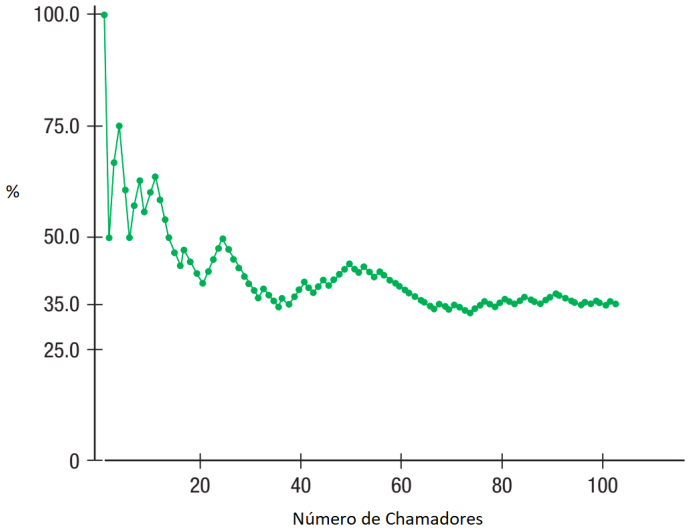
Big Data Real-Time Analytics com Python e Spark

O Fenômeno Aleatório

Quando um cliente liga para o número 0-800 de uma empresa de cartão de crédito, ele é solicitado a fornecer um número de cartão antes de ser conectado a um operador. À medida que a conexão é feita, os registros de compra desse cartão e as informações demográficas do cliente são recuperados e exibidos na tela do operador.

Se a pontuação (score) do cliente for alta o suficiente, o operador pode ser solicitado a "vender" outro serviço - talvez um novo cartão "platinum" para clientes com uma pontuação de crédito de pelo menos 750. Claro, a empresa não sabe quais clientes vão ligar. As chegadas de chamadas são um exemplo de um **fenômeno aleatório**.

Com fenômenos aleatórios, não podemos prever os resultados individuais, mas podemos esperar compreender as características dos seus comportamentos de longo prazo. Não sabemos se o próximo cliente se qualificará para o cartão platinum, mas quando as chamadas entrarem no call center, a empresa descobrirá que a porcentagem de chamadores qualificados para o cartão platinum se estabelecerá em um padrão, como mostrado no gráfico da figura abaixo.



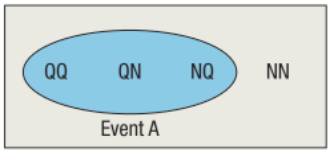
À medida que as chamadas chegam ao call center, a empresa pode registrar se cada chamador se qualifica. O primeiro chamador hoje é qualificado. Em seguida, as cinco qualificações dos próximos chamadores foram não, sim, sim, não e não. Se traçarmos a porcentagem que se qualifica em relação ao número da chamada, o gráfico começaria em 100% porque o primeiro chamador se qualificou (um de um, para 100%). O próximo chamador não se qualificou, então o percentual acumulado caiu para 50% (um em cada dois). O terceiro chamador qualificou-se (dois de três ou 67%) e assim por diante. A cada nova chamada, o novo dado é uma fração menor da experiência acumulada, de modo que, no longo prazo, o gráfico se estabiliza. À medida que se instala, parece que, de fato, a fração de clientes que se qualifica é de cerca de 35%.

Quando se fala em comportamento de longo prazo, ajuda a definir nossos termos. Para qualquer fenômeno aleatório, cada **tentativa ou teste** gera um resultado. Para o call center, cada chamada é uma tentativa. Algo acontece em cada tentativa, e chamamos o que quer que aconteça de **resultado**. Aqui, o resultado é se o chamador se qualifica ou não. Usa-se o termo mais geral para se referir a resultados ou combinações de resultados.

Por exemplo, suponha que categorizemos os chamadores em seis categorias de risco e enumeremos esses resultados de 1 a 6 (aumentando a capacidade de obtenção de crédito). Os três resultados 4, 5 ou 6 podem compor o evento "o chamador é pelo menos uma categoria 4".

Às vezes falamos sobre a coleta de todos os resultados possíveis, um evento especial ao qual nos referimos como o **espaço de amostra**. Denotamos o espaço amostral com a letra S. Mas qualquer que seja o símbolo que usamos, o espaço amostral é o conjunto que contém todos os resultados possíveis. Para as chamadas, se deixarmos Q = qualificado e N = não qualificado, o espaço de amostragem será simples: S = {Q, N}.

Se olharmos para duas chamadas juntas, o espaço amostral tem quatro resultados: S = {QQ, QN, NQ, NN}. Se estivéssemos interessados em pelo menos um chamador qualificado das duas chamadas, estaríamos interessados no evento (chame de A) consistindo nos três resultados QQ, QN e NQ, e escreveríamos A = {QQ, QN, NQ} conforme figura abaixo:



Isso nos leva ao conceito de Probabilidade Empírica, assunto da próxima aula. Até lá!