



**Data Science
Academy**

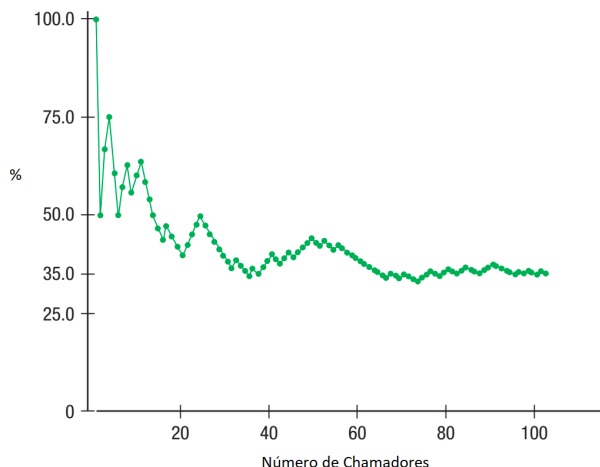
www.datascienceacademy.com.br

**Big Data Real-Time Analytics com
Python e Spark**

**Probabilidade Empírica e a
Lei dos Grandes Números**

Embora possamos não ser capazes de prever um determinado resultado individual, como qual chamada recebida representa uma potencial venda, podemos dizer muito sobre o comportamento de longo prazo.

Olhe mais uma vez para a imagem abaixo. Se você for perguntado sobre a probabilidade de que um chamador aleatório se qualifique, você pode dizer que é de 35%, porque, no longo prazo, a porcentagem dos chamadores que se qualificam é de cerca de 35%.



E é exatamente isso que queremos dizer com probabilidade. Como se baseia na observação repetida do resultado do evento, essa definição de probabilidade é geralmente chamada de **probabilidade empírica**.

Isso realmente simplifica as coisas se as tentativas individuais forem independentes. Grosso modo, independência significa que o resultado de um teste não influencia ou altera o resultado de outro. Lembre-se dos capítulos anteriores em que chamamos duas variáveis independentes se o valor de uma variável categórica não influenciou o valor de outra variável categórica (verificamos a independência comparando as distribuições de frequência relativa entre as variáveis.) Não há razão para pensar que a qualificação de um chamador influencia a qualificação de outro chamador, portanto, esses são testes independentes.

Felizmente, para eventos independentes, podemos depender de um princípio chamado **Lei dos Grandes Números**, que afirma que, se os eventos forem independentes, quando o número de chamadas aumentar, ao longo de dias ou meses ou anos, a longo prazo a frequência relativa de chamadas qualificadas fica cada vez mais próxima de um único valor. Isso nos dá a garantia de que precisamos e torna a probabilidade um conceito útil.



Como a Lei dos Grandes Números garante que frequências relativas se estabelecem a longo prazo, podemos dar um nome ao valor que elas abordam. Nós chamamos isso de a **probabilidade desse evento**. Para o call center, podemos escrever $P(\text{qualificado}) = 0,35$. Como se baseia na observação repetida do resultado do evento, essa definição de probabilidade é frequentemente chamada de **probabilidade empírica**.