



# Data Science Academy

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

## Microsoft Power BI Para Data Science

### Amostragem



Amostragem é a técnica, processo ou a pesquisa que podem ser realizadas para obter uma amostra. A amostragem usa a coleta, organização, apresentação e análise dos dados como meio de estudar os parâmetros de uma população.

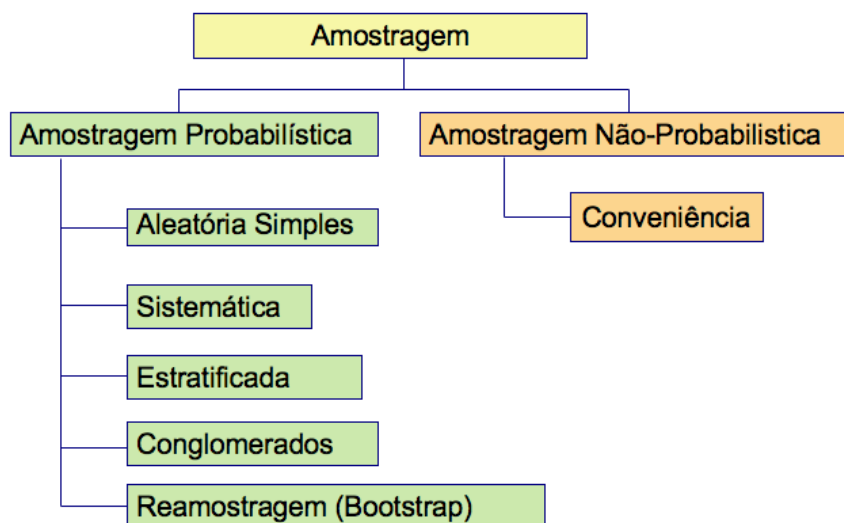
- **Amostragem:** é a técnica que seleciona apenas alguns elementos da população para análise.
- **Censo:** é a técnica que seleciona e avalia todos os elementos da população para análise.

Mas por que não medir uma população inteira, ao invés de medir apenas uma amostra? Dependendo das circunstâncias, medir uma população inteira seria caro demais ou até mesmo inviável. E em muitos casos, mesmo que fosse possível medir a população inteira, seria um desperdício de tempo e dinheiro. Se a amostra for selecionada corretamente e a análise sobre ela for feita seguindo as metodologias estatísticas, esta informação pode ser usada para fazer uma avaliação precisa sobre a população inteira.

Mas existem riscos envolvidos em tomar decisões baseadas em amostragem, pois a amostragem pode ser exposta a erros, que podem levar a decisões incorretas.

Existem basicamente 2 tipos de Amostragem: Probabilística e Não-Probabilística. Amostragem Não-Probabilística é **subjetiva**, pois é influenciada pela pessoa que está conduzindo a pesquisa. Ela se baseia nas decisões pessoais do pesquisador. Amostragem Probabilística é objetiva, pois não é influenciada pela pessoa que está conduzindo a pesquisa ou experimento.

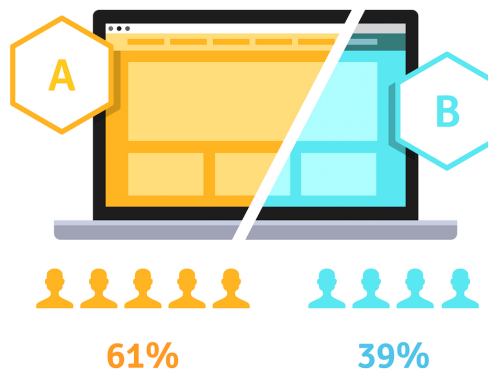
Obviamente, a Amostragem Probabilística leva a resultados mais precisos e consistentes na maioria das vezes. Os elementos da amostra são selecionados aleatoriamente e todos eles possuem probabilidade conhecida de serem escolhidos. Tal seleção ocorre através de uma forma de sorteio não viciado, com seleção aleatória dos dados. Emn resumo, nós temos:



Vamos estudar a Amostragem Probabilística em mais detalhes.

## Amostragem Probabilística

Amostragem probabilística é uma técnica em que gera uma amostra em que todos os membros da população tem chance de pertencer a amostra, o que não significa que cada membro tenha exatamente a mesma chance de ser selecionado.



Com Amostragem Probabilística, nós temos a possibilidade de realizar uma variedade de testes de estatística inferencial, que permitirão extrair conclusões confiáveis sobre a população. Vejamos quais são os tipos de Amostragem Probabilística:





## Amostragem Aleatória Simples

Uma amostra aleatória simples é a amostra em que cada membro de uma população tem igual chance de ser selecionado, ou seja, todos os elementos da população têm a mesma probabilidade de pertencerem à amostra. Essa amostragem pode ser sem reposição, que é quando o elemento que já foi sorteado não continua no sorteio ou com reposição, quando o elemento sorteado continua no sorteio, podendo ser escolhido novamente.

## Amostragem Sistemática

Outra forma de garantir que uma amostra é selecionada randomicamente, é usar Amostragem Sistemática. A Amostragem Sistemática consiste em selecionar as unidades elementares da população em intervalos pré-fixados. Para funcionar, a técnica requer a listagem prévia da população, que deverá ser homogênea e uma atenção especial na periodicidade dos dados. A Amostragem Sistemática é semelhante à aleatória simples, mas a listagem é ORDENADA.

O processo basicamente é:

- Divide-se o tamanho da população (**N**) pelo tamanho da amostra (**n**), obtendo um intervalo de retirada (**k**).
- Sorteia-se o ponto de partida. A cada **k** elementos retira-se um para a amostra (**k** pode ser o segundo, quarto, vigésimo elemento, etc...).

Uma das vantagens de se utilizar esta técnica, é evitar amostra tendenciosa. Como o membro da amostra será selecionado sempre na mesma posição a partir da população, evitamos uma seleção direcionada a membros específicos. Entretanto, deve-se ter cuidado com o conceito de periodicidade, quando um padrão da população coincide com a posição que estamos selecionando.

## Amostragem Estratificada

Nós dividimos nossa população em grupos mutuamente exclusivos (chamados estratos) e randomicamente selecionamos membros de cada grupo para nossa amostra.

## Amostragem Por Conglomerado

Na amostragem por conglomerados, a população é dividida em conglomerados. Selecionam-se aleatoriamente os conglomerados que farão parte da amostra, ao passo que todos os elementos dos conglomerados selecionados serão amostrados. É muito utilizada quando há necessidade de se realizar entrevistas ou observações em grandes áreas. Na amostragem por conglomerados, também criamos grupos mutuamente exclusivos, cada um representativo da população. Ao invés de selecionarmos membros de cada grupo, selecionamos randomicamente grupos inteiros para nossa amostra.

## Reamostragem

**Reamostragem** é uma **técnica estatística** em que várias amostras são repetidamente extraídas da **população**. Um tipo específico de técnica de Reamostragem é conhecido como método Bootstrap, desenvolvido por Bradley Efron, membro do Departamento de Estatística da universidade de Stanford na Califórnia nos EUA. O método Bootstrap consiste em usar software de computador para extrair diversas amostras (com reposição), para estimar determinados parâmetros da população, tais como média e desvio padrão.

