



# Data Science Academy

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

## Microsoft Power BI Para Data Science

### Intervalo de Confiança



Conforme nós já vimos, a inferência estatística tem como objetivo tirar conclusões acerca da população a partir de dados extraídos da amostra. A amostra deve ser representativa da população. Um dos grandes objetivos da inferência estatística corresponde à estimação dos parâmetros da população.

Utilizar amostras tem como pontos positivos: a eficiência, o ganho de tempo e o custo reduzido, além de dados mais confiáveis e facilidade para realizar operações. Temos como um exemplo disso o censo, pois é difícil coletar informações de toda a população, tendo em vista o controle dos entrevistadores, o custo e o prazo para isso. Imagine se em uma pesquisa eleitoral fosse necessário entrevistar toda a população, seria um verdadeiro transtorno e absurdamente caro. Mas, em contrapartida, quanto menor a amostra maior os erros. Por exemplo: se a amostra for pequena demais, não será capaz de gerar os resultados precisos para os parâmetros da população.

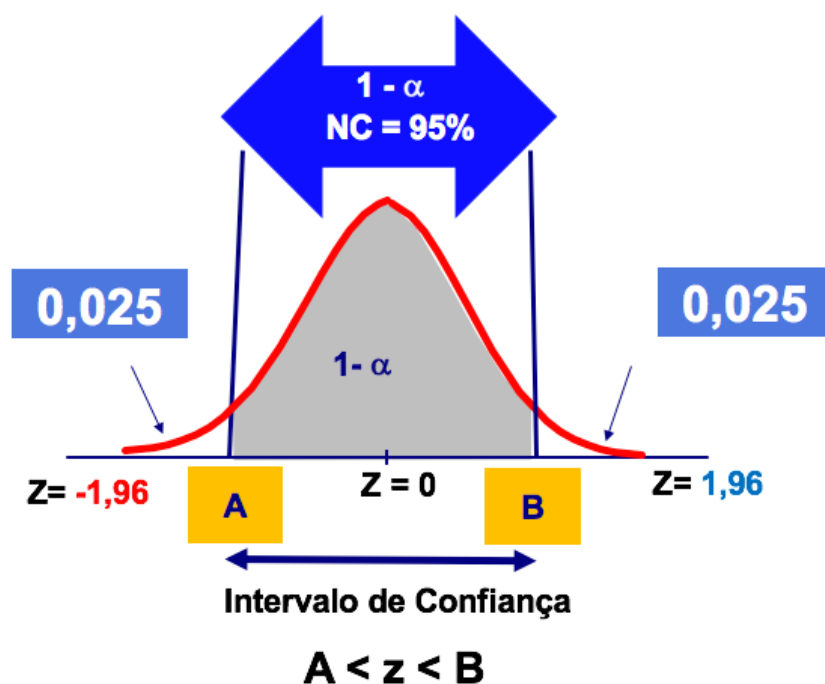
Uma das mais importantes funções da Estatística, é usar informações de amostras, para fazer análises sobre a população, da qual a amostra foi extraída. Usamos amostras para estimar a população. Mas qual é a precisão destas estimativas? A resposta a esta pergunta é muito importante e o conceito de intervalo de confiança nos ajudará a respondê-la. Mas antes precisamos saber o conceito de Estimação de Parâmetros.

A Estimação é o processo de se utilizar dados amostrais para estimar os valores para a média, o desvio padrão de uma população e a proporção populacional dos parâmetros populacionais que são desconhecidos. Os parâmetros podem ser estimados pontualmente, por meio de um único ponto (estimação pontual) ou por meio de um intervalo de valores (estimação por intervalos). Os principais métodos de estimação pontual são: o estimador dos momentos, o método dos mínimos quadrados e o método da máxima verossimilhança. Já os principais métodos de estimação intervalar ou intervalos de confiança (IC) são: IC para a média populacional quando a variância é conhecida, IC para a média populacional quando a variância é desconhecida, IC para a variância populacional e IC para a proporção. Vamos nos concentrar na estimação por intervalo.

Intervalo de confiança é uma amplitude (ou um intervalo) de valores que tem a probabilidade de conter o valor verdadeiro da população. Intervalo de Confiança está associado a um grau de confiança que é uma medida da nossa certeza de que o intervalo contém o parâmetro populacional. Observe que na definição de intervalo de confiança, está associado uma probabilidade. A esta probabilidade chamamos de nível de confiança.

**A estimação por intervalos é utilizada quando se deseja conhecer um intervalo de possíveis valores no qual o parâmetro estimado esteja presente, com um determinado nível de confiança  $(1 - \alpha)$ .**

Um Intervalo de Confiança funciona como um indicador da precisão da sua medida e indica qual o grau de estabilidade da sua estimativa, a qual pode ser calculada para determinar o quanto você está próximo de sua estimativa original quando realiza um ou mais experimentos. O **nível de confiança** é expresso percentualmente e por isso usamos: **100 (1 -  $\alpha$ )**.



$$\text{Probabilidade } P = \{A \leq \mu \leq B\} = 1 - \alpha$$

- O intervalo (A, B) é chamado de intervalo de confiança.
- $\alpha$  (alfa) é o **nível de significância**
- $100 (1 - \alpha)$  é o **nível de confiança**
- $1 - \alpha$  é o **coeficiente de confiança**
- Probabilidade P é a probabilidade da média da população estar no intervalo de confiança.

Normalmente utiliza-se Nível de Confiança de:

**90%**  
 $\alpha = 0,10$

**95%**  
 $\alpha = 0,05$

**99%**  
 $\alpha = 0,01$

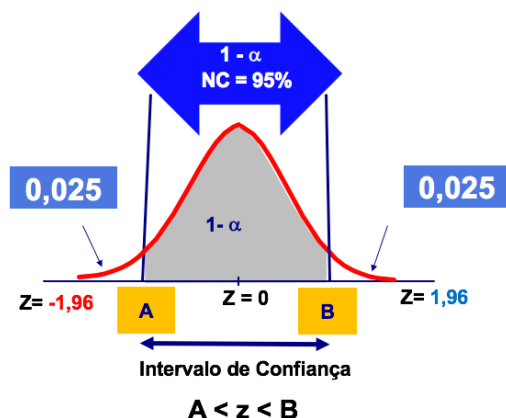
Ou seja, se coletarmos várias amostras e construirmos um intervalo de confiança para cada uma, a longo prazo, **95%** destes intervalos conteriam efetivamente a **média da população**  $\mu$ . Vale ressaltar que podemos utilizar o percentual que desejarmos, 60%, 40%, 87,33%. Não há obrigatoriedade em utilizar somente 90, 95 ou 99%. Elas são apenas escolhas mais comuns.

Um Analista/Cientista de dados irá utilizar os recursos de intervalo de confiança diariamente para:

- Calcular a variabilidade de tempo de atendimento à clientes (bancos, centrais de atendimento, estabelecimentos em geral).
- Tempo de aterrissagem de voo.
- Tempo de substituição de aparelhos eletrônicos em uma central de processamento de dados inteligentes.
- Vida útil de componentes elétricos e eletrônicos em ambientes de trabalho e domésticos.
- Resultados de pesquisas eleitorais ou outras pesquisas em geral.
- Dentre inúmeros outros recursos.

Você como analista de dados irá coletar os dados, extrair as informações desses dados e transformar essas informações em conhecimento, para que elas possam ser usadas para a tomada de decisões.

E qual a função do Escore Z (também chamado valor crítico z)?



O Escore z corresponde ao valor da fronteira da área  $\alpha/2$  nas caudas direita ou esquerda da distribuição normal. O valor crítico é o número na fronteira que separa os valores estatísticos amostrais prováveis de ocorrerem dos valores que tem pouca chance de ocorrer. O **Intervalo de Confiança da Média** é um intervalo de estimativa em torno da **média da amostra**, que provê um range de valores no qual está a **média da população**. De fato, a média da população raramente é conhecida, assim o **intervalo de confiança** é a **única evidência** que nós temos sobre a **média da população**.