



**Data Science
Academy**

www.datascienceacademy.com.br

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

**Intervalo de Confiança e
Significância Estatística**



Intervalo de Confiança é uma amplitude (ou um intervalo) de valores que tem a probabilidade de conter o valor verdadeiro da população. Observe que na definição de intervalo de confiança, está associado uma probabilidade. A esta probabilidade chamamos de:

Nível de Confiança
Grau de Confiança, ou
Coeficiente de Confiança

Essas probabilidades podem vir a partir de escolhas comuns do grau de confiança que se deseja alcançar, dentre os mais comuns temos:

Grau de Confiança	Nível de Significância α	Valor Crítico Z
90%	0,10	1,645
95%	0,05	1,96
99%	0,01	2,575

Considerando:

$$\text{Probabilidade } \{c1 \leq \mu \leq c2\} = 1 - \alpha$$

Onde:

- O intervalo (c1, c2) é chamado de intervalo de confiança.
- μ = média da população.
- α é o **nível de significância**.
- **100 (1 - α)** é nível de confiança.
- **1 - α** é o coeficiente de confiança.

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. Portanto, o 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.

O principal objetivo da análise estatística de dados é estabelecer se os resultados possuem ou não significância estatística, de acordo com os parâmetros estabelecidos.

Ao formular uma hipótese em relação a uma determinada característica de uma população a amostra pode pertencer ou não à população de origem. Portanto as diferenças observadas são decorrentes de variações normais ou a amostra pode não pertencer a população e as diferenças encontradas representam um efeito real, não podendo ser atribuídas ao acaso. Vale ressaltar que devemos ter os níveis de significância previamente escolhidos.

O nível de significância é o limite que se estabelece para afirmar que um certo desvio é decorrente de acaso ou não.



Os níveis de significância mais aceitos são $P = 0,05$ e $P = 0,01$, ou seja, 5% e 1% respectivamente.

A partir de um nível de significância convencionado (alfa) os desvios são ocasionados pela lei do acaso, e o resultado é considerado não significativo.

Na prática é considerado satisfatório o limite de 5% de probabilidade de erro. Não sendo significativas as diferenças que tiverem uma probabilidade acima desse limite. O nível de significância deve ser estabelecido antes do experimento ser realizado e corresponde ao risco que se corre de rejeitar uma hipótese verdadeira ou aceitar uma hipótese falsa. A Significância de um resultado também é denominada de p-value ou valor-p.

Uma das mais importantes funções da Estatística no mundo atual é coletar informações sobre uma amostra e então usar esta informação para analisar a população da qual a amostra foi extraída. O que fará um Cientista de Dados com essas informações? O Cientista de Dados irá utilizar recursos de intervalo de confiança diariamente para:



- Calcular a variabilidade de tempo de atendimento a clientes (bancos, centrais de atendimento, estabelecimentos em geral).
- Calcular o tempo de aterrissagem de voo.
- Calcular o tempo de substituição de aparelhos eletrônicos em uma Central de Processamento de Dados.
- Calcular a vida útil de componentes elétricos e eletrônicos em ambientes de trabalho e domésticos.
- Calcular os resultados de pesquisas eleitorais ou outras pesquisas em geral

A Estatística oferece ferramentas para analisar dados e resolver problemas. E o que as empresas mais procuram nos dias de hoje é por profissionais que saibam resolver problemas. Seja um desses profissionais.

Referências:

The Logic of Science: Principles and Elementary Applications Vol 1
E. T. Jaynes