

A Importância da Estatística para o Profissional de Ciência de Dados

Olá! Seja bem-vindo(a) à nossa newsletter semanal sobre Ciência de Dados!

Nesta edição, vamos destacar a importância do conhecimento de estatística na atuação profissional do cientista de dados. A estatística é a base fundamental para muitas das técnicas e métodos utilizados na ciência de dados. Desde a coleta de dados até a interpretação dos resultados, o domínio da estatística permite ao cientista de dados realizar análises precisas e tomar decisões informadas. Mas, por que o conhecimento de estatística é tão importante? Destacamos os cinco principais motivos: 1) **Compreensão dos Dados:** A estatística permite que os cientistas de dados compreendam e descrevam os dados que possuem. Isso inclui a identificação de padrões, tendências e anomalias, o que é essencial para qualquer análise subsequente; 2) **Inferência Estatística:** Com a estatística, é possível fazer inferências sobre uma população com base em uma amostra de dados. Isso é crucial para prever tendências futuras e tomar decisões estratégicas baseadas em dados históricos; 3) **Modelagem e Testes:** Muitos modelos preditivos e algoritmos de machine learning são baseados em princípios estatísticos. Conhecer bem a estatística ajuda na escolha e ajuste dos modelos mais adequados para cada situação; 4) **Validação de Resultados:** A estatística oferece ferramentas para validar os resultados de uma análise, garantindo que as conclusões sejam robustas e confiáveis; e 5) **Comunicação de Resultados:** Saber estatística ajuda na interpretação e na comunicação dos resultados para as partes interessadas, que muitas vezes não possuem formação técnica.

A seguir, apresentamos as cinco perguntas mais frequentes realizadas pelos entrevistadores sobre estatística para preenchimento de vagas na área de Ciência de Dados.

#	Pergunta	O que o entrevistador busca saber	Resposta possível
1	Qual é a diferença entre uma amostra e uma população?	Se você entende os conceitos básicos de estatística	Uma população é o conjunto total de itens que estamos interessados em estudar, enquanto uma amostra é um subconjunto dessa população, selecionado para análise.
2	Explique a diferença entre média, mediana e moda.	Seu conhecimento sobre medidas de tendência central	Média é a soma de todos os valores dividida pelo número de valores, mediana é o valor central quando os dados são ordenados, e moda é o valor que aparece com mais frequência.
3	O que é um intervalo de confiança?	Sua compreensão sobre a estimativa de parâmetros populacionais	Um intervalo de confiança é um intervalo estimado a partir dos dados da amostra que, com um certo nível de confiança, contém o valor verdadeiro do parâmetro populacional.
4	Como você lida com valores ausentes ou outliers em um conjunto de dados?	Suas habilidades em preparação e limpeza de dados	Valores ausentes podem ser imputados ou removidos, dependendo da quantidade e da importância. Outliers podem ser investigados para determinar se são erros de dados ou pontos válidos.
5	O que é regressão linear e como ela é usada?	Seu conhecimento sobre técnicas de modelagem	Regressão linear é um método estatístico usado para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes, com o objetivo de prever valores e entender as relações entre as variáveis.

Esperamos que você tenha achado estas informações úteis. Seja você um iniciante no mundo da Ciência de Dados ou um profissional experiente, entender a importância da estatística pode aumentar suas chances de sucesso em entrevistas de emprego e na carreira.

Se você tiver alguma dúvida ou sugestão, não hesite em entrar em contato. Queremos garantir que nossos conteúdos sejam sempre relevantes e úteis para você.

Fique atento à nossa próxima edição, onde continuaremos a explorar ferramentas técnicas e dicas valiosas para aprimorar suas habilidades em Ciência de Dados. Até lá, continue explorando e aprendendo!

Saudações,

Prof. Dr. Dilermando Piva Jr

Coordenador de Ciência de Dados para Negócios / Fatec Votorantim

E-mail: f301.cdn@fatec.sp.gov.br