**Estatística Inferencial**

Estatísticas é uma das habilidades chaves fundamentais requeridos para deep learning. Qualquer profissional em deep learning certamente recomendaria aprender e aprimorar em estatísticas.

1. Porquê precisamos de estatística inferencial?

Suponhamos que você quer saber a média do salário de profissionais de Data Science na Índia. Quais dos métodos podem ser usados para calcular?

* 1. Encontrar todo profissional de data Science na Índia. Anotar os seus salários e calcular a média?
  2. Ou escolher um número de profissionais em uma cidade como Gurgaon. Anotar os seus salários e calcular a média.

Bem, o primeiro método não é impossível mas requer uma enorme quantidade de recursos e tempo. Mas hoje, empresas querem tomar decisões rápidas e de forma rentável, então o primeiro método não tem chance.

De outro modo, o segundo método parece possível. Mas, há um problema. Se a população de Gurgaon não for a mesma população da India inteira? Tem boas chances de você estimar errado o salário de Indianos profissionais em Data Science.

Agora, qual método pode ser usado para estimar a média de salário de todos os profissionais em Data Science na India?

Em termos simples, estatísticas inferenciais podem ser usadas para designar inferências além da data imediata disponível.

Com a ajuda de estatísticas inferenciais, nós podemos responder as questões abaixo:

* Extrair inferências sobre a população a partir de um exemplo.
* Concluir se um exemplo é significante diferente da população. Por exemplo, vamos dizer que você colecionou os detalhes de um salário de profissionais de Data Science em Bangalore. E você observou que a média de salário de profissionais de Data Science em Bangalore é maior que a média de salário na Índia. Agora, podemos concluir se a diferença é estaticamente significante.
* Se adicionando ou removendo uma característica de um modelo vai realmente ajudar a melhorar o modelo.
* Se um modelo é significativamente melhor que o outro?
* Testando hipótese em geral.

1. Pré-requisitos

Para começar com Estatísticas Inferenciais, é necessário ter um conhecimento em Estatísticas Descritivas. Você pode ver por [este curso](https://www.youtube.com/playlist?list=PLAwxTw4SYaPn22DmaF6x8JtG4TeWOJk_1) no youtube para aprender sobre estatísticas descritivas.

Aqui está uma breve descrição de algumas terminologias que vamos estar usando:

* **Estatística –** Uma única medida de algum atributo de um exemplo. Por exemplo: Média/Mediana/Modo de um exemplo de Cientistas de Dados em Bangalore
* **Estatística Populacional –** A estatística da população inteira em contexto. Por exemplo: Média populacional do salário de toda a população de cientistas de dados pela Índia.
* **Estatística de Amostra –** A estatística de um grupo retirado de uma população. Por exemplo: Média dos salários de todos os cientistas de dados em Bangalore.
* **Desvio padrão –** É a variação total de um dado populacional. É dado por σ.
* **Erro padrão –** É a variação total em um dado de exemplo. É relacionado ao desvio padrão como σ/√n, onde n, é o tamanho de exemplo.

1. Distribuição