

Entrega 3 - Estatística Inferencial

Objetivo do Projeto

O projeto de Ciência de Dados do grupo consiste em analisar a diversidade e as tendências do mercado de TI, ligado à área de Dados no Brasil, focando em como as vagas de emprego estão sendo distribuídas entre diferentes grupos, como por exemplo, gênero, raça, orientação sexual e deficiência. Além disso, queremos filtrar quais tecnologias estão sendo mais demandadas e quais níveis de senioridade estão sendo mais buscados.

O objetivo geral do nosso projeto é gerar análises eficientes para entender como as empresas filtram candidatos, gerando assim maior entendimento na hora dos processos seletivos. Além disso, queremos saber como esses profissionais desempenham seu trabalho, seja por senioridade, horas trabalhadas e entre outros aspectos.

Nas entregas da matéria de Análise Inferencial de Dados, resolvemos filtrar nossa pesquisa em faixa salarial e idade dos participantes.

Construção e Interpretação do Intervalo de Confiança

```
> #ENTREGA 03
>
> # Usar a mesma planilha e colunas
> # Calcular a média, desvio padrão, e tamanho da amostra para "salario_Convertido"
> media_salario <- mean(dados_pesquisa_2023$`salario_Convertido`, na.rm = TRUE)
> desvio_padrao_salario <- sd(dados_pesquisa_2023$`salario_Convertido`, na.rm = TRUE)
> n <- length(na.omit(dados_pesquisa_2023$`salario_Convertido`))
>
> # Nível de confiança (por exemplo, 95%)
> nivel_confianca <- 0.95
> z <- qnorm(1 - (1 - nivel_confianca) / 2)
>
> # calcular o erro padrão
> erro_padrao <- desvio_padrao_salario / sqrt(n)
>
> # calcular o erro (Epsilon)
> epsilon <- z * erro_padrao
>
> # Intervalo de confiança
> limite_inferior <- media_salario - epsilon
> limite_superior <- media_salario + epsilon
>
> # calcular o erro usando a largura do intervalo
> epsilon_calculado <- (limite_superior - limite_inferior) / 2
>
> # Exibir resultados
> cat("Intervalo de Confiança para salario_Convertido:", limite_inferior, "a", limite_superior, "\n")
Intervalo de Confiança para salario_Convertido: 8617.899 a 9011.442
> cat("Erro (Epsilon) calculado pela largura do intervalo:", epsilon_calculado, "\n")
Erro (Epsilon) calculado pela largura do intervalo: 196.7715
>
```

Conclusão

Com os dados amostrais coletados pelo grupo, obtemos o intervalo de confiança e podemos confirmar que, pegando uma quantidade de dados consideráveis, os valores estarão entre uma média de R\$8.617 e R\$9.011. Junto a análise do intervalo de confiança, também obtivemos um erro amostral de R\$196, ou seja, a média salarial (R\$8.815) pode variar para mais ou para menos deste valor.