

Entrega 2 - 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.

Explorar Relações

```
# Criar um gráfico de dispersão entre "Salario" e "Idade"
plot(dados$`Idade`,
     dados$`Salario`,
     main = "Gráfico de Dispersão: Idade vs Salario",
     xlab = "Idade",
     ylab = "Salario",
     col = "blue",
     pch = 19)

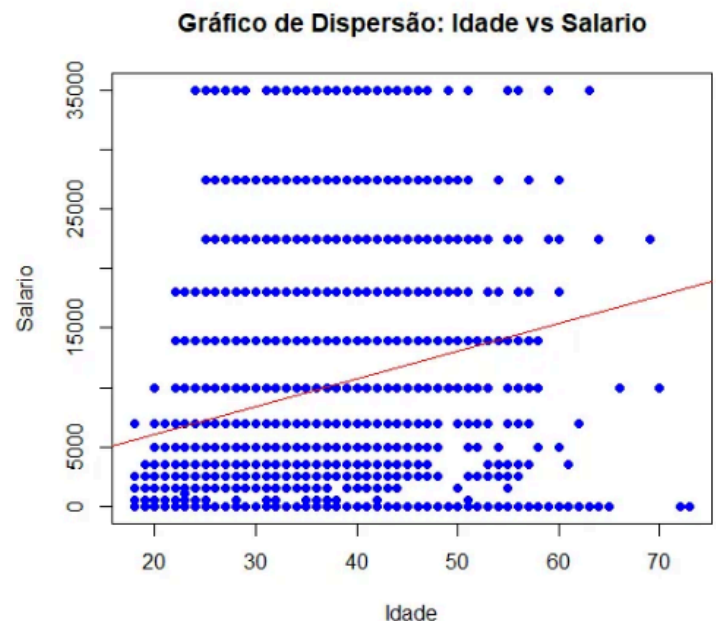
# Adicionar uma linha de tendência (regressão linear)
abline(lm(dados$`Salario` ~ dados$`Idade`), col = "red")

# Ajustar um modelo de regressão linear
modelo_regressao <- lm(dados$`Salario` ~ dados$`Idade`)

# Obter os coeficientes da regressão
coeficientes <- coef(modelo_regressao)

# Mostrar a equação de regressão
cat("Equação de Regressão: Salario = ",
    round(coeficientes[1], 2), "+",
    round(coeficientes[2], 2), "* Idade\n")

# Ajustar um modelo de regressão linear
modelo_regressao <- lm(dados$`Salario` ~ dados$`Idade`)
```



Escolher o Modelo

O modelo melhor adaptado para o grupo foi a regressão linear.

Ajustar e analisar - Coeficiente R^2

```
# Calcular o  $R^2$ 
r2 <- summary(modelo_regressao)$r.squared

# Exibir o  $R^2$  no console
cat("R² do Modelo de Regressão: ", round(r2, 4), "\n")
```

Valor do $R^2 \rightarrow 0.0591$

Conclusão

Após a análise crítica dos gráficos e valores obtidos, podemos concluir que provavelmente o modelo utilizado pelo grupo é fraco, possivelmente causado pela “pouca aproximação da reta”, ou seja, não há uma relação exata e causal entre os dados obtidos em relação a idade e a média da faixa salarial, havendo então alta variação entre os dados.