**Resumo do Projeto: Análise de Relação entre Absenteeism e Wage**

**Pergunta de Pesquisa:**

A questão central do meu projeto foi: **"Será que o número de dias de ausência (Absenteeism) tem impacto no salário (Wage) dos funcionários?"**

**Definição das Hipóteses:**

Para testar essa relação, defini duas hipóteses:

* **Hipótese Nula (H₀):** Não há interferência entre **Absenteeism** e **Wage**. Ou seja, **Absenteeism** não afeta o salário dos funcionários.
* **Hipótese Alternativa (Hₐ):** Há uma interferência entre **Absenteeism** e **Wage**. Ou seja, o número de dias de ausência afeta o salário dos funcionários.

**Testes de Normalidade:**

Antes de realizar o teste de hipótese, foi importante verificar se a variável **Absenteeism** segue uma distribuição normal. Isso é necessário porque alguns testes estatísticos assumem que os dados vêm de uma distribuição normal.

Escolhi realizar o **teste de normalidade Shapiro-Wilk**, pois tenho apenas 40 dados (uma amostra relativamente pequena), e esse teste é ideal para amostras pequenas.

* **Hipótese do teste de normalidade:**
  + **Hipótese Nula (H₀):** Os dados de **Absenteeism** seguem uma distribuição normal.
  + **Hipótese Alternativa (Hₐ):** Os dados de **Absenteeism** não seguem uma distribuição normal.

O **p-value** do teste de normalidade indicou que **os dados de Absenteeism seguem uma distribuição normal**, permitindo que eu prosseguisse com os testes paramétricos.

**Escolha do Modelo de Teste de Hipótese:**

Decidi usar o **teste de correlação de Pearson** para investigar a relação entre **Absenteeism** e **Wage**, pois:

* Ambas as variáveis são **contínuas** e podem ter uma relação linear.
* O **teste de correlação de Pearson** é adequado para medir a força e a direção dessa relação linear.
* **Hipóteses do teste de correlação de Pearson:**
  + **Hipótese Nula (H₀):** Não há correlação significativa entre **Absenteeism** e **Wage**. Ou seja, **Absenteeism** não tem impacto em **Wage**.
  + **Hipótese Alternativa (Hₐ):** Existe uma correlação significativa entre **Absenteeism** e **Wage**. Ou seja, **Absenteeism** tem impacto em **Wage**.

**Nível de Significância e Interpretação do p-value:**

Escolhi um **nível de significância de 0.05**, o que significa que estou disposto a aceitar uma probabilidade de 5% de cometer um erro tipo I (rejeitar a hipótese nula quando ela é verdadeira).

Ao realizar o teste de correlação de Pearson, obtive os seguintes resultados:

* **Coeficiente de Correlação (r): -0.5014**
* **p-value: 0.0010**

**Resultado Final:**

Com base nos resultados do teste de correlação de Pearson:

* O **coeficiente de correlação de -0.5014** indica uma correlação negativa moderada entre **Absenteeism** e **Wage**. Isso sugere que, à medida que os dias de ausência aumentam, o salário tende a diminuir.
* O **p-value de 0.0010** é muito menor que o nível de significância de 0.05, o que nos leva a **rejeitar a hipótese nula (H₀)**.

Portanto, **há uma correlação significativa entre Absenteeism e Wage**. Isso significa que **Absenteeism interfere no Wage**.

ANALISE CONJUNTA DAS VARIAVEIS:

Exatamente! Essa é uma ótima forma de interpretar seus resultados. Vamos organizar essa explicação de forma clara e didática para que você possa apresentar com segurança:

✅ **Isoladamente**, todas as variáveis (Absenteeism, Education e Experience) mostram algum grau de influência no salário. Isso ocorre porque, ao analisar uma variável por vez, você não está controlando o impacto das outras.

✅ **Quando analisadas juntas**, apenas **Education** se destaca como significativa, indicando que sua influência no salário é mais robusta. As variáveis **Absenteeism** e **Experience** perdem sua significância, o que sugere que seu efeito pode estar sendo absorvido por outros fatores — ou que, na presença de **Education**, seu impacto é mínimo.

✅ O valor de **R² ajustado = 0.54** indica que seu modelo explica **54% da variação nos salários**, o que é uma explicação moderada. Isso é comum em dados reais, especialmente em temas sociais e econômicos, onde muitos outros fatores podem estar em jogo.

✅ Para melhorar essa análise, você pode explorar outras variáveis relevantes, como:

* **Setor de atuação**
* **Cargo ou nível hierárquico**
* **Tamanho da empresa**
* **Localização geográfica**
* **Habilidades técnicas específicas**

**💬 Resumo para seu relatório**

"A análise inicial indicou que todas as variáveis analisadas apresentavam alguma relação com o salário quando testadas isoladamente. No entanto, ao aplicar um modelo que considera todas as variáveis simultaneamente, apenas **Education** se mostrou significativa. Esse modelo explica aproximadamente **54%** da variação nos salários, sugerindo que há outros fatores importantes que ainda não foram incluídos. Para uma análise mais robusta, recomenda-se explorar variáveis adicionais e testar modelos alternativos."

Essa abordagem mostra que você entende tanto os resultados quanto suas limitações — o que é essencial para uma análise sólida. 🚀