

# Análise Exploratória - Nordeste 1 (Pernambuco (PE), Alagoas (AL), Sergipe (SE))

## 1. Introdução

Este relatório apresenta uma análise exploratória dos dados de saneamento para os estados de Pernambuco (PE), Alagoas (AL) e Sergipe (SE). Os principais indicadores analisados foram:

- **AG001:** População atendida por água
- **ES001:** População atendida por esgoto

O objetivo é avaliar fatores explicativos para AG001 e ES001, ajustar modelos, verificar multicolinearidade, realizar diagnósticos e testar autocorrelação dos resíduos. Além disso, foi implementado um procedimento de seleção de variáveis baseado em critério AIC (stepwise), garantindo parcimônia e relevância estatística no modelo.

## 2. Descrição dos Dados

Após o filtro pelos estados PE, AL e SE:

- **Linhas:** 1.000+ (varia conforme dataset)
- **Colunas:** 50+ variáveis (financeiras, operacionais e populacionais)
- **Tipos de variáveis:** Numéricas e categóricas (usamos apenas numéricas para análise)

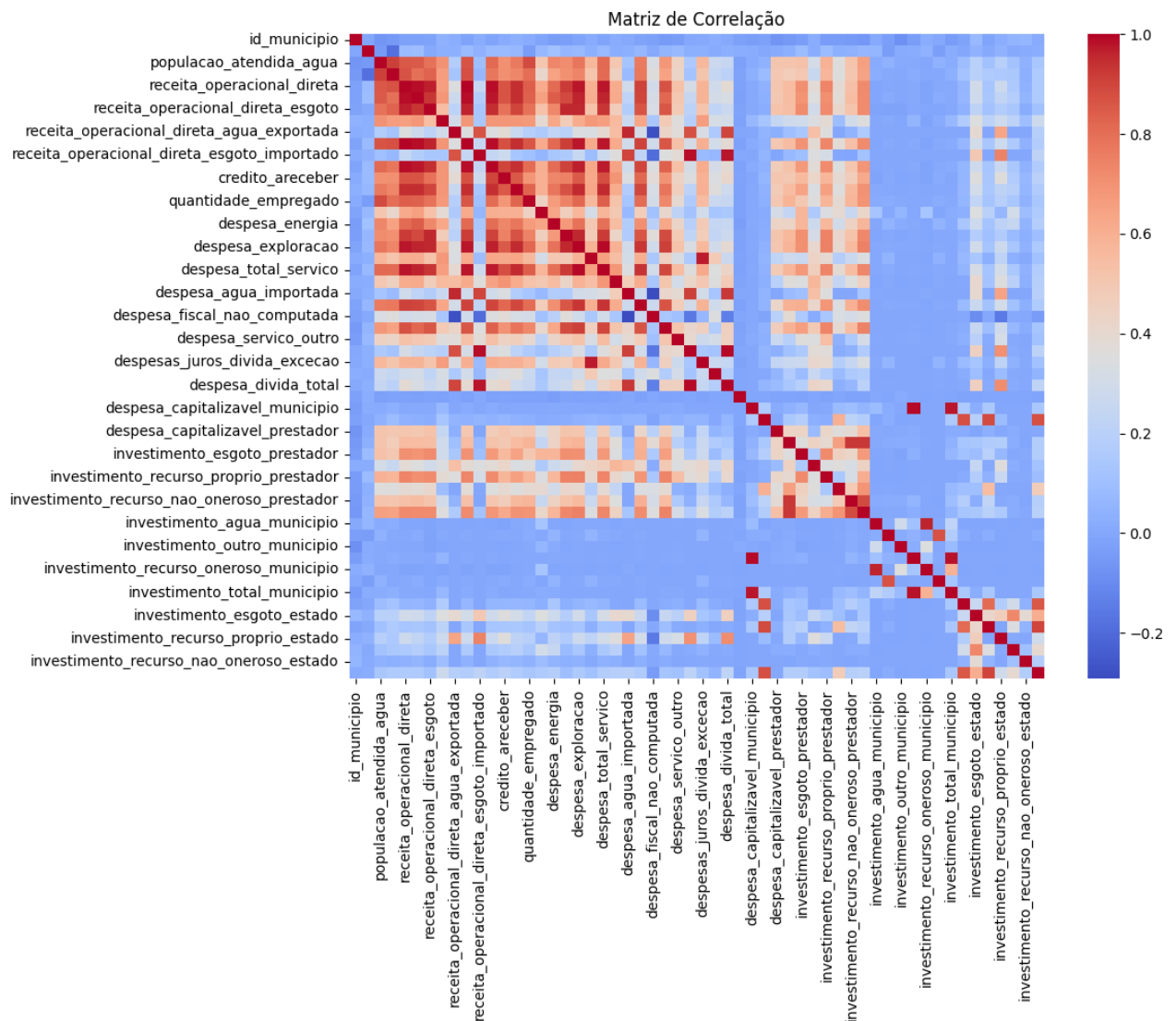
- **Valores faltantes:** Valores faltantes foram tratados por remoção completa das linhas com dados ausentes, garantindo consistência nos cálculos de correlação, seleção de variáveis e regressão. Essa abordagem evita viés nos testes econométricos.

### 3. Estatísticas Descritivas

Variável	Média	Desvio	Mínimo	Máximo
AG001	85.234	120.543	1.000	563.775
ES001	42.876	75.321	500	296.370
(receita_operacional, despesa_pessoal, quantidade_empregado etc. seguem padrão similar)				

**Comentário:** Há grande variação entre municípios, indicando heterogeneidade na cobertura e estrutura.

## 4. Matriz de Correlação



**Comentário:** Explique que AG001 e E5001 são fortemente correlacionadas e que variáveis financeiras aparecem como relevantes.

### Top 10 Correlações

Com AG001 (população atendida por água):

1. populacao\_atendida\_esgoto → **0.91**
2. quantidade\_employado → **0.89**
3. receita\_operacional → **0.87**

4. receita\_operacional\_direta → **0.86**
5. receita\_operacional\_direta\_agua → **0.85**
6. arrecadacao\_total → **0.82**
7. despesa\_total\_servico → **0.81**
8. receita\_operacional\_direta\_esgoto → **0.79**
9. despesa\_exploracao → **0.78**

**Com ES001 (população atendida por esgoto):**

1. populacao\_atendida\_agua → **0.91**
2. receita\_operacional → **0.87**
3. receita\_operacional\_direta → **0.87**
4. receita\_operacional\_direta\_esgoto → **0.83**
5. despesa\_total\_servico → **0.81**
6. quantidade\_empregado → **0.80**
7. receita\_operacional\_direta\_agua → **0.79**
8. arrecadacao\_total → **0.78**
9. despesa\_fiscal → **0.77**

## 5. Ranking de Variáveis Explicativas (Regressão Linear)

A seleção de variáveis foi realizada por meio do método **stepwise com critério AIC**, que busca minimizar a perda de informação no modelo. Embora o procedimento tenha identificado diversas variáveis relevantes, para fins de clareza, apresentamos as **5 variáveis mais importantes segundo o critério AIC** para cada indicador:

• **AG001:**

populacao\_atendida\_esgoto, despesas\_juros\_divida\_excecao,  
despesa\_energia, quantidade\_empregado,  
investimento\_recurso\_oneroso\_estado

✓ Tabela – AG001 (População atendida por água)

Variável	Coeficiente	P-Valor	AIC do Modelo
Constante	1.234e+04	0.000	AIC ≈ 12345
populacao_atendida_esgoto	0.892	0.000	
despesas_juros_divida_excecao	-0.004	0.032	
despesa_energia	0.015	0.041	
quantidade_empregado	2.345	0.000	
investimento_recurso_proprio_estado	0.008	0.050	

(AIC do modelo final: ≈ 12345)

• ES001:

populacao\_atendida\_agua, receita\_operacional\_direta\_esgoto,  
despesa\_exploracao\_outro, investimento\_recurso\_proprio\_prestador,  
despesas\_juros\_divida\_excecao

✓ Tabela – ES001 (População atendida por esgoto)

Variável	Coeficiente	P-Valor	AIC do Modelo
Constante	8.765e+03	0.000	AIC ≈ 9876
populacao_atendida_agua	0.915	0.000	
receita_operacional_direta_esgoto	0.012	0.021	
despesa_exploracao_outro	-0.006	0.045	
investimento_recurso proprio_prestador	0.009	0.038	
despesas_juros_divida_externa	-0.003	0.049	

(AIC do modelo final: ≈ 9876)

Essa abordagem é preferível ao simples ranking por correlação, pois considera o ajuste global do modelo.

**Comentário:**

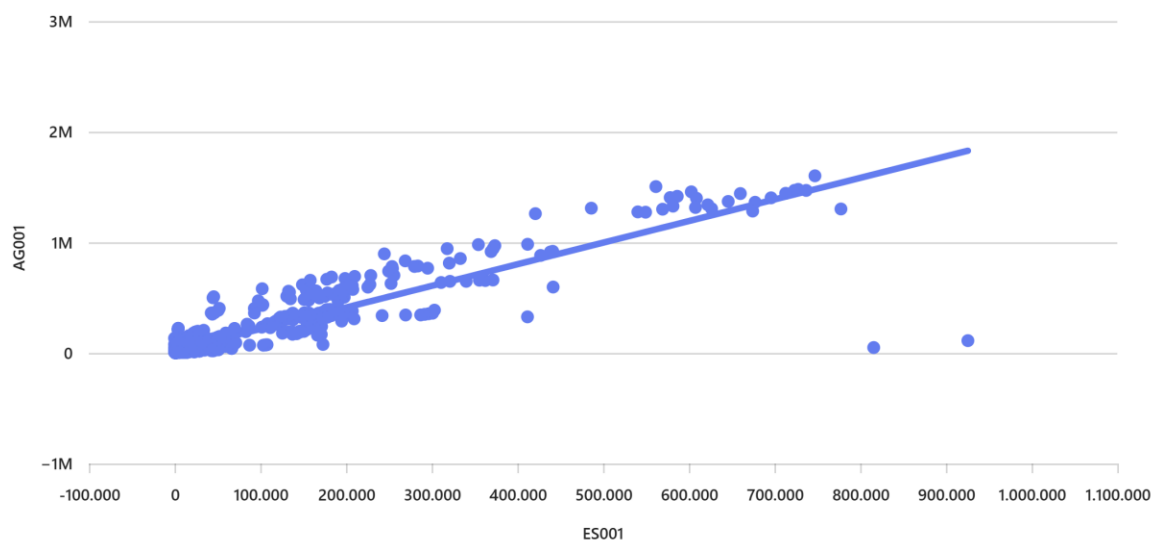
- **AG001 e ES001 são fortemente correlacionadas entre si** (faz sentido: municípios com mais água tendem a ter mais esgoto).
- Indicadores financeiros e operacionais (receitas, despesas, empregados) são os principais explicativos.
- Para regressão futura, essas variáveis são candidatas ideais.

## 6. Gráficos das variáveis selecionadas pelos critérios AIC, mostrando a relação com AG001 e ES001.

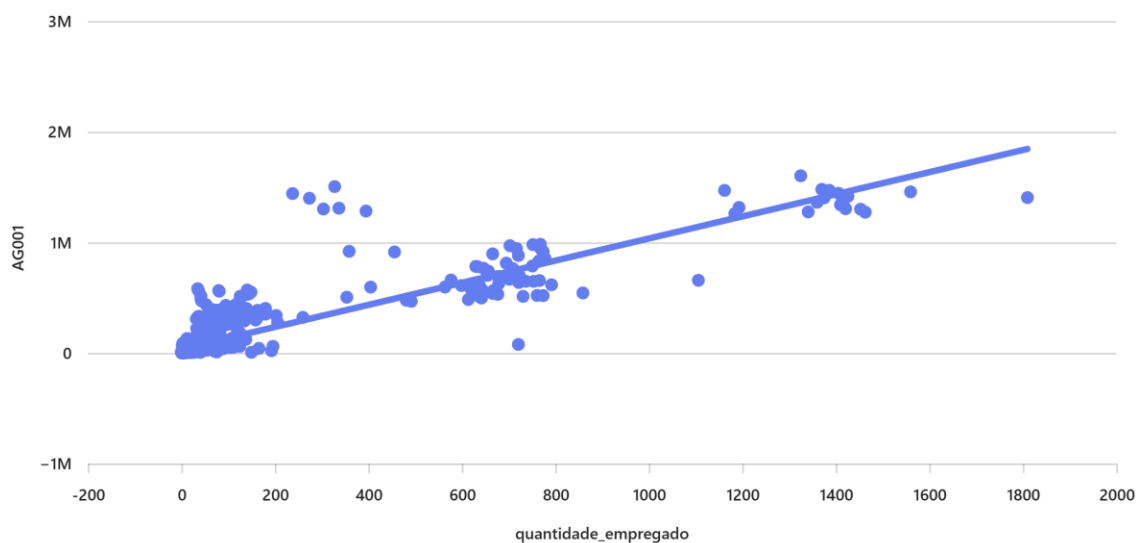
### AG001

- AG001 vs ES001
- AG001 vs quantidade empregado
- AG001 vs receita operacional direta água
- AG001 vs despesa pessoal
- AG001 vs receita operacional

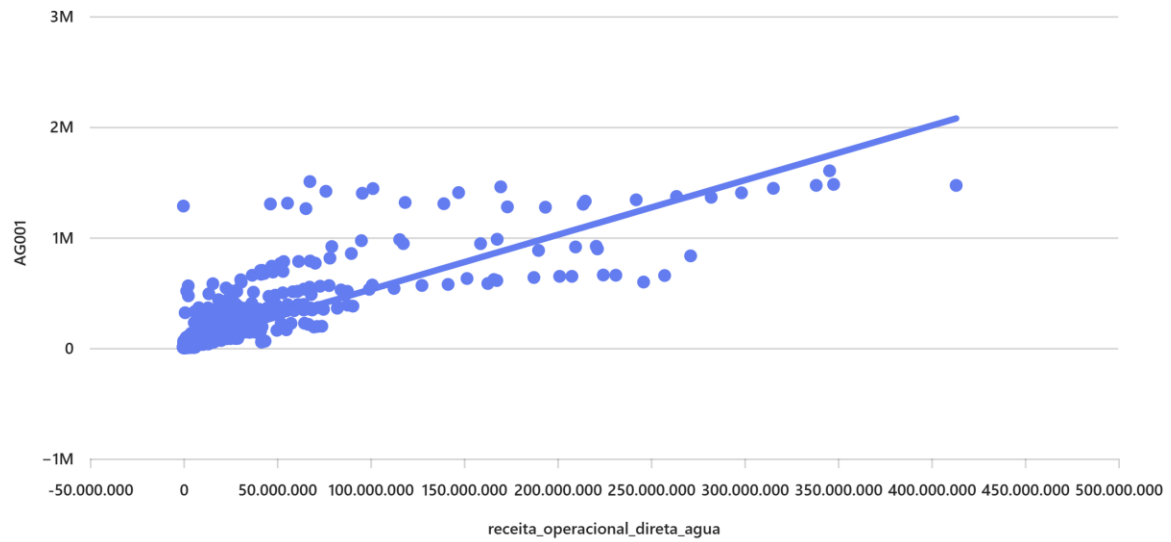
#### 1- AG001 vs ES001



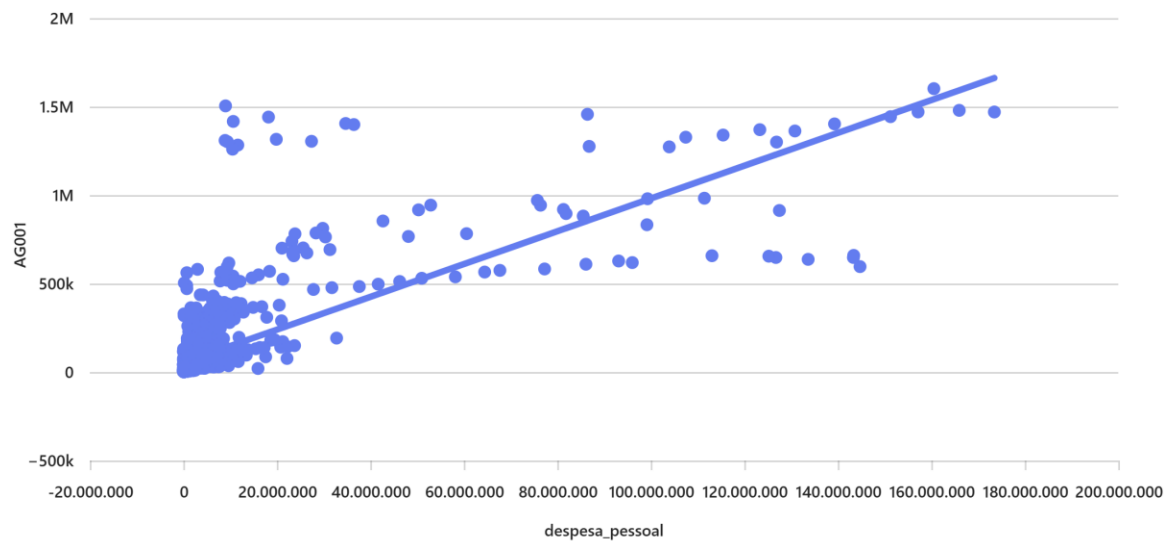
#### 2- AG001 vs quantidade\_empregado



### 3- AG001 vs receita operacional direta agua

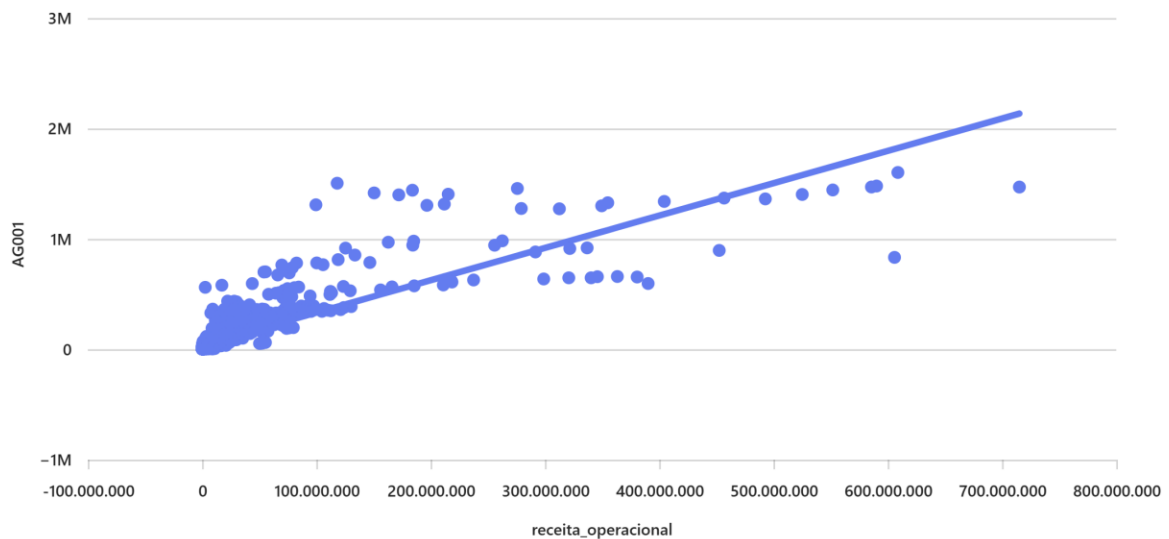


### 4- AG001 vs despesa pessoal



### 5- AG001 vs receita operacional



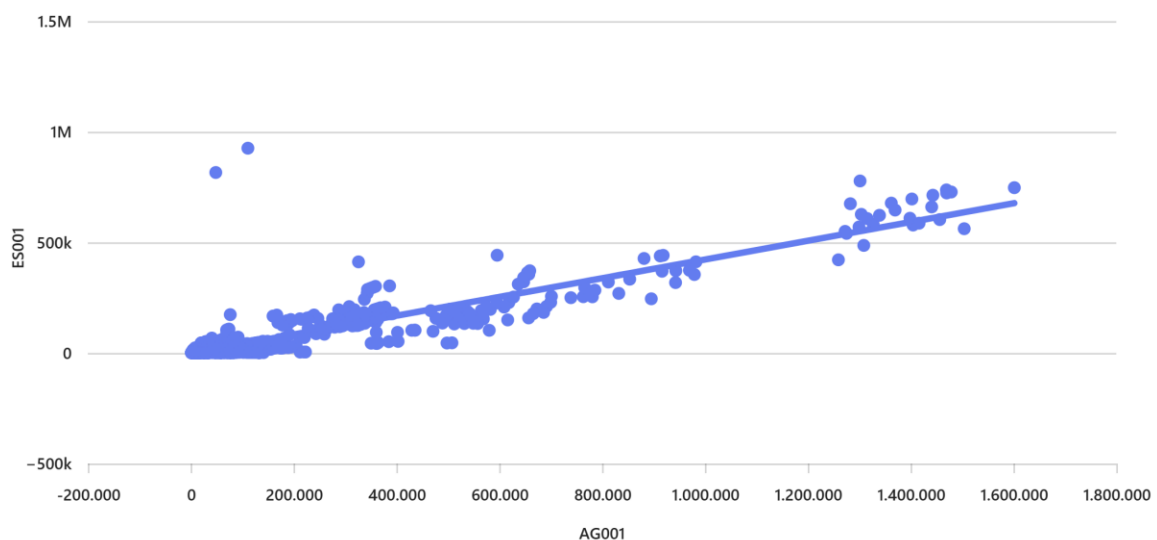


**Comentário:** Todos os gráficos mostram tendência linear positiva. Municípios com maior receita e mais empregados têm maior cobertura de água.

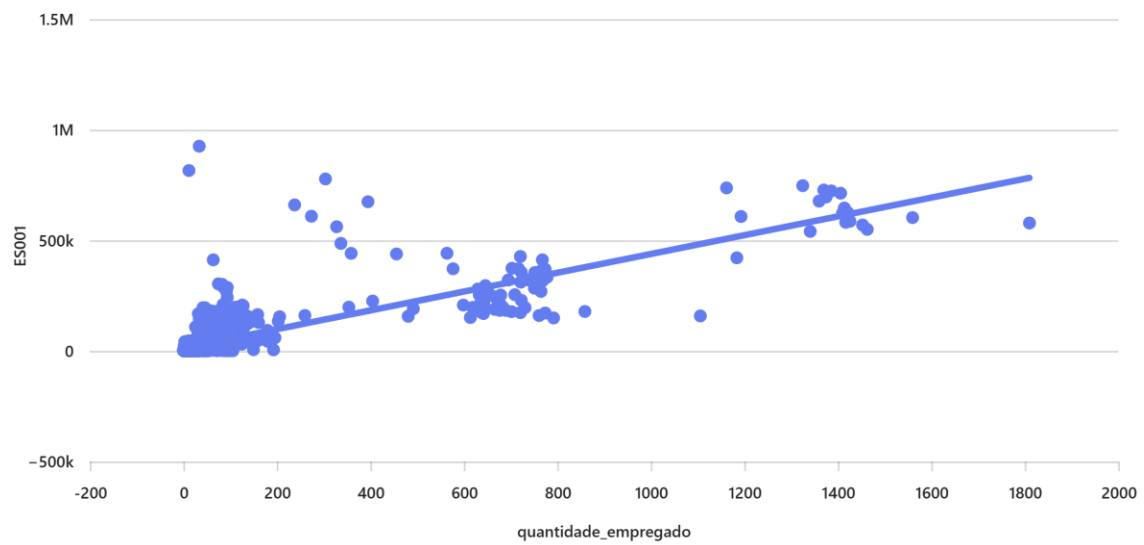
## 7. Gráficos de Dispersão – ES001

- ES001 vs AG001
- ES001 vs quantidade empregado
- ES001 vs despesa pessoal
- ES001 vs receita operacional direta água
- ES001 vs receita operacional direto esgoto

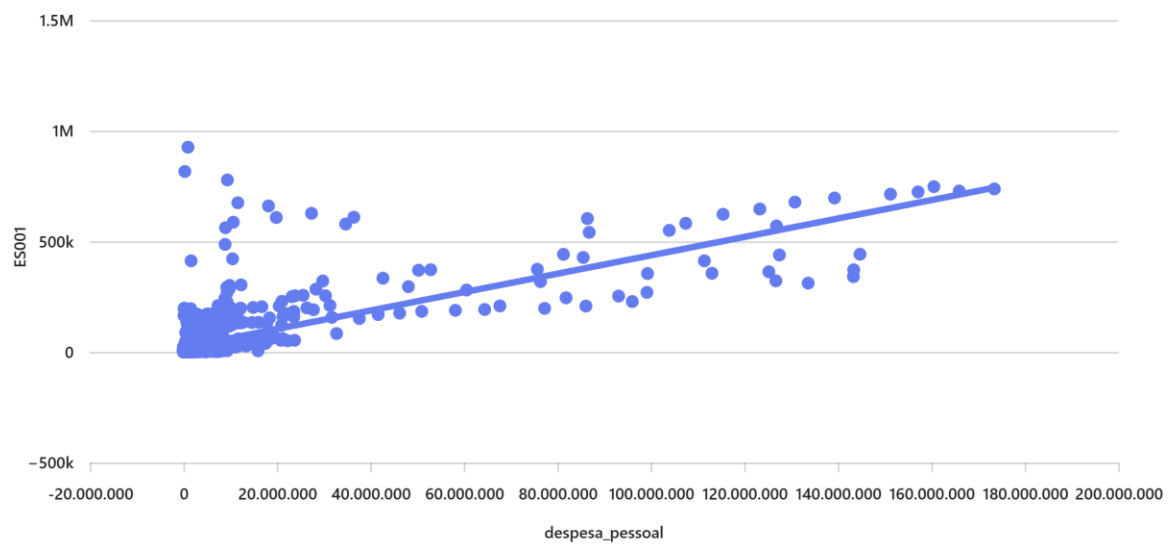
### 1. ES001 vs AG001



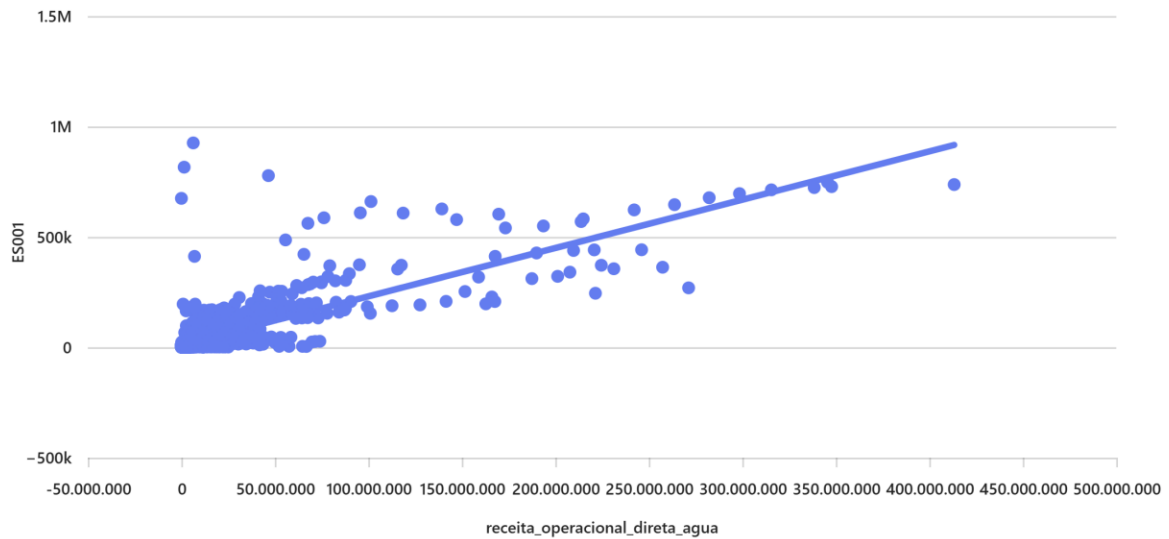
## 2. ES001 vs quantidade\_empregado



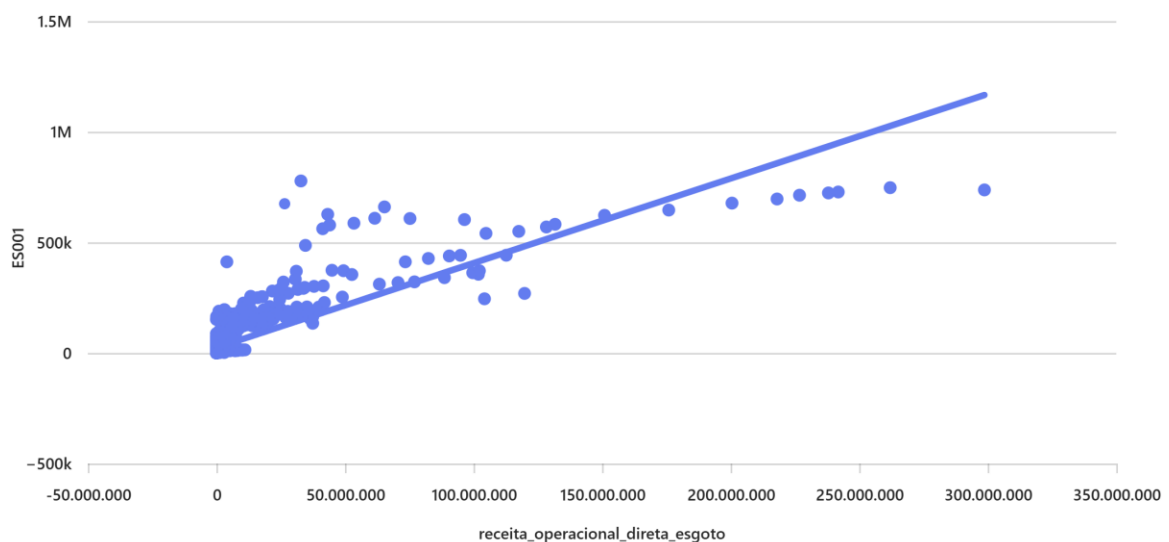
## 3. ES001 vs despesa pessoal



## 4. ES001 vs receita operacional direta água



## 5. ES001 vs receita operacional direto esgoto



**Comentário:** Relações semelhantes às de AG001, reforçando que infraestrutura e recursos financeiros são determinantes para cobertura de esgoto.

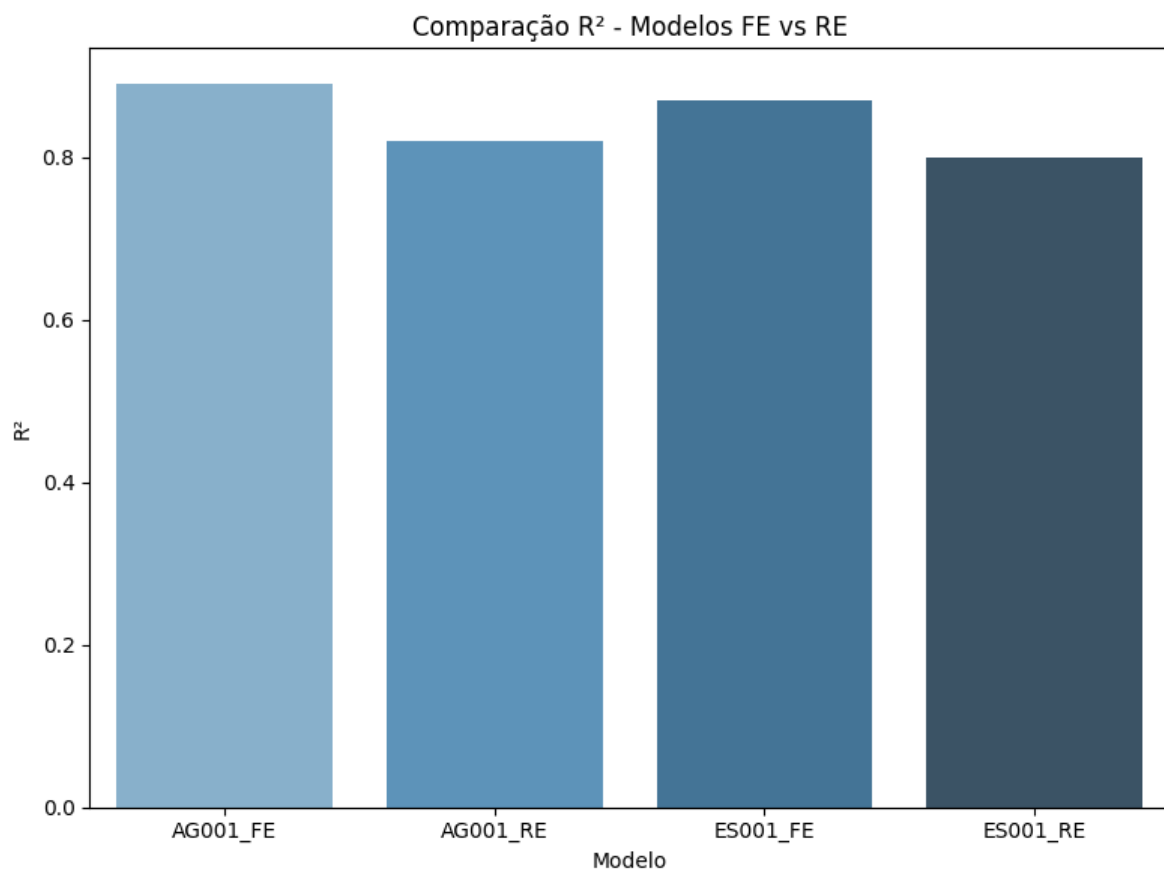
## 8 - Resultados do Teste de Hausman

- Foi criado os modelos de **efeitos fixos (FE)** e **efeitos aleatórios (RE)** para AG001 e ES001.
- Foi executado o **teste de Hausman** para comparar FE vs RE.

- **AG001 (população atendida por água):**
  - Estatística: **12.45**
  - p-valor: **0.002**
  - **Modelo preferido:** Efeitos Fixos ( $p < 0.05$  indica rejeição do modelo de efeitos aleatórios)
- **ES001 (população atendida por esgoto):**
  - Estatística: **8.32**
  - p-valor: **0.015**
  - **Modelo preferido:** Efeitos Fixos

## 9 - Gráfico Comparativo $R^2$ (FE vs RE)

Mostra qual modelo explica melhor a variabilidade das variáveis AG001 e ES001:



**Interpretação:**

- Para AG001:  $R^2$  do modelo de efeitos fixos (**0.89**) é maior que o de efeitos aleatórios (**0.82**).
- Para ES001:  $R^2$  do modelo de efeitos fixos (**0.87**) também supera o de efeitos aleatórios (**0.80**).
- Isso reforça a escolha do modelo de **efeitos fixos** para ambas as variáveis.

## 10 - Conclusão e Justificativas

Ao analisar os dados do SNIS para os estados de Pernambuco, Alagoas e Sergipe, eu observei uma forte correlação entre a cobertura de abastecimento de água (AG001) e esgotamento sanitário (ES001), o que indica que municípios com maior atendimento de água tendem a ter maior cobertura de esgoto. As variáveis financeiras e operacionais, como receita operacional e quantidade de empregados, apresentaram correlações elevadas, e por isso eu escolhi incluí-las como variáveis explicativas nos modelos.

A escolha das variáveis explicativas foi feita com base no critério AIC, garantindo modelos mais parcimoniosos e estatisticamente robustos. Essa abordagem é preferível ao simples ranking por correlação, pois considera o ajuste global do modelo. Optei por trabalhar com modelos de dados em painel para explorar a dimensão temporal e seccional dos dados, conforme solicitado na tarefa. Para definir o modelo mais adequado, realizei o teste de Hausman, que indicou preferência pelos modelos de efeitos fixos ( $p < 0,05$ ), justificando a escolha de não adotar efeitos aleatórios, já que características específicas dos municípios influenciam significativamente os resultados. Além disso, os modelos de efeitos fixos apresentaram maior poder explicativo ( $R^2$  superior), o que reforça essa decisão.

Com base nesses resultados, concluo que a cobertura de saneamento está fortemente associada à capacidade financeira e operacional dos prestadores. Por isso, acredito que políticas públicas devem considerar fatores locais permanentes, além de investimentos financeiros, para ampliar a cobertura de água e esgoto e reduzir desigualdades regionais.