

Prova 1 - Análise do Aluguel de Bicicletas

Contexto

Você foi contratado para ajudar uma empresa de compartilhamento de bicicletas a entender os fatores que afetam o número de bicicletas alugadas diariamente. A empresa coletou um conjunto de dados com informações sobre o número de bicicletas alugadas por dia, bem como características climáticas e sazonais.

Dicionário de Dados

- **instant**: índice do registro.
- **dteday**: data.
- **season**: estação (1: primavera, 2: verão, 3: outono, 4: inverno).
- **yr**: ano (0: 2018, 1: 2019).
- **mnth**: mês (1 a 12).
- **holiday**: indica se o dia é feriado ou não.
- **weekday**: dia da semana.
- **workingday**: indica se o dia é útil (1: não é fim de semana nem feriado, 0: caso contrário).
- **weathersit**:
 - 1: Céu limpo, poucas nuvens, parcialmente nublado.
 - 2: Névoa + Nublado, Névoa + Nuvens dispersas, Névoa + Poucas nuvens, Névoa.
 - 3: Neve leve, Chuva leve + Trovoadas + Nuvens dispersas, Chuva leve.
 - 4: Chuva forte + Granizo + Trovoadas + Névoa, Neve + Nevoeiro.
- **temp**: temperatura em graus Celsius.
- **atemp**: sensação térmica em graus Celsius.
- **hum**: umidade.

- **windspeed**: velocidade do vento.
- **casual**: contagem de usuários casuais.
- **registered**: contagem de usuários registrados.
- **cnt**: contagem total de bicicletas alugadas, incluindo usuários casuais e registrados.

Objetivo

Utilizando um **modelo linear tradicional**, responda às seguintes perguntas:

1. Análise descritiva dos dados

- Promova uma análise exploratória inicial:
 - Resuma estatisticamente as variáveis numéricas (use medidas como média, mediana, desvio-padrão, ou outras de sua preferência).
 - Explore as distribuições das variáveis.
 - Visualize a relação entre o número de alugueis e cada variável explicativa, como temperatura e dia da semana.
 - Gráficos são bastante úteis nesse momento.

2. Definição do modelo

- Proponha um modelo linear para explicar o número de bicicletas alugadas, considerando as variáveis fornecidas e a sua análise exploratória, bem como qualquer conhecimento prévio sobre o problema.
- Apresente a fórmula do modelo e justifique a escolha das variáveis incluídas. É razoável se você preferir utilizar o método stepwise.
- Exemplo de modelo sugerido:

$$\text{alugueis} = \beta_0 + \beta_1 \cdot \text{temperatura} + \beta_2 \cdot \text{vento} + \beta_3 \cdot \text{precipitacao} + \beta_4 \cdot \text{feriado} + \varepsilon$$

3. Ajuste do modelo

- Ajuste o modelo linear aos dados fornecidos, estimando os coeficientes utilizando qualquer apoio computacional que você escolher.

4. Avaliação da qualidade do ajuste

- Avalie a qualidade do modelo ajustado:
 - Analise os resíduos do modelo (gráficos de resíduos).
 - Interprete o coeficiente de determinação R^2 .
 - Comente sobre a validade das suposições do modelo linear (normalidade, homocedasticidade, etc.).

5. Interpretação dos efeitos estimados

- Interprete os coeficientes estimados (β):
 - Qual é o impacto da temperatura no número de alugueis?
 - Como os dias de feriado afetam os alugueis em comparação com dias normais?
 - Há uma relação clara entre velocidade do vento e o número de bicicletas alugadas?

Entregáveis

1. Um relatório contendo:
 - A análise descritiva dos dados.
 - A fórmula do modelo linear proposto e sua justificativa.
 - Os resultados do ajuste do modelo, incluindo os coeficientes estimados.
 - A avaliação da qualidade do ajuste.
 - Interpretação dos efeitos estimados.
2. Gráficos e tabelas que suportem as análises realizadas:
 - Gráficos de dispersão, boxplots e histogramas para análise descritiva.
 - Gráficos de resíduos, ou quaisquer outros, para avaliação do modelo.
3. Conclusões gerais baseadas nos resultados da análise.

O relatório será entregue de forma digital através do sigaa. Você pode escrever o seu relatório em qualquer formato (Word, L^AT_EX, pdf, markdown, jupyter, etc.)