

**ASSUNTO:** Análise de correlação linear; análise de regressão linear simples e múltipla: estimativa e interpretação dos coeficientes de regressão, significância dos coeficientes e predição; análise de resíduos; análise de multicolinearidade; seleção de modelos.

**GRUPOS:** No máximo 2 (dois) alunos por grupo. Portanto, pode ser feito individualmente.

**DADOS:** Deverá ser definido e coletado pelo grupo para realizar análise de regressão linear múltipla ( $Y, X_1, \dots, X_q$ ). Como o objetivo é aplicação de regressão linear múltipla, pelo menos inicialmente, note que **é necessário escolher um conjunto de dados com, no mínimo, duas variáveis explicativas, isto é,  $q \geq 2$** . A data limite para definição da base de dados e encaminhamento por e-mail para verificação da adequação deve ser feita **até o dia 15/10/2022, impreterivelmente**.

**ENTREGA:** O relatório contendo as análises (PDF) e o conjunto de dados utilizado devem ser enviados através do e-mail [guilopes2110@gmail.br](mailto:guilopes2110@gmail.br) **até o dia 21/10/2022, impreterivelmente**.

**ROTEIRO:** Para o seu banco de dados específico, cada grupo deve realizar uma análise de regressão linear e produzir um relatório referente à análise dos dados, com 5 a 8 páginas (descontando elementos como capa e sumário), contemplando os seguintes elementos:

1. **Capa** contendo o título do trabalho e nomes dos alunos;
2. **Resumo**, com 10 a 15 linhas, descrevendo brevemente o contexto e o objetivo do estudo, a metodologia, resultados principais e discussão;
3. **Introdução**, em que deverá ser apresentado, em linhas gerais, o contexto e informações adicionais sobre os dados e os objetivos da análise. Caso tais informações não estejam disponíveis na fonte dos dados, os próprios alunos podem elaborar uma breve motivação para o problema;
4. **Materiais e Métodos**, em que deverão ser descritos com maior detalhamento os dados (origem, tamanho da amostra, forma como foi selecionada ou delineado o experimento (se disponível), descrição das variáveis e suas escalas), bem como os recursos computacionais e métodos estatísticos utilizados.

**Atenção:** Não é necessário escrever na metodologia a teoria de regressão linear, mas sim descrever adequadamente o processo de modelagem adotado.

5. **Resultados e Discussão**, em que deverão ser apresentados os principais resultados da análise e a interpretação/discussão dos mesmos. **Atenção:**

- Seja parcimonioso quanto aos resultados incluídos no relatório. Obviamente, nem todos os resultados produzidos na análise precisam ser relatados. Algumas representações fundamentais:
  - Gráficos e ou tabelas resultantes de uma análise descritiva exploratória;
  - Os resultados referentes ao(s) modelo(s) ajustado(s) na forma de gráficos ou tabelas;

- Figuras (que podem ser compostas por múltiplos gráficos) referentes ao diagnóstico do ajuste.
  - A depender da análise, figuras, quadros ou tabelas para outros tipos de resultados podem ser necessários. Alguns resultados (como medidas e testes de qualidade de ajuste) podem ser inseridos no próprio texto;
  - Todas as tabelas e figuras deverão ter títulos e numeração. Todos eles deverão ser mencionados, em algum momento, no texto, com a discussão dos respectivos resultados;
  - Os resultados deverão ser devidamente editados. Saídas cruas do R ou de qualquer outro software serão desconsideradas;
  - Não incluir códigos de programação dentro do texto! Se desejado, encaminhar os códigos para o professor por e-mail ou incluir em Apêndice.
6. **Conclusão**, fechamento do trabalho, com um apanhado geral do trabalho realizado, destacando limitações e possíveis investigações futuras.

É esperado que todos os itens abaixo sejam contemplados ao longo do processo de análise:

- a) Descrição das variáveis e do problema relacionado ao seu conjunto de dados, com clara definição da variável resposta e das variáveis explicativas de interesse.
- b) Investigação e análise da correlação linear entre a variável resposta e cada uma das variáveis explicativas.
- c) Escolha, aplicação e descrição de método de seleção de variáveis aplicado, o que idealmente vem acompanhado de uma análise de multicolinearidade. Análise de medidas de comparação entre modelos concorrentes e análise dos resultados dos testes de significância. Realizar e analisar o Teste da Falta de Ajuste caso ache réplicas no conjunto de variáveis explicativas.
- d) Investigar se faz sentido a inclusão de termos de interação entre as variáveis explicativas ou outros termos de ordem superior (regressão polinomial). Analisar necessidade/viabilidade de transformação nas variáveis explicativas.
- e) Interpretação dos parâmetros do modelo final escolhido e análise da porcentagem da variabilidade da variável resposta que é explicada pela(s) variável(is) explicativa(s) no seu modelo final.
- f) Fornecer estimativa para a variância  $\sigma^2$  do termo de erro do modelo.
- g) Escolha de um conjunto de valores para a(s) variável(is) explicativa(s) presentes no seu modelo final e realização de previsão aplicando estes valores no modelo ajustado. Gráfico com bandas de confiança e de predição caso o modelo final contenha apenas uma variável explicativa.
- h) Análise gráfica da adequação do modelo às suposições utilizando os resíduos do seu modelo final ajustado. Testes de hipóteses para agregar à análise gráfica dos resíduos. Caso alguma suposição não tenha sido satisfeita, comentar a respeito de possíveis transformações nos dados ou mudanças na definição do modelo que poderiam melhorar o ajuste. Verificar se as propostas de melhoria são efetivas.
- i) Analisar se existem observações discrepantes/influentes que, potencialmente, podem estar influenciando no ajuste obtido. Se for o caso, ajustar um modelo de regressão retirando tais observações e analisar os resultados obtidos.