

Classwork 07

Instruções:

1. Este Classwork (CW) poderá ser resolvido em duplas;
2. O CW precisará conter o(s) **nome(s) completo(s)** e a(s) **matrícula(s)** do(s) discente(s);
3. **Serão considerados apenas** os Classworks resolvidos por discentes presentes;
4. O CW deverá ser enviado via **Google Classroom até as 09h51 do dia 01/06/2026** em **um único arquivo** e em **PDF** (*Portable Document Format*);
5. **Não serão considerados** Classworks com péssima qualidade visual ou em desacordo com as instruções anteriores.

Conjunto de dados. *Tokusatsu* é o termo japonês para classificar séries de super-heróis. Esse gênero foi muito popular no Brasil entre o final da década de 1980 e os primeiros anos da década de 1990, destacando-se “O Fantástico Jaspion” (*Kyojuu Tokusou Juspion*). Apesar de ter perdido popularidade no país, essas séries ainda reverberam, batizando nomes de ruas (CEP 26.292-366), cervejas, produção de livros, mangás e são referenciadas no “mundo do entretenimento”, como canais sobre essa temática no **Youtube** (*Alphabet Inc.*), que vão desde debates à produção de **curtas metragens**. O banco de dados a ser analisado foi construído baseado em métricas dos videoclipes que fazem referências aos *Tokusatsu* exibidos nos anos 1980/90 postados pelo cantor Ricardo Cruz em seu **canal** no **Youtube**. O banco de dados contém as seguintes variáveis:

- *Posted* - data da postagem do vídeo;
- *Collected* - data da coleta dos dados;
- *Lifetime* - número de dias entre a coleta e a postagem dos vídeos;
- *Timespan* - tempo de duração do vídeo;
- *Length* - tempo de duração do vídeo em segundos;
- *Views* - número de visualizações do vídeo;

- *Likes* - número de curtidas por mil do vídeo;
- *Comments* - número de comentários do vídeo;
- *Viewssub* - Razão entre o número de visualizações e o número de inscritos no canal;
- *Likesviews* - Razão entre o número de curtidas e o número de visualizações;
- *Singers* - número de cantores ou pessoas creditadas no vídeo;
- *Studio* - se o videoclipe foi gravado em um estúdio;
- *Pandemic* - se o videoclipe foi postado durante a pandemia de Covid-19;
- *Seriejp* - título original da série referenciada no vídeo;
- *Seriebr* - título brasileiro da série referenciada no vídeo;
- *Label* - título da série encurtado;
- *Year* - ano de lançamento da série no Japão;
- *Category* - em qual categoria/subgênero a série se encaixa;
- *Metalhero* - se o videoclipe é da categoria/subgênero *metal hero*;
- *Broadcast* - canal de televisão brasileiro em que a série foi transmitida primeiramente;
- *Video* - título do vídeo no Youtube;
- *Link* - endereço eletrônico do vídeo.

Exercício. Seu objetivo é ajustar um modelo de regressão linear que explique o número de visualizações dos vídeos (*Views*), utilizando as demais como possíveis variáveis como explicativas. Ao selecionar as variáveis explicativas, verifique se elas não são derivadas diretamente da variável resposta (*Views*) ou calculadas a partir dela, pois isso pode introduzir dependência artificial no modelo e comprometer a interpretação dos resultados.

Tarefas.

1. Realize uma análise exploratória dos dados. Apresente gráficos e medidas descritivas que permitam compreender o comportamento das variáveis. Discuta possíveis padrões, relações entre as variáveis e a presença de valores atípicos.
2. Ajuste o modelo completo e avalie a significância estatística dos coeficientes (considere, por exemplo, um nível de significância de 5%). Caso existam variáveis não significativas, proponha um modelo reduzido, justificando a exclusão dessas variáveis.
3. Verifique a adequação do modelo ajustado por meio da análise de resíduos e de diagnósticos apropriados. Discuta a presença de possíveis problemas, como observações influentes ou padrões sistemáticos nos resíduos. Quando apropriado, proponha uma transformação da variável resposta que possa melhorar o ajuste do modelo.
4. Interprete os resultados do modelo final, incluindo os coeficientes estimados. Discuta o efeito de cada variável explicativa sobre a variável resposta, mantendo as demais variáveis constantes.