

Disciplina: Estatística Aplicada
Professor: Gilberto S. Matos

Período: 2021.2e

Entrega: em equipe formada por **até 3 Integrantes (Os mesmos do 2o Trabalho)**.

Prazo: Até às **23:59h do dia 27 de agosto (Sáb)** de 2022.

Aluno(a): Vinícius Sousa Azevedo Nota: _____

Aluno(a): Kilian Macedo Melcher Nota: _____

Aluno(a): Karen Anne Aciole Alves Nota: _____

Aluno(a): Vinícius Trindade Rocha Ribeiro Nota: _____

3º Estágio / Trabalho

Desenvolvimento de uma Pesquisa com Análise de Correlação e Regressão Linear Simples

- 1) Apresente/Descreva um conjunto de dados que contenha duas (2) variáveis do tipo quantitativa (numérica) cujo interesse é investigar sobre a existência de uma relação linear entre elas. Descreva o contexto ao qual a base de dados está inserida.

Para o trabalho, foi escolhida a base de dados do campeonato brasileiro desde 2003. Dados de 2003-03-29 até 2006-12-03 possuem somente as seguintes informações: data, horario, rodada, estadio, time_man, time_vis, gols_man, gols_vis pelo fato de que o site que fornece esses dados não disponibilizou outras informações. A partir de 2007-05-03, informações mais completas foram disponibilizadas, contudo com algumas exceções. Outro ponto é que tecnico_man e tecnico_vis possuem alguns valores vazios devido a mudança de padrão. Geralmente técnicos possui id = 0. Contudo quando o técnico é um ex-jogador, ele possui um id específico. Em consequência da inconsistência no padrão do id dos técnicos, linhas em que esse erro ocorre as colunas valor_equipe_titular_man, valor_equipe_titular_vis, idade_media_titular_man, idade_media_titular_vis, também são afetadas e resultando em vazio. Na data de 2016-12-11, rodada 38, a partida entre Chapecoense e Atlético-MG não

ocorreu em decorrência do desastre que atingiu a equipe do Chapecoense na época, resultando assim em W.O. duplo; Fonte: <https://www.transfermarkt.com/campeonato-brasileiro-serie-a/startseite/wettbewerb/BRA1>. Duas variáveis do tipo quantitativa da presente base que tenham uma potencial existência de relação linear entre si é o caso das variáveis chutes_man e gols_man

A tabela do campeonato brasileiro inclui dados desde 2003 com detalhes como número de gols, estádio, árbitro, técnicos, e mais.

2) A partir das duas (2) variáveis adotadas para análise:

- a) Desenvolva uma breve análise exploratória/descritiva das mesmas;
- b) Desenvolva e **interprete de forma prática** uma análise de correlação.
- c) Desenvolva e **interprete de forma prática** uma análise de regressão linear simples, incluindo a análise de resíduos e previsões para alguns valores estabelecidos para a variável independente, $X = x$.

DICAS / SUGESTÕES!

- a) **Anexe; separadamente; a base de dados** em *excel* ou *csv* ou disponibilize o link de acesso à base.
- b) Entregue o **relatório** da análise estatística **em arquivo pdf**.
- c) Procure sempre apresentar os **resultados** com um **olhar crítico** e ou com **insights** relevantes para o problema em estudo.
- d) Precisando de mais ajuda, entrem em contato por e-mail (gilberto.silva@professor.ufcg.edu.br), chat do PVAE ou estando presente nas aulas síncronas.

***Bons Estudos e Tragam suas Dúvidas
para
Discutirmos nas Aulas Síncronas!!!***