

## Estudo de Caso 5: Teste de Hipóteses I

Analisaremos a planilha “bundesliga.rds” extraída (e editada) a partir do banco de dados *kaggle*. A planilha contém informações sobre os jogos do campeonato de futebol masculino alemão de 2011 a 2021. As variáveis são a data das partidas, o nome do time da casa e de fora, a quantidade de gols marcados no primeiro tempo pelo time da casa e de fora, a quantidade de gols marcados no jogo todo pelo time da casa e de fora e o resultado da partida<sup>1</sup>. **Lembre-se de comentar e tirar conclusões dos exercícios realizados.**

### 1. Teste de Hipóteses

- a. Analise o desempenho de gols dentro e fora de casa da equipe Dortmund. Represente os conjuntos por uma análise de boxplot e compare os resultados. Além da comparação qualitativa e visual das caixas utilize as medidas de tendência central de cada um dos dados (utilize a função `summary` em cada um dos conjuntos para facilitar a obtenção de tais valores).

Um antigo jornalista, após anos de cobertura do campeonato alemão, afirma que o Dortmund **jogando em casa** possui uma média de **2.6 gols** marcados a favor. Tomaremos, em um primeiro momento, essa afirmação como “verdade atual”. Após derrotas recentes do time, há quem questione essa quantidade de gols, imaginando uma média **menor** que os 2.6 afirmados pelo jornalista. Podemos testar o “descrédito recente” nos gols em casa a partir de um teste de hipótese para média de gols (nível de significância de 5%) utilizando a amostra de jogos que dispomos.

- b. Construa o teste definindo as hipóteses nula e alternativa relativas à média de gols do clube jogando em casa.
- c. Escreva a equação analítica para a estatística utilizada no teste (valor crítico), considerando desvio padrão populacional conhecido.
- d. Calcule a estatística do teste para desvio padrão populacional igual a 0.5. Compare o valor crítico com o obtido na amostra e conclua o teste de hipótese.

---

<sup>1</sup> O resultado da partida é assinalado com A caso o resultado seja favorável ao time de fora (Away), H caso o resultado seja favorável ao time da casa (Home) ou D caso haja empate (Draw).

- e. Escreva a equação analítica para a estatística utilizada no teste (valor crítico), considerando desvio padrão populacional desconhecido.
- f. Calcule a estatística do teste para desvio padrão populacional desconhecido. Compare o valor crítico com o obtido na amostra e conclua o teste de hipótese.
- g. Compare os valores críticos para o caso de desvio padrão populacional conhecido e desconhecido. Qual dos dois apresenta maior incerteza para o teste? Por que?
- h. Ao final das análises, o que se pode dizer sobre a afirmação do jornalista?
- i. Considere nível de significância de 1%. Refaça o item 2 e compare os resultados.

### 3. Entrega no Moodle.

Os cases devem ser enviados no e-disciplinas em um arquivo .pdf com o script do R anexo ao final do próprio PDF, de forma a possibilitar o Ctrl c, Ctrl v do mesmo para efeitos de correção.

Lembre-se de que dissertações e conclusões acerca dos resultados são mais importantes que a própria construção do código em R. Indique todos os resultados da maneira mais expositiva possível.

O prazo de entrega é domingo, 14/05, às 23h59.