3rd Quiz ADA

4th Trimester

António Fidalgo   Miguel Salema

May 17, 2024

Instructions

- You have 70 minutes to complete the quiz.

- This is an individual exam.

- You may not use the internet. You must have your Wi-Fi disconnected.

- All relevant R code and written answers must be answered in a .qmd file and then rendered to pdf. In case you cannot render to pdf, you may render it to html, but you will have a 10 points discount.

- Upload the full directory of your assignment to Moodle by the indicated deadline as a .zip file. The .zip file should include the .qmd, the .pdf/.html and the data.

- Your code must run on the grader's computer.

- The detection of any form of plagiarism in your work will result in a grade 0.

- You must separate all your questions and sub-questions using headers in the .qmd document.

- You can access all class materials but internet access is forbidden.

Benfica's Data

To answer this quiz, you must load the file `df_benfica.csv` from Moodle. Benfica is a professional football club based in Lisbon that competes in the "Primeira Liga", which is the top-tier league in Portugal. The data contains all games played by Benfica in that league from the season 2000/01 to 2018/19. Each observation is a game. Table 1 contains the description of the variables. **Assume that you know nothing else about Benfica's past performance and draw no conclusions that cannot be drawn from this data alone**.

Table 1: Description of the Variables

Variable	Description
season	Starting year of the season; e.g: season 20/21 will show as 20.
home	Name of the team playing at home.
away	Name of the team playing away.
win	Which team won the game (H-home; D-draw; A-away)
goal_home	How many goals were scored by the home team.
goal_away	How many goals were scored by the away team.
date	Day when the game was played.
goal_diff	Goal difference for Benfica in each game, which is the goals scored by Benfica minus the goals scored by the other team.

Questions

1. (20 points) Read the data into `R` and name the object `df`.

2. (35 points) Create a sample named `df_09_10` containing only the seasons of 2009/10 and 2010/11.

3. (60 points) In this exercise, you'll analyse the average goal differences of Benfica over time.

   a. (30 points) Create a dataset named `df_av_goal` that contains the average goal difference of Benfica by each season.

   b. (30 points) Use `df_av_goal` to draw a line plot, showing the evolution of the average goal difference of Benfica per season. Your plot must also contain the points, not only the lines. Write the code even if you didn't manage to do exercise 3.a). We did it and it looks like this, but you don't have to perfectly mimic our plot:
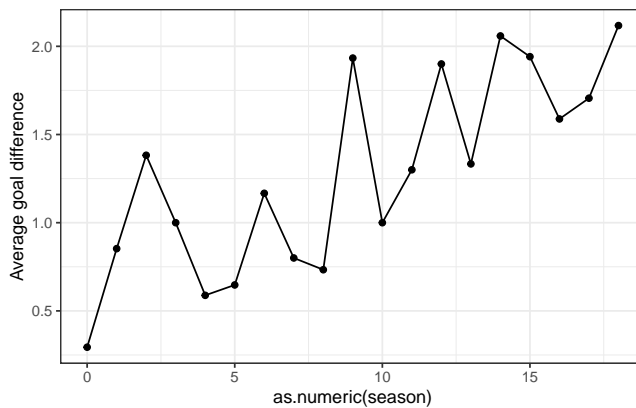
4. (35 points) Copy the following sentence into your quarto file and, using the dataset `df`, fill the "xxx" using inline `R` code: "Benfica has an average goal difference of xxx in the games played against Porto." (Hint: remember that Benfica can play against Porto away and at home.)

5. (50 points) Jorge Jesus was Benfica's coach in the 2009/10 and 2010/11 seasons. You have a friend who said the following in 2011:

   > Benfica performed so poorly this year. It seemed like Jorge Jesus was no longer the coach! He must be fired. It's clear from the goal difference. Last year's average goal difference is bigger than this year's by almost 1 goal!

   Do you think your friend is correct in drawing such strong conclusions from the goal differences? (hint: use the `lm` function and the dataset `df_09_10`.)