

# INTRODUCTION TO SPORT ANALYTICS (20630)

Exam Mock

Allowed time 90 minutes

The dataset **Teams\_overall2023.csv** extends the one used in the lectures by including the most recent completed NBA seasons.

Using this database generate a .R or .Rmd file containing all the relevant answers. Please name this file with your surname and the appropriate extension (.R or .Rmd)

**Q1 (3 points)** Clear the environment, set the correct default directory, import the data in R, install and run all the relevant packages for the programme.

**Q2 (3 points)** Provide two tables, the first one showing the number of season played by each team, the second one showing the number of teams playing in each season.

**Q3 (3 points)** Provide a time-series plot of the number of games played in each season by the Atlanta Hawks over the period 1980-2022

**Q4 (3 points)** Provide a Table with the ranking of Teams based on their performance in season 2022. Use a measure of performance WP, i.e. the ratio of wins to games played

**Q5 (3 points)** Using all the available data for all seasons provide a cross-plot of the winning percentage (the variable you have created) and the average age of teams ( a variable available in the database and called AGE)

**Q6 (6 points)** Exploiting the graphical evidence derived in the answer to question 5 specify a model to explain the relationship between AGE and winning percentage, then

1. estimate all unknown parameters
2. comment on the significance of single coefficients and on the overall performance of the regression
3. provide a plot that add to the cross-plot from Q5 the fitted values from your model
4. indicate the model predicted winning percentage of two hypothetical teams with AGE of 20 and 40 years.

**Q7 (8 points)** By referring to the course material create three variables measuring "Employed Possessions"(empl\_poss), Field Goal Attempted differences (FGAD) and "Acquired Possessions"(acq\_poss) Create a new database that includes variables FGAD, OFT and FTA from season 2000 to season 2010. Run a regression of FGAD on a constant OFT and FTA and test the hypothesis that the coefficient on OFT is 0.45 and the coefficient of FTA is -0.45. Provide a cross-plot of the variable *TEAM\_R* on the sum of the regression fitted constant and residuals. Briefly comment on your results.

**Q8 (4 points)** Propose and briefly discuss measure of "Competitive Balance" in a sport league.