

Homework 1 (Statistical Regression & R)

Instructions:

- Submit your answer to the task below as a clean report of no more than four pages.
Warning: Only the first four pages will be marked if you submit more than four pages.
- Submit a separate R file containing your syntax/code (the tutors will check your answers by running your R syntax file).

Task:

An institute of demographic research has hired you to analyse and compare aggregated death rates by age. You are provided with the dataset file “death_rates.csv”. This dataset contains three variables: Age (in years), Sex (M for males and F for females), and DeathRates (number of deaths in 10,000).

- 1) Undertake a descriptive analysis of the variables in this dataset. Feel free to use appropriate graphics to spice your description.
- 2) Explore (and comment on) the association between Death rates and Age, separately for males and females.
- 3) Build a simple linear regression model to explain the death rates of females as a function of age.
Note: Do not use the lm/glm functions in R.
- 4) Use the estimated parameters of your model in 3) to predict the death rate of females aged 51.

- 5) Comment on the overall quality of your model in 3) and suggest ways to improve it.