

Deadline 15 February 2021**NAME:**

1. (a) Using function `dataGen()` from library `rospca`, simulate one dataset with 100 observations containing 9 variables. The data generating function must be specified such that the first three variables have same standard deviation equal to 20, the next three variables same standard deviation equal to 5 and the last three variables same standard deviation equal to 10; the first three variables have inner correlations equal to 0.7, the next three variables inner correlations equal to 0.9 and the last three variables inner correlations equal to 0.8. The remaining arguments are left as a default. [3]
- (b) Check that the estimated correlations and standard deviations are close to those of the data generating function. [2]
- (c) Simulate a dataset with the same settings as reported in 1. with the only difference that all inner correlations are set to zero. [2]
- (d) Perform principal component analysis on both datasets. Comment on the results. [3]

2. Dataset UNSY97.txt available on Moodle contains the values of five economic and demographic indicators for a sample of 25 countries (Albania, Argentina, Australia, Austria, Benin, Bolivia, Brazil, Cambodia, China, Colombia, Croatia, El Salvador, France, Greece, Guatemala, Iran, Italy, Malawi, Netherlands, Pakistan, Papua New Guinea, Peru, Romania, USA, Zimbabwe). The data refer to 1990 and they come from the United Nations Statistical Yearbook of 1997. The indicators are annual percentage population growth rate (*Increase*), life expectancy in years (*Life*), infant mortality rate per 1000 (*IMR*), total fertility rate (*TFR*), and Gross Domestic Product per capita in US dollars (*GDP*). The data have been recorded following the alphabetical ordering.

Perform MDS with 1 dimension and 2 dimensions on this dataset and write down your conclusions (no more than 250 words excluding R codes, R output etc.). [10]

3. Read the paper "The Online Dating Intensity Scale: Exploratory Factor Analysis in a Sample of Emerging Adults" by Bloom & Taylor (2020) available on Moodle. Write a report (not more than 3 pages) summarising the goals of the work, the data source, the methods used, the results of the analysis and the conclusions. Also, comment on the robustness and generality of the results, and limitations of the analysis. [Marks: 10]