Statistical Methods II: Week 7 Assignment (due 10-11 March, 2025)

1. The data set `worldpop` available in the R package `lmreg` contains data on the midyear population of the world (in billions) for the years 1981-2000. Assuming that the regression of the world population (`Pop.billion`) on the year variable (`Year`) is a continuous and piecewise linear function (with a single change point at the year 1990), fit a suitable multiple linear regression model through an R code, and identify from the output the following: (a) the estimated regression coefficients, (b) the estimated standard deviation of the error and (c) the value of R-square.

2. Continuing with the data of Question 1, assume that the regression of `Pop.billion` on `Year` is a continuous and piecewise quadratic function, with a single change point at the year 1990. Specify a multiple linear regression model that captures the above characteristics, clearly identifying the variables and the regression coefficients.

3. Continuing with the data of Question 1, fit the model of Question 2, identify the estimated regression coefficients and interpret them.

4. Identify (a) the estimated standard deviation of the error and (b) the value of R-square in the fit of Question 3, compare with the corresponding values in Question 1 and comment.

5. For which values of the parameters of Question 2 will you conclude that the piecewise linear model is adequate?

6. Modify the answer to Question 2 with the additional assumption that the derivative of the piecewise quadratic function is continuous at the change point.

7. Continuing with the data of Question 1, fit the model of Question 6, identify the estimated regression coefficients and interpret them.

8. Identify (a) the estimated standard deviation of the error and (b) the value of R-square in the fit of Question 7, compare with the corresponding values in Questions 1 and 4 and comment.

9. For which values of the parameters of Question 2 will you conclude that the derivative of the piecewise quadratic function is continuous at the change point?

10. The data set `leprosy` available in the R package `lmreg` (taken from Senedecor and Cochran, 1967) contains pre- and post-treatment scores on abundance of leprosy for patients receiving treatments A, D or F (Placebo). Fit a regression model for the post-treatment score, with pre-treatment score as an explanatory variable and the treatment regarded as a 'factor' with three levels. According to the fitted model, what is the expected post-treatment score of a patient who has pre-treatment score x and belongs to treatment group (a) A, (b) D or (c) F? [Three different answers are asked for.]

11. For the data set of Question 10, replace the 'factor' treatment by a pair of binary variables that are indicators of groups A and D, and fit this modified model. According to the fitted model, what is the expected post-treatment score of a patient who has pre-treatment score x and belongs to treatment group (a) A, (b) D or (c) F? [Three different answers are asked for.]

12. How will the answers to Question 11 change if the binary variables are chosen as indicators of groups A and F?