

Please work today in mini-groups of two. We will be considering three data sets in this session, which will serve as a basis for your group work on Thursday and Friday.

Task 3.1: (The Age/Income data)

- (a) Load the `age.income` data from R package **SemiPar** and read the help file.
- (b) Plot the data and fit any nonparametric smoother.
- (c) Use SiZer methodology to answer the question: *Is the dip at age ≈ 45 really there?*

Task 3.2: (The Galaxy data)

- (a) Load the `galaxy` data from R package **ElemStatLearn** and read the help file
- (b) Visualize the data set, using tools like `plot` or `pairs`. Hint: For the best 3-dimensional appearance, plot `velocity` against the *negative* (not spherical) coordinate axes using R package **scatterplot3d**.
- (c) Try to fit simple one-dimensional smoothers to this data set.
- (d) Look at the relation between `north.south` and `velocity`. Does the fit improve using a median smoother (see course web page)?

Task 3.3: (The Fetal data)

- (a) Load the *fetal* dataset via

```
fetal <- read.table('T:MATHS/DMAOJE/fetal.dta', header=TRUE, sep=' ', ' ')
```

(the '0' in DMAOJE it is a zero!) and read the documentation (Handout 2) provided on the course web page . Using `na.omit(...)`, remove all days from the data set which contain one or more missing values.
- (b) Does the mortality NATMOR vary over time (DAY)? To answer this question informally, create the corresponding scatterplot, and fit a nonparametric smoother, taking the response distribution into account. The result should look like Figure 1 (left) in the Documentation. *Hint:* Use `locfit` or `gam` and specify `family = 'poisson'`.
- (c) Use the function `factor()` to generate indicator variables for MONTH and WEEK (e.g. `month <- factor(MONTH)`). Then fit the "CORE"-Model as specified on Handout 2 (something like `gam(NATMOR~month+week+ TEMP+UMID, family=..., data=...)`). Visualize and give an interpretation of the result.

Task 3.4: (Further procedure)

- (a) At the end of today's session (around 3.45pm), we will hand out new sheets. Your mini-group will then merge with another mini-group and you will work in **groups of four** from then on. You new group will continue to work on **only one** of the data sets from Sheets 2 or 3 **for which you decide in your group** right now or tomorrow morning. **Each data set can only be worked on by at most two groups** on a first-come first-serve basis. Please confirm with a tutor that your project is still available!

- (b) **Please print, note, or save your key results/plots of this sheet. They will be useful for discussion of the results in your group, and may feed into your project presentation!**