Statistical Inference — Lab 4

Date of issue: December, 6^{th} 2018 Due Date: December, 20^{th} 2018, 6:00 pm

Family Name: _____

First Name: _____

Student ID: _____

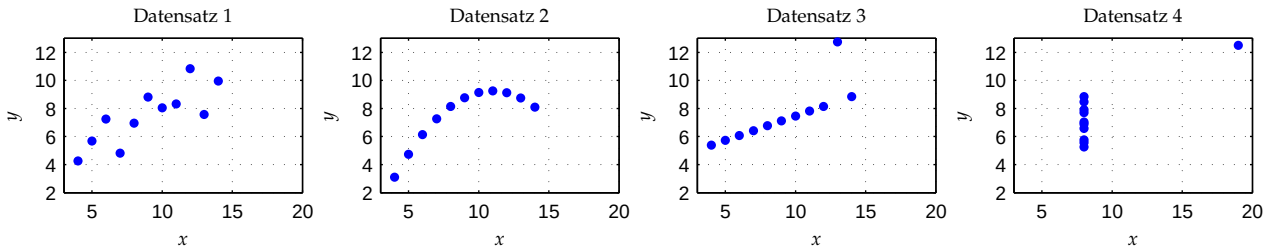
Linear and polynomial regression

Four different data sets (`AnscombeQuartet.mat` or the data file `AnscombeQuartet.dat`, both available on Ilias → Statistical Inference → Lab 4) plotted below in Fig.1 are to be fitted in a least-squares sense (A-model) to a linear function $y = a + b\,x$ and a parabola $y = a + b\,x + c\,x^2$. Such modellings are called respectively *simple linear regression* and *polynomial regression.* (Please note that the use of built-in MATLAB functions for the regression is not allowed!)

In this respect, the following quantities, if they exist, must be determined and reported for each case:

- the design matrix A
- The normal matrix N and its inverse N^{-1}
- the estimator $\hat{a}$, $\hat{b}$ resp. $\hat{a}$, $\hat{b}$, $\hat{c}$ of the regression functions
- the sum of squared residuals $\hat{e}^\mathsf{T}\hat{e}$, the number of degree of freedom (or redundancy r) and the normalised sum of squared residuals $\frac{\hat{e}^\mathsf{T}\hat{e}}{r}$.

All the results must be critically analysed, compared and when required, fully explained. In particular, a plot containing the initial data set and the curve of the regression functions must be made for each scenario.



F. J. Anscombe (1973): Graphs in Statistical Analysis. The American Statistician 27: 17–21