Computer Workshop 2

- You are expected to come prepared to the computer workshop. This means that you have to make sure that you understand the theoretical concepts behind the questions.
- Solutions will NOT be published on Blackboard. You are expected to raise your questions in the workshops.

This computer session uses `clothing.csv` which contains annual sales data and other characteristics of 400 Dutch men's fashion stores in 1990. The variables are:

Variable	Description
tsales	Annual sales in Dutch guilders
sales	sales per square meter
margin	Gross-profit-margin
nown	Number of owners (managers)
nfull	Number of full-timers
npart	Number of part-timers
naux	Number of helpers (temporary workers)
hoursw	Total number of hours worked
hourspw	Number of hours worked per worker
inv1	Investment in shop-premises
inv2	Investment in automation.
ssize	Sales floor space of the store (in m2).
start	year start of business

(a) Plot a histogram for tsales. What are the mean and the median? Plot tsales against ssize.

(b) Redo a) for sales.

(c) Regress sales on ssize. Interpret.

(d) Regress sales on ssize and ssize squared. Interpret. Is there evidence for a nonlinear relationship? If yes, what type of extremum do you find?

(e) Regress sales on nown, nfull, npart, naux, inv1, inv2, ssize and ssize squared.

(i) Interpret your results.

(ii) Is the regression significant?

(iii) Test whether $\beta_{\text{inv1}} = 0$.

(iv) Test whether $\beta_{\text{nown}} = 1000$.

(v) Test whether $\beta_{\text{nfull}} = 2\beta_{\text{npart}}$.

(vi) Use a Chow test to see whether the relationship is the same for stores with start ≤ 40 and start > 40.

(f) Plot the squared residuals from the original regression in (e) against the explanatory variables. Do you find evidence of heteroskedasticity? How could you test for heteroskedasticity using a regression?