Terminology and Models

Terminology

Our book tries to emphasize applications of our material, which are called here "authentic situations" (any other book would call them "applications"). The terminology is harmless, but you should be aware that most examples, even if using real data, are not treated as they would in a truly authentic situation, but, rather, are used as raw material for whatever the topic of the book section is at that point. Many of the questions would never be asked in that form in a true application situation.

Data points are often presented graphically, as most humans are visual, and get a better handle on things if they are represented by a picture. Plotting data points in a graph results in what is usually called a "scatterplot", but our book uses the much less common term of "scattergraph". This is not a problem, but be aware of the fact that most any other book you may look at will use the common terminology.

What Are Models?

The book refers repeatedly to models as tools for handling data. These are mathematical expressions used to summarize and, more generally, analyze application data. We should be aware that this can be done in different ways, depending on the specific problem, with different results.

- One way (which is the stricter version of the term "model") is to rely on a theory, or at least on some theoretical insight, to choose a functional form for the modeling function – linear, quadratic, exponential, and so on. The data will be used to choose a specific numerical form for our model that matches the data as closely as possible. It is not realistic to expect that observation data will fit our model exactly. Nature can often be well approximated by simple models like the ones we will look at, but there will be many small, but real, effects which will cause real data to deviate from our simple model. If the deviation is relatively small, the simplicity of our tool more than makes up for it, and predictions, while not completely exact, should be close enough to the actual outcome to be worthwhile.

- Sometimes we don't have a theory that can help us to come up with an efficient model, and will fall back on writing an expression that matches reasonably well the data at hand, without pretending to point to a deeper structure of the phenomenon we are studying. These formulas have practically no validity outside the range of data under scrutiny and do not allow for real predictions. For example, plotting the stock price of a company in time will not allow for a theoretical model, as the dynamics of stock prices is extremely difficult to model (and, in any case, while people have worked out generic models for stock behavior, none is expected to be useful when looking at a particular stock). It goes without saying that trying to predict the future price of a stock using a "model" obtained by approximating the price over a fixed amount of time is not a recommended practice.

Examples of the first kind are many physical laws, like *Hooke's Law* (describing with a linear function the elongation of an elastic spring when pushed or pulled by a force), *Ohm's Law* (describing, again with a linear function, the relation between voltage and current in an electrical circuit), and the law of the motion of an object thrown in the presence of gravity (as long as we are close enough to the Earth to treat gravity as a constant force), and in a vacuum (the presence of air complicates the description considerably), which turns out to be described by a quadratic function.

Examples of the second kind are most of the examples of polynomial (usually quadratic) models you will see in the book and in the assignments, as they simply try to fit data points displaying, for example, a turn around with no hope of a reasonable expectation of a *theoretical* quadratic model. For example, several quadratic models suggested in the book (women in the workforce, investment in security by the Government, Profits of a company) are for practice only, and using them as predictors for the future (or the past) would lead to completely unreliable results.