Chapter 6
Epilog

Now that you have nearly made it through the whole book, let me give you a little food for further thought and some additional ideas on the way. Ironically, some of these will probably shake up a bit what you have learnt so far, but I hope they will also stimulate some curiosity for what else is out there to discover and explore.

One thing to point out again here is that especially the sections on (generalized) linear models (ANOVAs and regressions) are very short. For example, we have not talked about count/Poisson regressions. Also, we skipped the issue of repeated measures: we did make a difference between a *t*-test for independent samples and a *t*-test for dependent samples, but have not done the same for ANOVAs. We have not dealt with the difference between fixed effects and random effects. Methods such as mixed-effects / multi-level models, which can handle such issues in fascinating ways, are currently hot in linguistics and I pointed out some references for further study above.

Another interesting topic to pursue is that of (cross) validation. Very often, results can be validated by splitting up the existing sample into two or more parts and then apply the relevant statistical methods to these parts to determine whether you obtain comparable results. Or, you could apply a regression to one half of a sample and then check how well the regression coefficients work when applied to the other half. Such methods can reveal a lot about the internal structure of a data set and there are several functions available in R for these methods. A related point is that, given the ever increasing power of computers, resampling and permutation approaches become more and more popular; examples include the bootstrap, the jackknife procedure, or exhaustive permutation procedures. These procedures are non-parametric methods you can use to estimate means, variances, but also correlations or regression parameters without major distributional assumptions. Such methods are not the solution to all statistical problems, but can still be interesting and powerful tools (cf. the libraries `boot` as well as `bootstrap`).

Recommendation(s) for further study
Good (2005), Rizzo (2008: Ch. 7, 8)

It is also worth pointing out that R has many many more possibilities of graphical representation than I could mention here. I only used the traditional graphics system, but there are other more powerful tools, which are available from the libraries `lattice` and `ggplot`. The website <http://addictedtor.free.fr/graphiques/> provides many very interesting and impressive examples for R plots, and several good books illustrate many of the exciting possibilities for exploration (cf. Unwin, Theus, and Hofmann 2006, Cook and Swayne 2007, and Sarkar 2008)

Finally, note that the null hypothesis testing paradigm that is underlying most of the methods discussed here is not as uncontroversial as this textbook (and most others) may make you believe. While the computation of *p*-values is certainly still the standard approach, there are researchers who argue for a different perspective. Some of these argue that *p*-values are problematic because they do in fact not represent the conditional probability that one is really interested in. Recall, the above *p*-values answer the question "How likely is it to get the observed data when H_0 is true?" but what one actually wants to know "How likely is H_1 given the data I have?" Suggestions for improvement include:

- one should focus not on *p*-values but on effect sizes and/or confidence intervals (which is why I mentioned these above again and again);
- one should report so-called p_{rep}-values, which according to Killeen (2005) provide the probability to replicate an observed effect (but are not uncontroversial themselves);
- one should test reasonable null hypotheses rather than hypotheses that could never be true in the first place (there will always be some effect or difference).

In addition, some argue against different significance steps of the kind "significant" vs. "very significant" vs. "highly significant" by saying that either a result is significant or not but that, once the null hypothesis gets rejected, no further distinctions are necessary. I find it hard to get excited about this latter kind of debate. Many R functions provide different numbers of asterisks so that one can use them, and as long as one is aware that "very significant" is part of a particular conventionalized terminology, I see no reason to make this a bone of contention.

Another interesting approach is the so-called Bayesian approach to statistics, which allows to include subjective prior knowledge or previous results with one's own data. All of these things are worth exploring.

Recommendation(s) for further study
– Cohen (1994), Loftus (1996), and Denis (2003) for discussion of the null hypothesis testing paradigm
– Killeen (2005) on p_{rep}-values
– Iversen (1984) on Bayes statistics

I hope you can use the techniques covered in this book for many different questions, and when this little epilog also makes you try and extend your knowledge and familiarize yourself with additional tools and methods – for example, there are many great web resources, one of my favorites is <http://www.statmethods.net/index.html> – then this book has achieved one of his main objectives.