

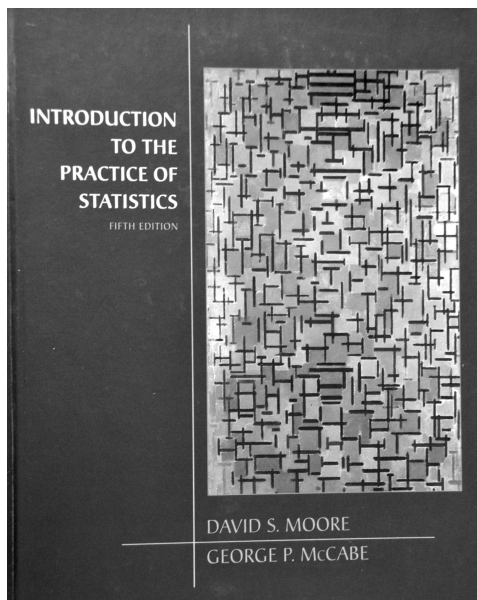
The Typesetting of Statistics

Abstract

The Dutch translation of the 750 page textbook 'Introduction to the Practice of Statistics' is typeset using a set of ConT_EXt macros. This article gives a short impression of the production process of this book, showing that the use of T_EX for the actual typesetting was perhaps the least cumbersome part of the process.

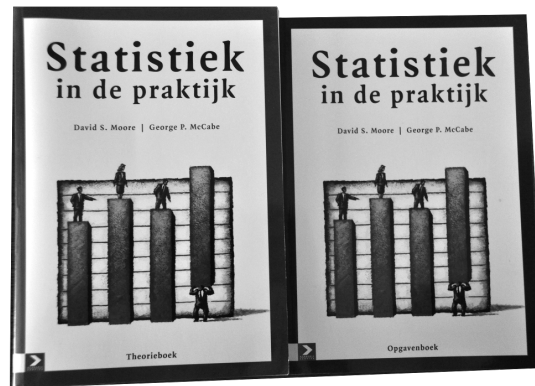
The original version

'Introduction to the Practice of Statistics' is a fairly hefty textbook in what I consider a typical 'United States' style: it is full-color throughout, in a fairly large format that combines theory and practice, bound in a single hardcover book, and is printed on rather thin paper. It has been typeset using L^AT_EX.



The Dutch version

'Statistiek in de praktijk' in turn is a fairly typical adaptation to the Dutch market: The body is printed in grayscale, it has a slightly smaller format, the original is split into two paperbacks (a textbook and a workbook) and it is printed on standard thickness paper. It is typeset using ConT_EXt.



History of the Dutch version

The first edition was printed elsewhere, and the original Dutch sources that I received used macroT_EX by Amy Hendrickson. This source was converted to plain T_EX for the second and third editions, using a bunch of special purpose extension macros.

During this time, there was one T_EX source file for each chapter body, one for each chapter title page, and one for each chapter preface.

The Dutch publisher skipped the fourth edition to save on expenses.

For the fifth edition, I switched to ConT_EXt for easier use of graphic objects and references. In the process, the source file layout was converted to the standard ConT_EXt model for projects so that a single texexec call can be used to generate each of the books. However, makeindex is still being used for the index generation instead of the normal ConT_EXt methods.

The fifth edition

The Dutch translation of the fifth edition was sent to us as a bunch of Microsoft Word files (one per original chapter), and we were lucky enough to also receive the original sources of the English version.

The Word data format itself was fine with us, but actual content posed a number of problems:

- There were quite a lot of vocabulary differences between the chapters because four different translators had been working simultaneously.

- Parts of the translation were missing: anything in the fifth edition that was identical to the text in the third edition was not translated. We were expected to copy these blocks over from the Dutch sources of the third edition.
- There were accidental omissions in the translation.
- The graphics also needed to be translated, converted to grayscale, and corrected in Adobe Illustrator, then exported as EPS, and finally converted to PDF.
- The page location remarks in the translation files referred to the Dutch version of the third edition, instead of to the actual page numbers in the English fifth edition.

The last item above was by far the most problematic, because after the first proofs came back, it became necessary to have five pretty large piles of paper on the desk at all times:

1. The printout of the original Word files (because of the meta data).
2. The Dutch third edition (for continuous reference).
3. The English fifth edition (also for reference).
4. The editor-corrected printout of the Dutch fifth edition proofs (for correction).
5. A stack of printouts of various email messages and documents with corrections and addenda sent in by the translators.

Handling Word files

For the main text, the Word files were exported to HTML in OpenOffice, followed by a cleanup script written by Siep Kroonenberg.

The tabular material was imported from the English L^AT_EX sources using a dedicated ConT_EXt module that implements the L^AT_EX tabular environment. Using that, much of the tabular data could be copied straight over. The exceptions were tables containing decimal points (those were converted to commas) and tables dealing with money (where all amounts had to be converted into euros).

Mathematical formulas were re-keyed. This was not a big problem, as almost all formulas needed editing for localization anyway.

Taco Hoekwater
Elvenkind BV
taco (at) elvenkind dot com