

Guidelines for Writing a Bachelor or Master Thesis

Research Group Methods for Big Data
Scientific Computing Center (SCC)
Karlsruhe Institute of Technology (KIT)

Contact: kleinlab@scc.kit.edu | <https://kleinlab-statml.github.io/index.html>

Qualifying requirements:

Bachelor:

- 1.) Exam pass in at least one course offered by the *Research Group Methods for Big Data*
- 2.) One satisfactory presentation and regular attendance in the chair's seminar.

Masters:

- 1.) Exam pass in at least two courses offered by the *Research Group Methods for Big Data* with a total of at least 7 ECTS. One of the modules must be a lecture.
- 2.) One satisfactory presentation and regular attendance in the chair's seminar.

Confirmation by the appropriate examination office is required for the modules passed.

Equivalent courses can qualify for meeting the requirements and needs to be discussed on an individual basis. Before officially registering for the thesis, a 2 to 4-page exposé with a tentative table of contents, schedule, and literature must be submitted.

Aim of the thesis:

The purpose of the final thesis is to show that you can appropriately solve a problem at hand using statistical methods. For a bachelor thesis, this typically means answering a research question using data-analytic methods. For a master thesis, the problem should be more complex, involving, for example, a method extension, a method implementation, a simulation to compare the performance of methods, and/or advanced methods that require independent literature research.

The evaluation of the thesis is based on the following aspects:

- 1.) *Form*: Have the formatting instructions been observed? Is the thesis well written and understandable? Are all figures and tables meaningful and correctly labelled? Is all notation introduced and consistent? Are there any typing mistakes, grammatical errors, or incomplete sentences? Are used references correctly cited and word-for-word citations correctly marked as such?
- 2.) *Statistics*: Are all used statistical methods clearly and correctly explained in the author's own words? Are the statistical methods appropriate for the problem at hand? If simulations are used, are these well conducted, clearly described and the results well summarised? If methods are extended, are the extensions reasonable, well-motivated and fitting to the problem?
- 3.) *Application*: Is the data well described and are the problems of interest clearly stated? Are the problems adequately translated into statistical questions? Have the problems presented been adequately answered? Is the interpretation of the results correct, complete and understandable?

Requirements of content and form of the thesis:

- 1.) The work should present the author's competence in the subject and their ability to independently handle and process statistical data with the help of statistical software (for example SPSS, Matlab, R, etc.).
- 2.) The thesis should include the following components:
 - Title page
 - Table of contents
 - List of abbreviations
 - List of figures & tables (if any)
 - List of formulae & symbols (if any)
 - Main body
 - Bibliography
 - Appendix
 - Declaration of Academic Honesty

3.) Formulae:

- If necessary, you should make required line breaks according to a formula generated by calculator and, where possible, not separate bracketed expressions.
- Intertemporal considerations require the indication of time indices for all dynamic variables.
- For complex parenthetical expressions, round, square and curly brackets may be used for better clarity of layers of presentation.
- The exact differentiation of symbols used must be guaranteed.

4.) Tables:

- Graphic design according to individual preferences.
- Correct factual, temporal, and spatial arrangement of the presented numbers.
- Table headings may contain subdividing blocks.
- Line by line breakdown by indent of subpoints up to a maximum of three steps.
- The first header element should also be used to identify a subdivision of the column elements; this can be divided diagonally and labelled accordingly.
- Tables should be comprehensible without accompanying text.
- If abbreviations, remarks, or sources are necessary, these should be reflected in corresponding footnotes and the words "derived from..." should be appended to the source. Complementary tables and figures, that are too large for presentation within the main body of text or that are not explicitly addressed in the text, should be moved to the appendices.

5.) The thesis should be written in LaTeX; use the defaults for your document class.

6.) The American referencing system for citations should be used.

7.) Important tables or figures should be integrated into the text (not into the Appendix). Each table or figure should not exceed one page.

8.) Any variation of the above regulations will only be permitted with prior approval.

9.) All sources, quotes, cited graphics or tables etc. should be clearly identified. The thesis will be checked for plagiarism, and if found, the work will be graded as "failed".

Guidelines for thesis submission:

- The duration of the thesis preparation (writing) period is defined by the current study regulations; the submission deadline must not be exceeded.
- Two printed (hardcopy) versions of the work must be submitted to the examination office (Prüfungsamt) before the deadline, with the title of the work and the author's name and student ID clearly visible on the cover (front) page.
- The work is also to be submitted to the Chair of Uncertainty Quantification and Statistical Learning in a pdf format by the author, thereby enabling the grading process to start as soon as possible.
- All analyses presented in the thesis must be reproducible. All code (including necessary explanations) is to be submitted electronically.