LaTeX Template for STAT 548 Qualifying Paper Report

Daniel J. McDonald

September 15, 2025

You can remove everything from here to the end of the document when you start writing your report.

But do read through it first to get an idea of how to use the style file and some best practices for writing the report.

1   Requirements

As described on my website https://dajmcdon.github.io/teaching/stat548.html, the report should contain three main sections:

1. **Summary (∼3 pages)** The first section of the report should provide a summary of the paper and the problem(s) it addresses, including its relationship to any previous work, its major contributions (e.g., novel techniques, algorithmic developments, problem formulations, theoretical contributions), and any limitations or shortcomings (e.g., restrictive assumptions, computational constraints, flawed methodology). The aim of this section is for you to synthesize the findings of a body of work and clearly present the important points.

2. **Mini-proposal(s) for research projects** Each proposal should describe a research project that applies, extends, generalizes, adapts, or addresses shortcomings of the QP. Seemingly unrelated ideas inspired by the original QP are also fine. You may write more than one proposal, but you must write at least one. A proposal should concisely describe: the primary problem to be addressed; an approach (or multiple approaches) for addressing the problem; any technical or conceptual sub-problems; the potential impact of the project. You are not expected to pursue any of these projects (though we can talk more if you would like to). The aim of this section is to get you thinking creatively about research, and to begin developing the skills necessary for writing research proposals. Each proposal should be no more than **2 pages max**.

3. **QP specific project results** Each potential QP listed below has a brief description of a related project. We will discuss the project in detail in our initial meeting, and we can meet again (as many times as necessary) before the report due date. Your grade will not be affected by how good the results look, whether your approach improves on past work, or whether you achieve the initial goal of the project. I will use this project to evaluate your research potential, which includes (among other aspects):

- clearly formulating a research question;
- setting up a useful mathematical framework for the problem;
- thinking creatively and independently to develop a solution;
- relating the problem to existing work, in other fields if necessary;
- being resourceful and asking questions when necessary;
- learning from and moving past the inevitable setbacks;
- reformulating the research problem when necessary;
- implementing new methods in code (when applicable);
- choosing appropriate experiments and metrics;
- communicating and reflecting on progress, setbacks, and results;
- thinking of future research directions.

The report should be submitted as a GitHub repository based on the template here. The template includes a LaTeX style file that should be used for the report. (Detailed instructions for usage can be found in the repository's README file.) Any experimental/numerical results should be reproducible. All code should be reusable, clearly commented/documented, and exist in the `src/` folder of the same GitHub repository to which you give me access as a collaborator. Code can be in any language you wish, though my strong preference is for `R` or `python`.

The remainder of this document demonstrates the style file and provides some best practices for writing the report.

2 Introduction

We write some math for fun:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,|\Sigma|^{n/2}} \exp\left\{-\frac{1}{2}(y-\mu)^\mathsf{T}\Sigma^{-1}(y-\mu)\right\}\mathsf{d}y = 1. \tag{1}$$

We encourage the use of the various macros defined in ShorTeX, so do your best. It makes things easier to read, but also provides lots of necessary mathematics definitions that render nicely.

Be careful with things like KL divergence and conditional probability statements. I find that

$$\mathrm{KL}(q \parallel p)$$

looks much better than

$$\mathrm{KL}(q||p),$$

and I similarly prefer

$$Y \mid X \sim \mathcal{N}(X, \ \sigma^2) \quad \text{to} \quad Y|X \sim \mathcal{N}(X, \sigma^2).$$

Note that the reals are $\mathbb{R}^p$. There is also

$$\widehat{\beta} = \operatorname*{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{2n} \|\boldsymbol{y} - \boldsymbol{X}\beta\|_2^2 + \lambda \|\beta\|_1,$$

and (with automatic sizing of the norm),

$$\widehat{\beta} = \operatorname*{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{2n} \left\| \frac{y}{1} - X\beta \right\|_2^2 + \lambda \|\beta\|_1.$$

Note that the indicator function looks like $\mathbb{1}\{\cdot\}$, but I sometimes prefer $\mathbf{1}\{\cdot\}$.

2.1 Cleveref

We prefer to use cleveref to get nice references to things. For example, you can say that Eq. (1) was printed in Section 2. No need to write out things like "Section".

3 Some best practices

Some of these are taken from Jacob Bien. Note the use of the ShorTeX itemize environment style below.

- Section titles should be all title case or all sentence case. Don't mix and match.

- I prefer data set to dataset.

- I prefer data to be singular. There remains debate on this point. When you use the word "datum" in a sentence, then we can argue. Data is a mass noun, like "information". We don't say "How many data are enough?", we say "How much data is enough." Enough said.

- Terminology is lower case, unless it's a person's name: Nyström extension and lasso.

- Equations are parts of the sentence. Displayed equations almost always have a comma or period after. Very rarely is there a colon or comma *before* a displayed equation.

- Don't start sentences with math ("Σ is the covariance of $\boldsymbol{X}$.") or the name of a software package that's lowercase, e.g., "`glmnet` is my favourite software".

- Don't use contractions.

- No need to put dollar signs around numbers: 12 versus 12.

- DO put dollar signs around math: p not p.

Ingredient	Quantity
Fusili	100 g
Eggs	2
Salt	1 tsp
Guanciale	50 g
Pepper	$1/2$ tsp
Grated parmesan	$1/4$ c

Table 1: This is a nice looking table. It might make carbonara.

- Use $x \gg y$ not $x >> y$.

- Careful with parentheticals and references. Wrong: (see, e.g., Akaike (1973)). Right: (see, e.g., Akaike, 1973).

- Never use `eqnarray`, always use `align`. Note that ShorTeX makes `\[    \]` into an align environment, so you can just use that always.

- For editing purposes, it is much better if the text is hard-wrapped rather than soft wrapped.

- asdfasdfsadf $\alpha\beta\hat{\gamma}$

3.1 Tables

Table 1 is a nice looking table. Strive for these.

4 Discussion

We made amazing contributions to the world of musical fractal pasta (McDonald, 2017; Tibshirani, 2013). We use Natbib, so be sure to use (Stein, 1981) for parenthetical references. Or you can say, according to Hastie et al. (2009), we should strive to balance truth and lies.

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In *Proceedings of the 2nd International Symposium of Information Theory* (eds. B. N. Petrov and F. Csaki), 267–281.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Verlag.

McDonald, D. J. (2017) Minimax Density Estimation for Growing Dimension. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)* (eds. A. Singh and J. Zhu), vol. 54, 194–203. PMLR.

Stein, C. M. (1981) Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, **9**, 1135–1151.

Tibshirani, R. J. (2013) The lasso problem and uniqueness. *Electronic Journal of Statistics*, **7**, 1456–1490.