Summary

The issue of missing data is encountered regularly in practice, for example, in surveys or experiments. Missing data is a problem because in most types of analyses we require an entry for each variable. The most common remedy is the *complete-case* analysis but it has its limitations. As a result, many alternative methods have been proposed to deal with missing data. In this thesis, we review the workings of the *Gibbs sampler* in `R` package BAMD, and investigate an alternative method, *multiple imputation.* We explore the function `aregImpute()` of `R` package Hmisc, which utilises *multiple imputation* via *predictive mean matching* to handle missing data.

Scope of thesis

Firstly, we provide an overview of missing data and methods for handling it. In the `R` package BAMD, a linear model with missing entries in the design matrix is solved using a Gibbs sampler. However, it is widely known that Markov Chain Monte Carlo (MCMC) methods suffer from issues like convergence and long computation time. Hence we apply multiple imputation via predictive mean matching to see if it will perform better. This review presents simulation outcomes to illustrate which method is more desirable.

Author's contributions

The author wrote the `R` programming codes for setting up the simulations used in this project. The author also wrote a function `generateSNP_LD()` which can be called within the function `generateData()` to generate SNPs that are in LD.