

# Chapter 1

## Introduction

The objective of this master thesis is to implement the variable importance measure LMG (named after the authors Lindeman, Merenda, and Gold ([Grömping, 2007](#))) in linear models estimated with Bayesian methods. Bayesian methods have gained popularity because they allow to quantify the uncertainty about parameters and they allow to include prior information.

Regression models are popular in many applied research areas ([Nimon and Oswald, 2013](#)). These models provide a tool to find an association between a response variable and a set of explanatory variables. The explanatory variables are also called predictors or covariates. Regression parameters provide information to what extent the response variable is expected to change when one predictor changes by one unit, given all other predictors in the model remain the same. Being aware of this last remark is very important for the correct interpretation of the regression parameters. It shows that the parameter value of a predictor is dependent on the other predictors in the model.

Because predictors are often correlated to some degree to each other, it is obviously not an easy task to find the most important predictors in a model. The first question is, what is meant by the importance of a predictor? There is no easy answer to this question and it is depending on the research issue. [Grömping \(2015\)](#) concludes that there may never be a unique definition of variable importance. There exist different metrics to quantify the importance of predictors. These metrics focus on different aspects and with correlated predictors they lead to different conclusions. A summary of variable importance metrics can be found in [Grömping \(2015\)](#).

A distinction should be made between the importance of predictors in regression models that are used to predict future data and in regression models, applied to find an association between predictors and the response variable. In the first case, the aim is only to reduce the error between the predicted values and the observable values. The underlying association between predictors is of minor importance. In the second case, the focus is on the strength of the relationship between the predictors and the response variable. A predictor may explain little of the response variable, given two other correlated predictors are already included in a regression model. However, this predictor that is unimportant from the regression output may be the main cause of the other two predictors. Therefore, it may be the most important predictor of this regression model ([Grömping, 2007](#)).

The causal relationship between the variables is missing in standard regression models. Studying a predictor, given other variables are already included or using models that contain only the

predictor itself, provide only some parts of the bigger picture about the predictor in a model. Which are the most useful variable importance metrics is still an open debate. A convincing theoretical basis is still lacking for all of those metrics. [Grömping \(2015\)](#) recommends to use the existing best practices, until a more profound solution will be found. For variance (or generally goodness-of-fit) decomposition based importance, she recommends to use LMG enhanced with joint contributions or dominance analysis ([Grömping, 2015](#)).

# Bibliography

Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *American Statistician*, **61**, 139–147. [1](#)

Grömping, U. (2015). Variable importance in regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, **7**, 137–152. [1](#), [2](#)

Nimon, K. F. and Oswald, F. L. (2013). Understanding the Results of Multiple Linear Regression: Beyond Standardized Regression Coefficients. *Organizational Research Methods*, **00**, 1–25. [1](#)