The Journal of Open Source Software

VarReg: An R package for regression in the location, shape and scale

Kristy P Robledo[1] **and Ian C MArschner 2**[1]

1 NHMRC Clinical Trials Centre, University of Sydney

DOI: 0000/joss.0000

Software

- Review
- Repository
- Archive

Submitted:
Published:

Introduction

Variance regression models arise in a variety of contexts, including measurement error and variance heterogeneity in standard linear regression analysis. Such models are necessary when the variance of the outcome measure changes as a covariate changes. The use of a variance regression model in these contexts allows the modelling of the variance in terms of covariates, either because the variance itself is of interest, or to increase the precision in estimation of the mean. In some areas, such as biomarker analysis, it may be that the variance itself is of interest.

We can consider the variance regression model as

$$X_i \sim N\left(f(x_i), g(x_i)\right), \quad i = 1, 2, ..., n \tag{1}$$

where $g()$ and $h()$ are known functions of a covariate x_i. This can also be extended to

$$X_i \sim N\left(\sum f(x_{ij}), \sum g(z_{ij})\right), \quad i = 1, 2, ..., n \tag{2}$$

where we consider multiple covariates, and also that the covariates in the mean model are not necessarily the same covariates as the variance model.

Generally, the variance regression model (Aitkin, 1987; Smyth, 2002; Verbyla, 1993) uses a log-link model, $h(z) = \log(z)$, and thus the covariates affect the variance multiplicatively. This is primarily because covariate effects in variance heterogeneity models can be negative as well as positive, and the use of a multiplicative model ensures that the overall error variance remains non-negative. Since this is simply a computational convenience, it may be that additive variance heterogeneity is more appropriate in some contexts. Indeed, additive decomposition of the variance is standard in other contexts, such as variance components models and genome-wide association studies. An attractive property of additive variance models is that spline models are effectively additive models, so semi-parametric models are obtained with little additional effort.

The Journal of Open Source Software

General model

Variance regression

Censoring

Location, Shape and Scale

Computational method

`VarReg` Package

Input arguments and examples

Computational methods

Output and plots

Discussion

Acknowledgements

?None

References

Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using glim. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *36*(3), 332–339. doi:10.2307/2347792

Smyth, G. K. (2002). An efficient algorithm for reml in heteroscedastic regression. *Journal of Computational and Graphical Statistics*, *11*(4), 836–847. doi:10.1198/106186002871

Verbyla, A. P. (1993). Modelling variance heterogeneity: Residual maximum likelihood and diagnostics. *Journal of the Royal Statistical Society. Series B (Methodological)*, *55*(2), 493–508.