

# A Multivariate Approach to Modelling Lifestyle Risk Factors of Children Myopia in the US

Stephen Su, STATS 767

## 1 Introduction

### 1.1 Background

The association between lifestyle factors and the development and subsequent progression of myopia among children has been long discussed and researched within the academic area. Among them, the Orinda Longitudinal Study of Myopia conducted research on children myopia spanning over 10 years. The research produced data that are both useful in exploring the lifestyle risk factors of myopia and a valuable case study for building and testing multivariate data models and analysis. This paper attempts to preliminarily analyse the OLSM myopia dataset and produce a multivariate model.

### 1.2 The Data

The dataset is from [ggeop/Myopia-Study](#) (Papachristou, 2018), which is a subset from the original data collected in 1989-1990 and 2000-2001. The dataset consists of 618 observations and 17 variables. The main focus of the paper is around numeric, lifestyle-related (non-definitional) variables and the logical variable indicating the prevalence of myopia, which are represented by the following variables:

Variable_Name	Unit	Description
myopic	boolean	Myopia within the first five years of follow up
age	years	Age at first visit
sporthr	hours per week	Time spent engaging in sports/outdoor activities
readhr	hours per week	Time spent reading for pleasure
comphr	hours per week	Time spent playing video/computer games or working on the computer
studyhr	hours per week	Time spent reading or studying for school assignments
tvhr	hours per week	Time spent watching television

Note: “Non-definitional” means the variables do not serve as an optometrical reference to myopia.

### 1.3 Outputs and Deliverables

This paper and the project presentation only includes selective outputs serving as the final deliverable, including a **GGally** plot (Schloerke et al., 2020) and model outputs from **base R** (R Core Team, 2021), the **MASS** (Venables & Ripley, 2002) and **mixOmics** (F et al., 2017) packages. Detailed model building, intermediary models, and materials, such as R code, required to reproduce this paper can be found at the Github repository [szmsu2011/stats767proj](#).

## 2 Exploratory Analysis

A preliminary visualisation of the selected data suggests heavy right-skewness except for the numeric variable **age**. As a convention, a log-transformation to all numeric variables except for **age** attempts to mitigate the skewness, yet the minima of the variables are zero. Instead, the **log1p** transformation is applied across the variables. Nevertheless, a subsequent plot of the transformed data indicates the **log1p** transformation seems to over-correct the skewness of variables **sporthr** and **tvhr**. Therefore, a final decision is made to transform variables **readhr**, **comphr** and **studyhr** by  $x \rightarrow \log(1 + x)$  and **sporthr** and **tvhr** by  $x \rightarrow \sqrt{x}$ .

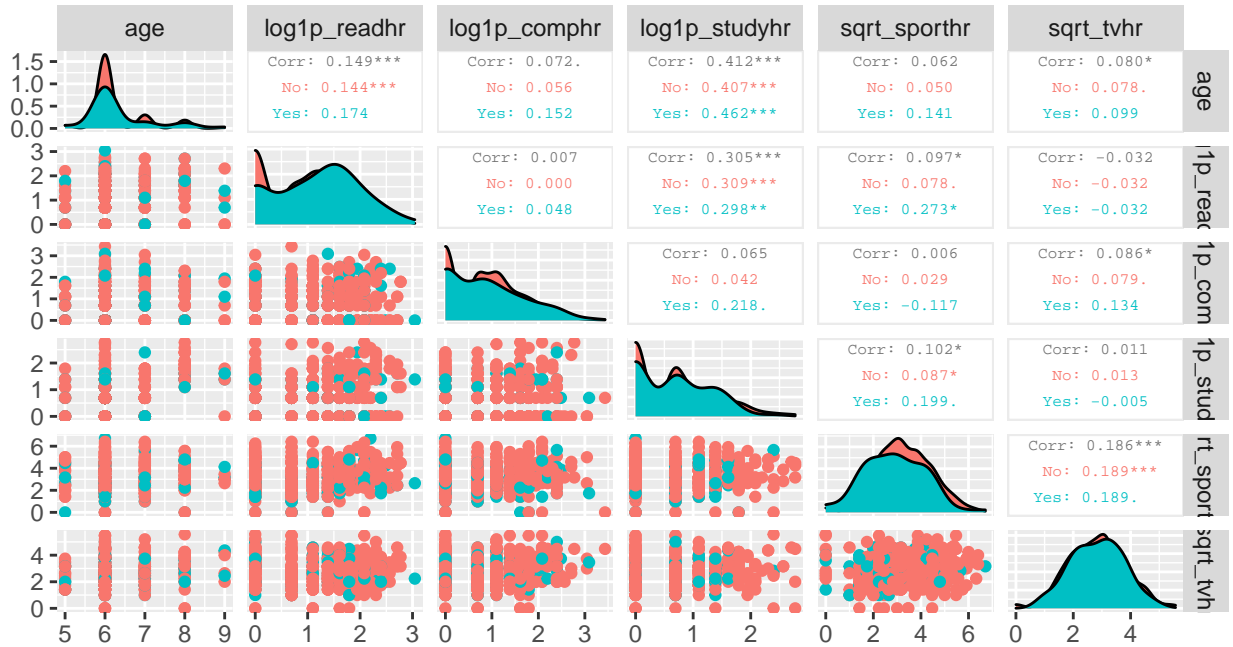


Figure 1: The transformed data is a considerable improvement from the original, notwithstanding a substantial departure from normality. All subsequent discussions are based on the transformed data.

Figure 2 ...

## 3 Models and Methodologies

## Univariate Association with Myopia

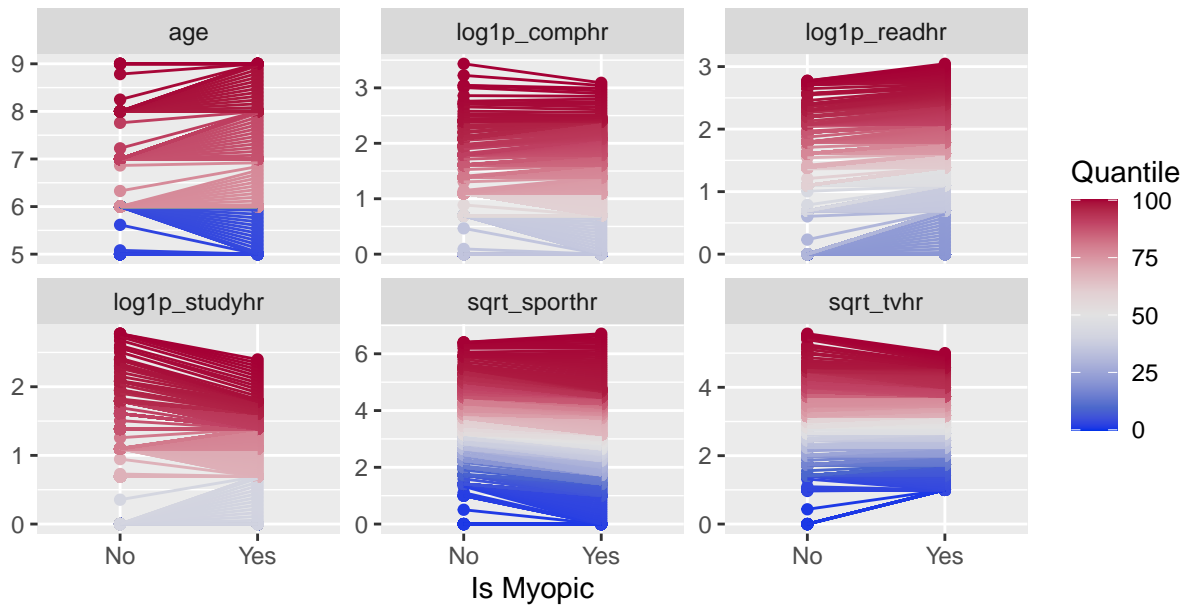


Figure 2:

## Bibliography

- F, R., B, G., A, S., & K-A, L. C. (2017). MixOmics: An r package for 'omics feature selection and multiple data integration. *PLoS Computational Biology*, 13(11), e1005752. <http://www.mixOmics.org>
- Papachristou, G. (2018). *Myopia study*. <https://github.com/ggeop/Myopia-Study>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., & Crowley, J. (2020). *GGally: Extension to 'ggplot2'*. <https://CRAN.R-project.org/package=GGally>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). Springer. <http://www.stats.ox.ac.uk/pub/MASS4>