

## Multi-level modeling

### Introduction

IQ levels are crucial to gain insights to many more information. To further understand family or peer siblings influence on IQ level , this project implies multilevel modeling concepts, The context of data is composed of two preliminary subjects which are the data explanatory and statistical data analysis in which we first understand the data and relations and then we perform our findings and methodology. For the analysis of the dataset, first I cleaned up the dataset, then I imply multi level modeling and analysis of each then I evaluate model based on criterion, hypothesis test is set up for possible attribute combinations and characteristics, assuming significance level of 0.05.

### Dataset

A sample of 60 families are selected, and on for each family 5 to 12 children's were evaluated on multiple subjects which are wordlist, cards, matrices, figures, animals and occupats. . In which mean of these scores give us the IQ score. Assumes no standardized curves.

Acknowledgements:

Third edition of Multilevel analysis: Techniques and applications (Hox, J. J., Moerbeek, M., & Schoot, R, 2018) retrieved from : <https://multilevel-analysis.sites.uu.nl/datasets/>, note that these data are open source no related analysis to our class lessons has been done to it.

### Explanatory Data Analysis

First I checked for null values, haven't found any null or missing values in the data set . No renaming was done. Feature engineering done is the addition of the IQ value which is mean score of sections. And then an aggregated data set per family is created by taking the average across all children in a family and addition of family size variable to maintain utility of data which indicates large or small in accordance to number of children being 6 or more for a family to be considered large

Variable Name	Type	Description
family	Discrete	family identification number
child	Discrete	child number in family
wordlist	Discrete	score in this section
cards	Discrete	score in this section
matrices	Discrete	score in this section
figures	Discrete	score in this section
animals	Discrete	score in this section
occupants	Discrete	score in this section
IQ	Discrete	mean of all sections score

Table 1: IQ data per child

Variable Name	Type	Description
family	Discrete	family identification number
wordlist	Continuous	mean score in this section
cards	Continuous	mean score in this section
matrices	Continuous	mean score in this section
figures	Continuous	mean score in this section
animals	Continuous	mean score in this section
occupants	Continuous	mean score in this section
IQ	Continuous	mean score of all children of all sections score
family.size	Factor	Small for family with less than 6 children and vice versa

Table 2: mean IQ data per family

## Visualization

Level one-covariates: Scores across each child for each section and IQ

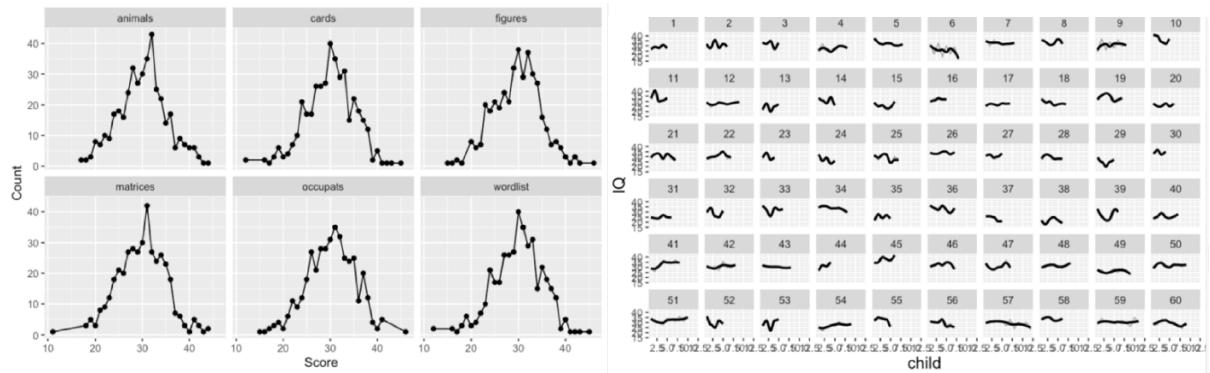


Figure 3 and 4: Histogram and Lattice plot

We can infer that mostly all predictors are approximately normally distributed against IQ score, we can infer that some families IQ score among children is roughly the same other may have some disruptions below mean levels

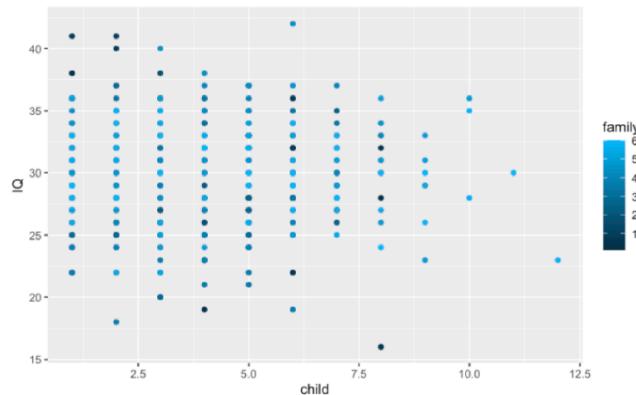


Figure 5: Scatter plot

Scatter plot shows that within family variation is possible to confirm our previous insight.

Level two-covaraites : Mean score among family for each section, IQ and family size

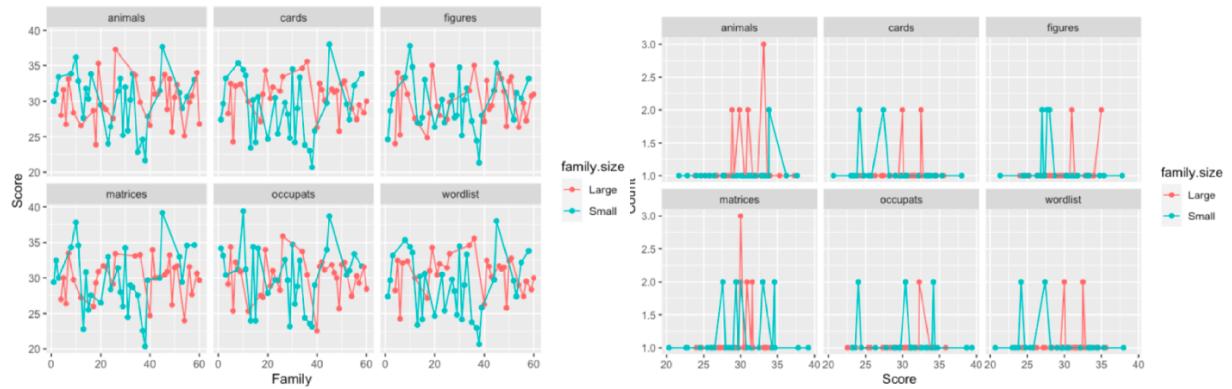


Figure 7 and 8: Level 2 Covariates Plot

Small families tend to have more variation in scores than large families for different sections but lower scores and large families tend to have less variations of scores but higher IQ values.

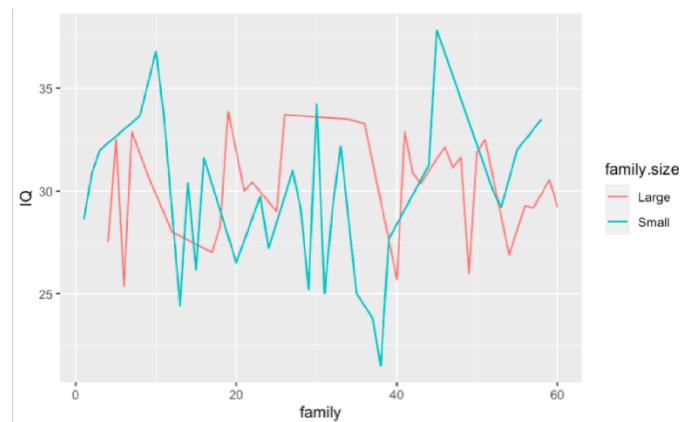


Figure 11 and 12 : Level 2 Covariate Plot

We can confirm above statements by looking at general plot as family ID progresses we can infer that IQ score for small families tend to have more variations but higher overall mean IQ score rather than individual sections evaluated.