Patient Data Analysis: What Factors Influence Early Child Development?

Nova IMS
Statistics
Spring 2019

DATA & OBJECTIVE

The data was gathered from the homes of different families that lived in low-income parts of Mexico City in the 1990's,. It corresponds to the development measures for 45 healthy-born babies of both sexes. A set of values are recorded at 3 month intervals from 12 to 36 months, totalling at 9 recordings per child. The recorded values correspond to different areas of interest of the mother's situation at that point in time, as well as 4 response variables that measure the child's progress in different areas of development.

The dataset, presented in panel format, contains recordings of the following features related to the child's mother:

- ***Child Age***: The child's age in months
- ***First Child***: Whether this is the mother's first child
- ***Mother Age***: The age of the mother at the time of recording
- ***Mother Education***: The number of years of Education after elementary school that the mother completed
- ***Income:*** The household income per capita in U.S Dollars
- ***Breastfeeding***: Does the mother breastfeed
- ***Home:*** score of home's environment and relations
- ***Mother Profile***: Score of mother's attention towards the child
- ***Average Increase***: Average monthly increase of the overall development

The objective is to identify the factors that influence early child development the most, measure the magnitude of their influence, and determine whether the data supports evidence of past studies. To achieved this, a multiple linear regression model will be fitted.

RESPONSE VARIABLES

Motor

Skills related to physical development of the child. Children who learn to use small and larger muscles develop strong motor skills.

Adaptive

Skills such as social responsibility and self-help skills, that allow the child to adjust and function in daily life.

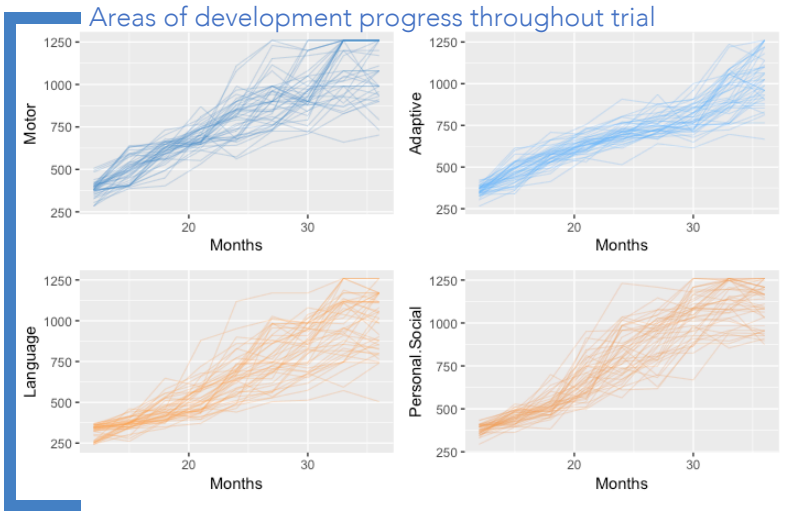
Language

The development of receptive and expressive verbal language skills.

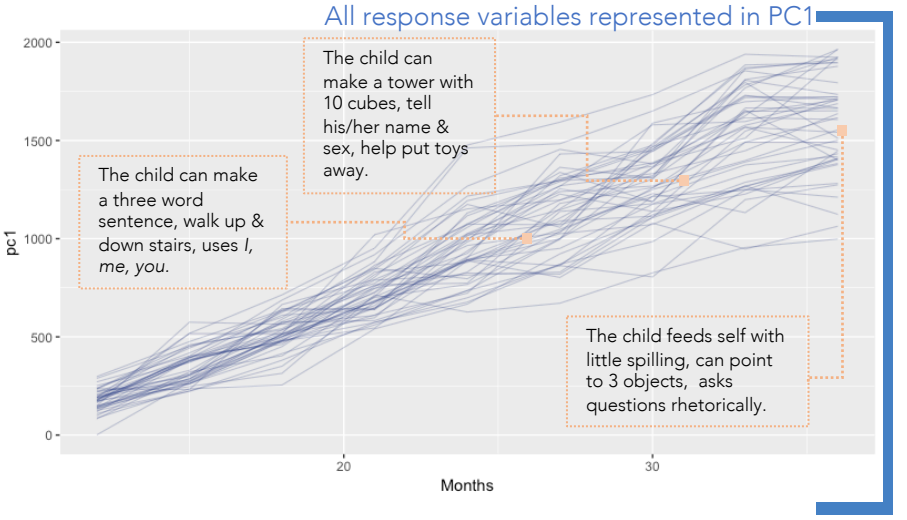
Personal-Social

The ability to take care of themselves and their interaction with others (children and adults).

Areas of development progress throughout trial



All response variables represented in PC1

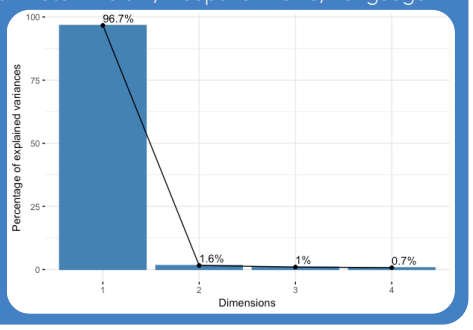


PCA

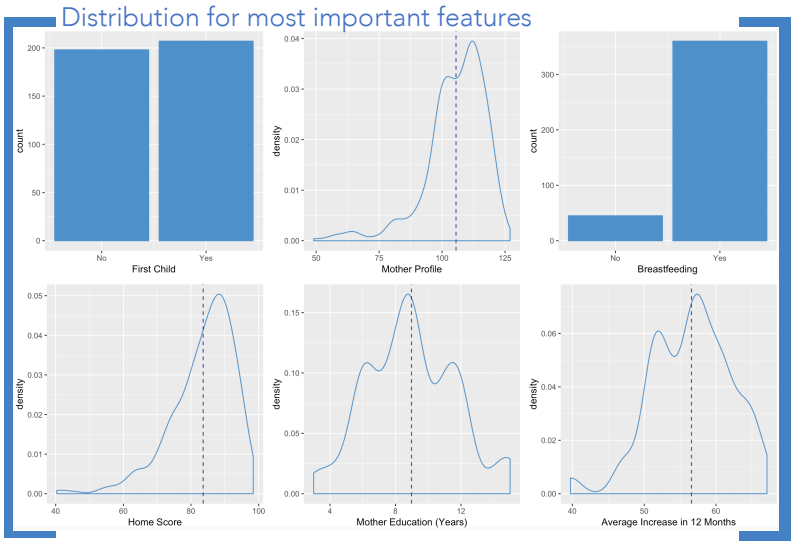
The responses are strongly correlated. In order to simplify the model and for the sake of parsimony, principal components analysis is used to reduce the dimensionality on the four response indicators of child development. Our new response variable will be called "*Overall Development*", represented by the first principal component. This PC is suitable as a response representing overall child development since it accounts for almost 97% (see figure below) of the total variance and has more or less equally distributed weights on the different indicators (loadings: Motor = 0.54 , Adaptive = 0.45, Language = 0.45 , Personal-Social = 0.54).

The previous technique will also help to *smooth* the response by reducing the "big jumps" that some individuals presented on a certain indicator for a given point in time.

The remaining PC's try to catch the differences (variance) in the child development related to each individual indicator. We kept just the first PC given that the marginal analysis of the different areas of child development goes out of the scope of this study.



Distribution for most important features



MULTIVARIATE LINEAR MODELS (MLMs)

Model 1: MLM with Box Cox transformation [4] on the response (lambda =0.7)
→ p value of Bartlett test of homogeneity of variances is under 0.01
→ We reject that the variances are constant

Heteroscedasticity of the variances

Model 2: Model 1 + weighted fitting [5]
→ p-value of Durbin Watson < 0.01
→ We reject that the errors are independent

First order **autocorrelation** on the errors

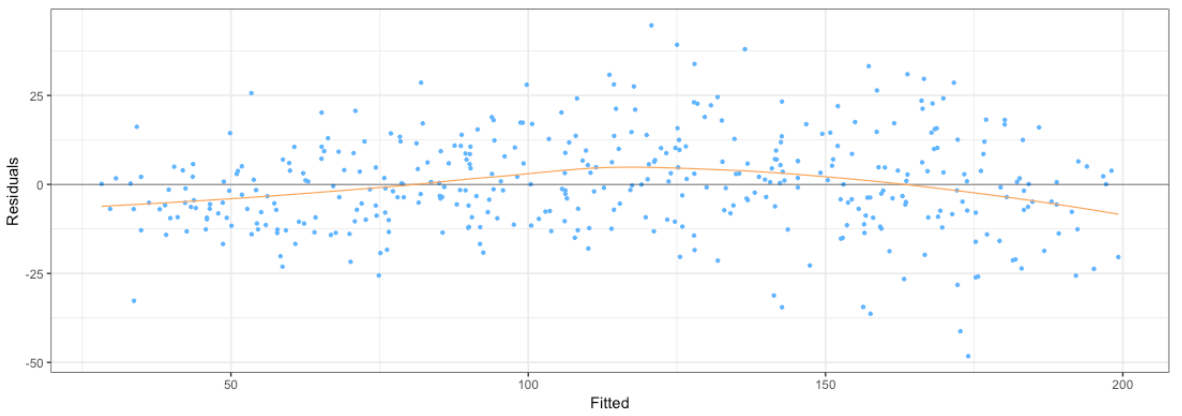
Model 3: Model 1 + Cochrane Orcutt Estimation [5]
→ p-value of Durbin Watson statistic = 0.3
→ Not strong enough to reject that the errors are independent.
→ Final Model

Assumptions \ Models	Model 1	Model 2	Model 3
Normality	✓	✓	✓
Homogeneity	✗	✓	✓
Independence	✗	✗	✓
No collinearity	✓	✓	✓
Adj. R-squared	0.9196	0.8822	0.9218

Model 3 Coefficients

Feature	Estimate	P-value (t-statistic)
Mother Age	0.23759	0.37846
Home	0.11471	0.08331
Mother Profile	0.070039	0.15815
Income	–0.00619	0.66572
Average Increase	0.78235	0.00015
Child Age	5.70277	< 2.2e-6
First Child	6.46969	0.00926
Breastfeeding	7.14153	0.03631
Child Sex	-1.15948	0.57985
Mother Education	0.97517	0.0177

Model 3 – Model 1 + Cochrane Orcutt Estimation



CONCLUSION

We can conclude from our analysis that Breastfeeding and First Child have the strongest influence on our regressors; being the first child has a positive relation on early age child development, this agrees with [2] and [3]. On average it gives 6 more points to the transformed overall development index. Breastfeeding helps increase the overall development, which concurs with [2]. On average, breastfeeding gives 7 more points to the transformed overall development index.

The mother's education, the average development increase in the first year, the home environment, and the mother profile have a positive relation with child development. Child sex, income, and the mother's age are not factors that influence child development on a statistically significant way.

Emilio Herreman - e20181378
Julian Kuypers – m20180409
Olivier Hofman - m20180417
Giacomo Zoppi - e20181375

References: [1] "*Birth Order and Human Capital Development; Evidence from Ecuador*", Monique De Haan Erik Plug José Rosero
[2] "*Patterns of Early-Life Social and Environmental Exposures and Child Cognitive Development, Rhea Birth Cohort, Crete, Greece*", Mariza Kampouri et. al.
[3] "*The Early Origins of Birth Order Differences in Children's Outcomes and Parental Behavior*", Jee-Yeon, K. Lehmann, Ana Nuevo-Chiquero Marian Vidal-Fernandez
[4] Chambers, J. M. (1992) Linear models. Chapter 4 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.
[5] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London Series A 160, 268–282.