

Project 1: PCA Analysis on a Birth Weight Dataset

Trevor Carpenter, Christina De Cesaris, Michelle Tran
STA 135: Multivariate Data Analysis
Professor Maxime Pouokam
University of California, Davis
May 4, 2021

I. Introduction

Although low birth weighted infants can be healthy, under improper care or other circumstances, it can cause serious long-term health problems. A low birth weight is defined as being below 2.5kg. It can be caused by a multitude of factors —poor socioeconomic situation, premature birth, a mother’s pre-existing health conditions, and more. In this project, we will highlight how smoking and other factors can cause low birth weights in infants in comparison to non-smoking parents.

[state question of interest]. In order to answer this, we will be using principal component analysis to analyze the data. PCA is a statistical technique that reduces a dimensionality of a dataset to make it interpretable while also preserving the variability of the data. [more explanation?] We can use it to see what variables group together, identify possible outliers, recognize correlations between variables, and more.

II. Summary of Data

III. Analysis

IV. Interpretation

V. Conclusion

Appendix: R Script

```
knitr::opts_chunk$set(echo = F)
smok = read.csv("data/Birthweight_reduced_kg_R.csv")
#subset(smok, smoker == 1, select = c("Length", "Birthweight"))
#Plots
```