https://docs.google.com/document/d/1ltWb6UROtAZ8m3OPb7\_toTnE175teBF-HZU02JnCp30/edit?usp=sharing

**Executive summary**

Low birth weight is one of the most common indicators of infant health. While it is not this report’s objective to establish causation or identify main factors, we are trying to find main associations which might cause low birth weight. A total of 1236 babies from a US hospital were evaluated in applied statistical methods including an unpaired t-test, analysis of variance (ANOVA) and a simple linear regression model.

We are aiming to measure mean baby birth weight of smoking and non-smoking mothers is compared using a t-test in order to investigate a difference effected by smoking habit during pregnancy.

The ANOVA compares the mean baby birth weight of three different ethnicities of mothers to explore differences between the considered ethnic groups 0-5=white 6=mex 7=black 8=asian 9=mixed 99=unknown. The linear regression model intends to firstly describe the relationship of weights and multiple variables and further on to create a prediction model that allows to foresee children’s birth weights with some significant variables.

@Srishti - about annova

Bootstrapping?

@Srishti and @Jacks

Despite this, future research using additional exploratory factors influencing weights of infants could be used in multivariate regression models.

All these findings are highly limited in its application because of the limited sampled population, as it is very small and collected in 1964. Further research in this topic could deepen our understanding of this connection but would fail to help in preventing low birth weights. Nonetheless, the findings in this report allow to point research into an adequate direction. The contribution of this analysis is to identify mother’s smoking habits and ethnicity as possible determinants of children birth weight along with other attributes that might affect birth weight.

**Introduction**

Previous studies shown low birth weight “is the single most important factor affecting neonatal mortality and a significant determinant of post-neonatal mortality”. Naturally, there is an elevated interest of parents to avoid and therefore also to explore the determinants of low birth weights. In this study we are aiming to test certain factors and their relationship to birth weights. This report investigates the effects of mothers’ attributes mainly smoking habit and ethnic race on the weight of newborn babies, in order to make possible future suggestions about prevention of low birth weights.

The dataset used for this report consists of 1236 observations of babies collected from a US hospital in the 1960s.

For the purpose of this study we decided to eliminate some of the variables which seem to

In this data analysis three main suggestions are that there is a difference on babies’ weight due to races, that there is evidence that smoking has a negative influence on babies’ weight and finally the relationship between mothers’ weight and babies’ weight is possibly not simple linear.

Summarasing the findings??

Lots of NAs 🡪 608 observations

**EXPLORATORY DATA ANALYSIS**

An initial descriptive statistic summary is conducted to explore potential relationships in the dataset. This section looks at the frequencies of baby’s weights across factors, and then the three possible relationships raised above. The exploratory data analysis was produced with the help of R software and plot features specifically.

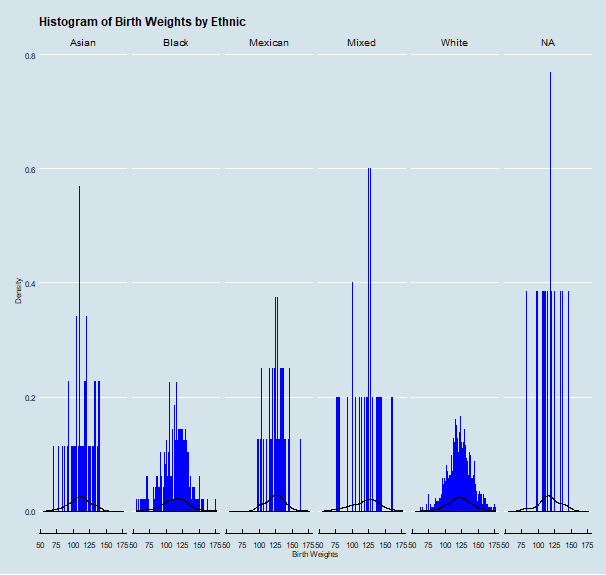
We assumed each baby is distinct therefore, assuming each baby to be independent seem to be reasonable to proceed with this analysis.

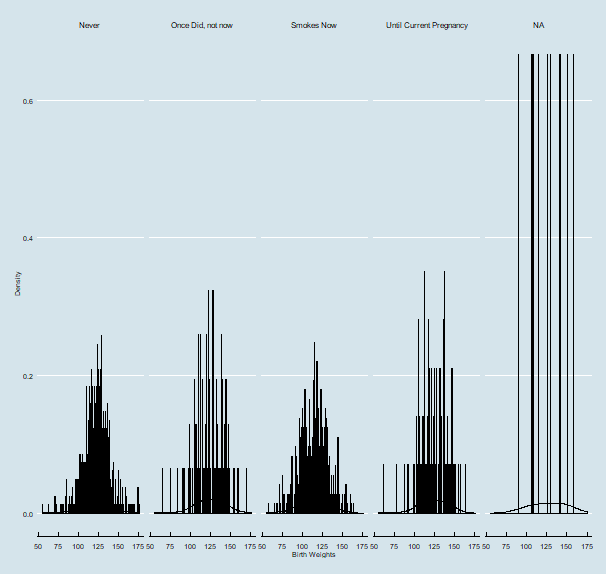
**ETHNICITY AND BIRTH WEIGHTS**

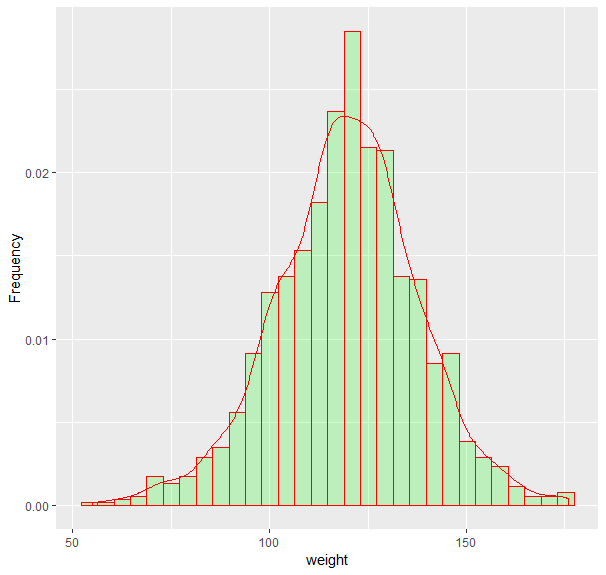
Below Table shows a summary babies weights across ethnic groups of their mothers. The mean birth weights for ethnic White and Mexican ethnic groups show a central tendency since their means are higher and close to each other. There is similarity of dispersion for each group with the figures obtained for standard deviations of the ethnic groups in terms variability in each ethnic group.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mother Race | Mean | Standard Deviation |
| 1 | Asian | 110.4318 | 15.99331 |
| 2 | Black | 113.2377 | 19.08851 |
| 3 | Mexican | 124.15 | 14.14313 |
| 4 | Mixed | 119.8 | 20.14324 |
| 5 | White | 121.6414 | 17.69685 |
| 6 | NA | 116.8462 | 16.73741 |

Figure below illustrates the distribution of the birth weights by race. The distributions for each group are roughly symmetric and unimodal. In the histogram there seems to be a difference in mean newborn weights across white and black ethnic groups, this will be tested in further sections using inferential tests.

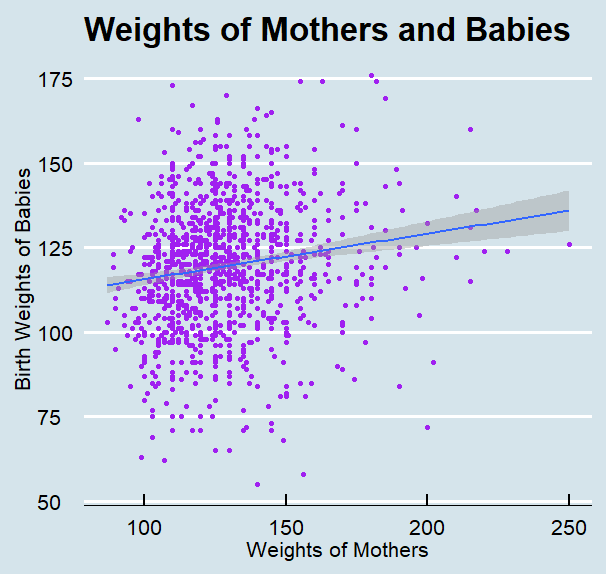






THE RELATIONSHIP BETWEEN THE WEIGHTS OF THE MOTHERS AND THE INFANTS

Investigating and exploring relationships between the weights of mothers and babies we managed to plot relationship between weight of mothers and babies delivered by below scatterplot. In below figure, the response variable, baby’s weight, is displayed against the mother’s weights. There is very weak relationship as there is no clear evidence that the points are tightly clustered, and several outliers interfere the overall pattern.



**Methods**

Anova

@srishti

**LINEAR REGRESSION**

@srishti

**Linearity Assumption**

**Normality Assumption**

**Findings**

**RESULTS**

**RESULTS -Anova**

**RESULTS- Linear Regression**

**Discussion**

**DISCUSSION- Anova**

@srishti

**DISCUSSION- Linear Regression**

@srishti

@Jacks

**Dataset validation**

**CONCLUSION**