Exploring Factors Associated with Low Birthweight

Prachi Tiwari, id = 23102057

17 October, 2023

# Exploratory data analysis

In this assignment you will carry out an **exploratory data analysis** of factors thought to be associated with low birthweight. The factors of interest are maternal smoking during pregnancy, history of premature labour, and number of visits to a physician in first trimester (the first trimester is the first three months of the pregnancy). The goal of the exploratory data analysis is to gain a better understanding of these factors and their associations with low birthweight in our dataset.

## Load the Libraries and Data Needed

The dataset you will be using is available in the aplore3 package and you will use the tidyverse package for data manipulation. You may not have these installed. If you get a yellow message to install one or both packages when you open this document in Rstudio, click on install. If you need to install them manually then remove the # and run following code in . You will only need to install the packages once.

#install.packages("tidyverse")   
 # install.packages("aplore3")

Load the required packages and make the birthweight dataset, lowbwt available using the data function as follows:

library(tidyverse)  
library(aplore3)  
  
data(lowbwt)

The low birthweight data is from the textbook “Applied Logistic Regression” by Hosmer and Lemeshow. The data are on 189 babies born at the Bayside Medical Centre in Massachusetts, in 1986. The following is a description of the variables in this dataset.

| Name | Description |
| --- | --- |
| subject | identification code |
| low | low birthweight (“< 2500 g” or “>= 2500 g”) |
| age | age of mother |
| lwt | weight at last menstrual period (pounds) |
| race | race (Black, White, Other) |
| smoke | smoked during pregnancy (Yes, No) |
| ptl | premature labour history (None, One, Two, etc.) |
| ht | history of hypertension (Yes, No) |
| ui | uterine irritability (Yes, No) |
| ftv | number of visits to physician during 1st trimester (three categories: None, One, Two, etc.) |
| bwt | birthweight (in grams) |

# Tabulations

The key variable of interest is low which represents whether a baby is born low birthweight, defined as a birthweight below 2,500 grams.

lowbwt %>% select(low) %>% table()

## low  
## >= 2500 g < 2500 g   
## 130 59

The tabulation shows 59 of the babies were low birthweight, and 130 were not low birthweight.

Let’s explore the association between smoking during pregnancy and low birthweight by cross-tabulating the two variables low and smoke.

lowbwt %>% select(smoke, low) %>% table()

## low  
## smoke >= 2500 g < 2500 g  
## No 86 29  
## Yes 44 30

From the cross-tabulation, we see that in this dataset 44+30 = 74 mothers smoked during pregnancy, and 86+29=115 did not. It seems the proportion of low birthweight babies is higher for mothers who smoked during pregnancy than those who did not smoke - let’s calculate row percentages so we can compare the proportion of low birthweight babies for mothers who smoked with the proportion of low birthweight babies for mothers who did not smoke.

lowbwt %>% select(smoke, low) %>% table() %>% prop.table(margin = 1)

## low  
## smoke >= 2500 g < 2500 g  
## No 0.7478261 0.2521739  
## Yes 0.5945946 0.4054054

We now explore the association between the mother having a history of any premature labour and low birthweight. First, in the following chunk, we recode the ptl variable into two categories, 0 for no premature labour and 1 for 1 or more premature labours, and call the new variable ptl\_any.

lowbwt<-lowbwt %>%   
 mutate(ptl\_any = ifelse(ptl =="None", 0, 1))

**Task 1**: In the following chunk explore the association between whether the mother had any previous premature labour (ptl\_any) and whether the baby was born with low birthweight (low), using both counts and appropriate proportions/percentages. Describe your findings in 1-2 sentences.

Answer: It looks like mothers who had premature labour history are more likely to give birth to babies who weigh less than 2500g which is 60% according to our dataaset. And mothers with no premature labour history are more likely to give birth to babies who weigh more than 2500g which is 74% according to our dataset. Thus, the percentage of babies with low birthweight is much higher with mother having premature labour history.

lowbwt %>% select(ptl\_any, low) %>% table()

## low  
## ptl\_any >= 2500 g < 2500 g  
## 0 118 41  
## 1 12 18

lowbwt %>% select(ptl\_any, low) %>% table() %>% prop.table(margin = 1)

## low  
## ptl\_any >= 2500 g < 2500 g  
## 0 0.7421384 0.2578616  
## 1 0.4000000 0.6000000

**Task 2**: In the following chunk explore the association between the number of visits to a physician in the first trimester (ftv) and whether the baby was born with low birthweight (low), using both the counts and appropriate proportions/percentages. Describe your findings in 1-2 sentences.

Answer: It looks like mothers who visited physician in the first trimester are less likely to give birth to babies with low birthweight (25%) compared to mothers who did not visit a physician at all (36%) Therefore, low birthweight is inversely related to the number of visits to a physician.

lowbwt %>% select(ftv, low) %>% table()

## low  
## ftv >= 2500 g < 2500 g  
## None 64 36  
## One 36 11  
## Two, etc. 30 12

lowbwt %>% select(ftv, low) %>% table() %>% prop.table(margin = 1)

## low  
## ftv >= 2500 g < 2500 g  
## None 0.6400000 0.3600000  
## One 0.7659574 0.2340426  
## Two, etc. 0.7142857 0.2857143

lowbwt<-lowbwt %>%   
 mutate(ftv\_any = ifelse(ftv =="None", 0, 1))  
lowbwt %>% select(ftv\_any, low) %>% table() %>% prop.table(margin = 1)

## low  
## ftv\_any >= 2500 g < 2500 g  
## 0 0.640000 0.360000  
## 1 0.741573 0.258427

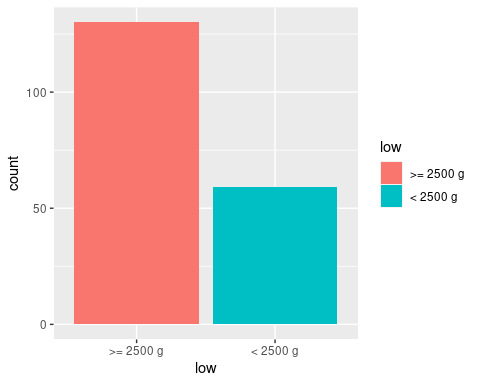
# Barcharts

Now we will create some barcharts.

## Barchart of Low Birthweight

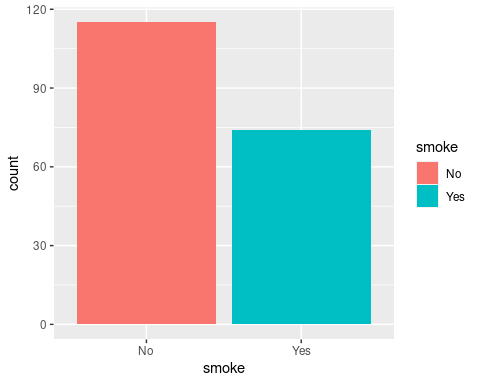
The following is a barchart of low birthweight status.

ggplot(lowbwt, aes(x = low, fill = low)) +  
 geom\_bar()



**Task 3**: In the following R chunk create a bar chart of smoking status of mothers.

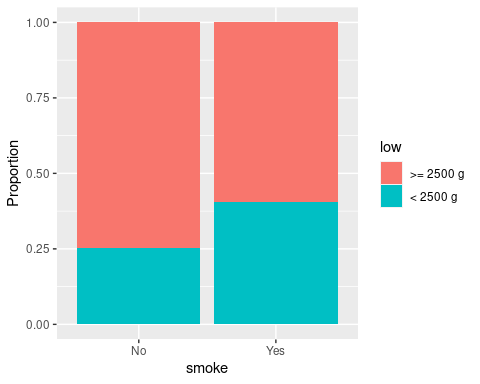
ggplot(lowbwt, aes(x = smoke, fill=smoke)) +  
 geom\_bar()



## Barchart of Low Birthweight by Smoking Status

Below is a stacked barchart of low birthweight of the babies by smoking status of mothers.

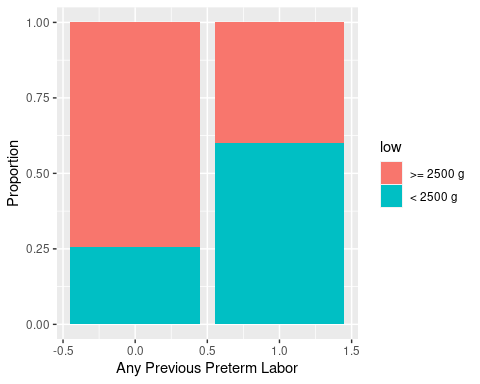
ggplot(lowbwt, aes(x = smoke)) +  
 geom\_bar(aes(fill = low), position = "fill") +  
 ylab("Proportion")



## Barchart of Low Birthweight by Preterm Labour

**Task 4**: Create a stacked barchart of low birthweight by any previous preterm labour (ptl\_any) by writing appropriate code in the R chunk below.

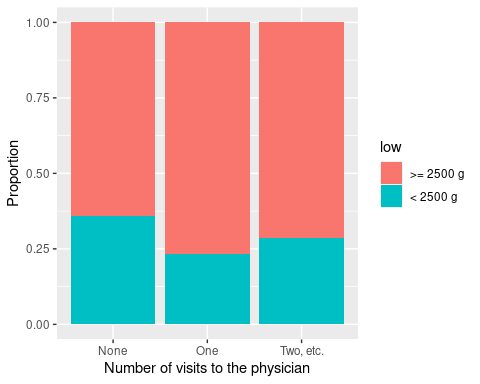
ggplot(lowbwt, aes(x = ptl\_any)) +  
 geom\_bar(aes(fill = low), position = "fill") +  
 ylab("Proportion") + xlab("Any Previous Preterm Labor")



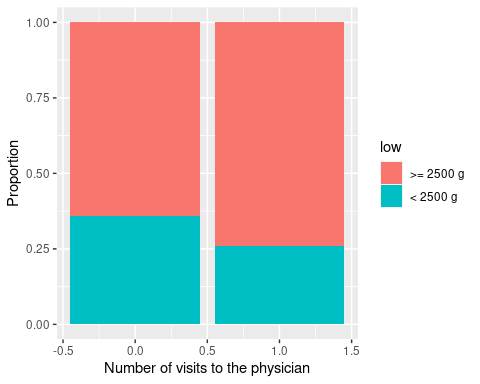
## Barchart of Low Birthweight by Number of Visits to Physician in first trimester

**Task 5**: Create a stacked barchart of low birthweight by number of visits to the physician in the first trimester by writing appropriate code in an R chunk below.

ggplot(lowbwt, aes(x = ftv)) +  
 geom\_bar(aes(fill = low), position = "fill") +  
 ylab("Proportion") + xlab("Number of visits to the physician")



ggplot(lowbwt, aes(x = ftv\_any)) +  
 geom\_bar(aes(fill = low), position = "fill") +  
 ylab("Proportion") + xlab("Number of visits to the physician")



## Comments on barcharts

**Task 6**: Once you have created the plot write your interpretation of the association between with low birthweight and each of the three factors, based on the three barcharts. You should mention the direction and strength of each association. (2-3 sentences)

ANSWER:

Smoking mothers are more likely to have higher proportion of low birthweight. Mothers having premature labour history are likely to have higher proportion of low birthweight. Number of visits to physician results in less chances of low birthweight compared to having no visits at all.

The trend is in favour of low birthweight, indicating that smoking during pregnancy increases that risk. And the strength is moderate means smoking mothers has moderately higher proportion of low birthweight. Preterm labour history is skewed in the direction that having no prior preterm labour is linked to a decreased risk of low birthweight. And the strength is strong means babies delivered to moms who have never experienced preterm labour have a higher chance of having low birthweight instances than those who have. For number of visits, trend is downward, meaning more visits lowers the risk of low birthweight.

# Conclusion

**Task 7**: Based on the tabulations and bar charts, write a brief conclusion on whether you think the three factors (1) mother smoking during pregnancy, (2) any preterm labour, and (3) number of visits to a physician in the first trimester, could be useful to predict if a baby might be born with low birthweight. Do you think the relationships between low birthweight and these three factors you have observed in this dataset could be used to make inferences about the relationships in a wider population, and if so, what population?

ANSWER:

Based on the tabulations and bar charts, smoking, preterm labour, and number of visits to a physician appear to be connected with low birthweight in newborns to varied degrees. They can be useful to predict low birth weight in newborns.

Inferences can be made in a comparable population similar to the one used in dataset. But for wider population, more research would be required. The wider population could be healthcare people, academics, public health officials and so on.

**Task 8**: “knit” the file as a Word (or PDF) document and submit the Word/pdf document to the Assignment 3 submission link on canvas before the deadline. **Do not submit the .Rmd document or you will lose marks**.