Investigating the Factors Affecting Birthweight

Chin Zhe Jing, id = 22221970

30 September, 2022

# Question of Interest

Are the factors of smoking, previous history of hypertension or urinary irritability associated with whether babies were born with low birthweight (less than 2,500 grams)?

## Load the Libraries and Data Needed

The dataset you need is available in the aplore3 package and we will also use the tidyverse package. These may not be installed already. If you need to install them then run following code in R. You will only need to do this once. After they have been installed you can just delete this chunk and text.

install.packages(c("tidyverse", "aplore3"),   
 repos = "https://cran.microsoft.com/", quiet = TRUE)

Load the required libraries so you can use them, and then make the birthweight data available (‘lowbwt’) as follows:

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

The low birthweight data is from the “Applied Logistic Regression” textbook by Hosmer and Lemeshow. 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 (None, One, Two, etc.) |
| bwt | birthweight (in grams) |

## Subjective Impressions

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

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

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

Let’s explore the association between history of hypertension and low birthweight by tabulating the data.

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

## ht  
## low No Yes  
## >= 2500 g 125 5  
## < 2500 g 52 7

It seems there were not many mothers with hypertension, but the proportions of low weight babies is very much higher for mothers suffering from hypertension status than those that were not.

lowbwt %>% select(low, ht) %>% table() %>% prop.table(margin = 2)

## ht  
## low No Yes  
## >= 2500 g 0.7062147 0.4166667  
## < 2500 g 0.2937853 0.5833333

Task: In the following R chunk explore the association between uterine irritability and whether the babies were born with low birthweight, using both the counts and appropriate percentages. Explain the results in words.

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

## ui  
## low No Yes  
## >= 2500 g 116 14  
## < 2500 g 45 14

lowbwt %>% select(low, ui) %>% table() %>% prop.table(margin = 2)

## ui  
## low No Yes  
## >= 2500 g 0.7204969 0.5000000  
## < 2500 g 0.2795031 0.5000000

print("There were less cases with mother having uterine irritability, but the proportions of low weight babies is the same for both mothers having uterine irritability and those that were not.")

## [1] "There were less cases with mother having uterine irritability, but the proportions of low weight babies is the same for both mothers having uterine irritability and those that were not."

Task: In the following R chunk explore the association between smoking status and whether the babies were born with low birthweight, using both the counts and appropriate percentages. Explain the results in words.

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

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

lowbwt %>% select(low, smoke) %>% table() %>% prop.table(margin = 2)

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

print("The proportions of low weight babies is much lower for mothers who smoke than those that were not.")

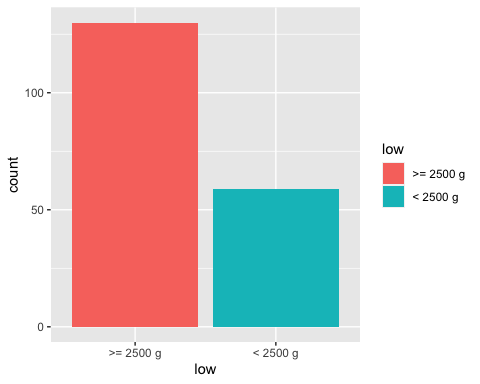
## [1] "The proportions of low weight babies is much lower for mothers who smoke than those that were not."

Now we will create some barcharts.

# Barchart of Low Birthweight

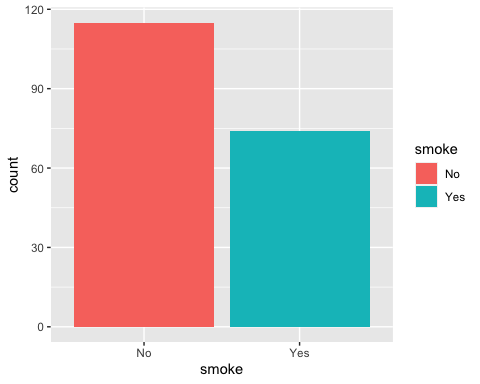
The following is a frequency plot of the low birthweight status.

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



Task: In the following R chunk create a frequency plot of the smoking status.

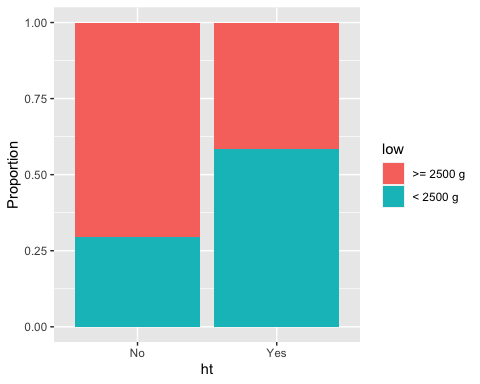
ggplot(lowbwt, aes(x = smoke, fill = smoke)) +  
 geom\_bar()



# Stacked Barchart of Low Birthweight by Hypertension Status

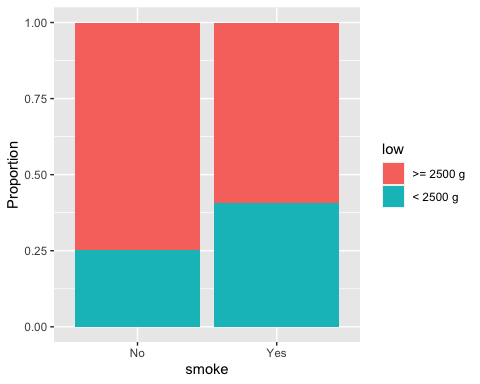
Below is a relative frequency plot of the low birthweight of the babies against the hypertension status of the mothers using a stacked barchart.

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



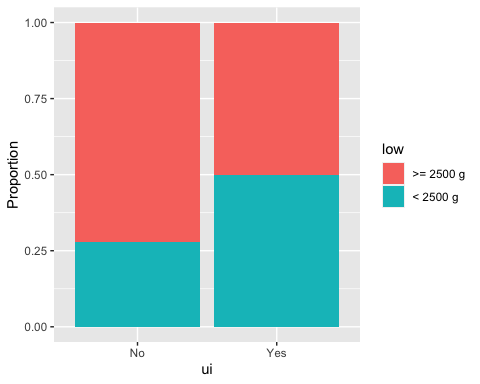
Task: Create a stacked barchart of low birthweight by smoking status by inserting an R chunk and relevant code below.

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



Task: Create a stacked barchart of low birthweight by uterine irritability by inserting an R chunk and relevant code below.

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

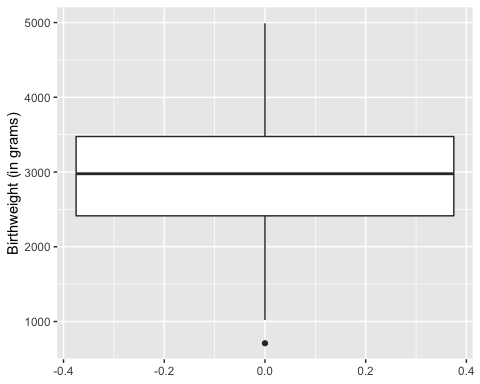


Task: Once you have created the plots, explain your interpretation of which factors are associated with low birthweight based on the three barcharts. State which factor you think is most associated with birthweight.

More mothers having hypertension is giving birth to lower weight babies while less smoking mothers are having low birthweight babies. The proportion of low birthweight babies between mothers having uterine irritability and those who were not are the same. Thus, based on the three barcharts, i think that hypertension is most associated with birthweight as mothers suffering from hypertension has higher proportion of lower weight babies.

The following R chunk produces a boxplot of the birthweight distribution.

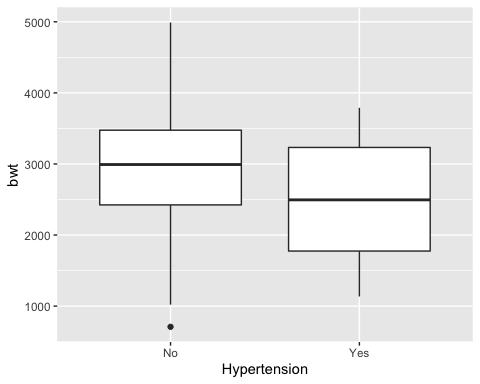
lowbwt %>% ggplot(aes(y = bwt)) +  
 geom\_boxplot() +  
 labs(y = "Birthweight (in grams)")



Task: In the previous task you stated which factor you believe was most associated with birthweight, so you can explore the impact on the distribution in more detail. Create a graph of side-by-side boxplots comparing the birthweight distribution for each level of that factor (e.g. comparing mothers who had uterine irritability and those who did not), by inserting an R chunk and relevant code below.

[Hint: we used side-by-side boxplots in the week 4 lab and in the Exploratory Data Analysis worksheet]

lowbwt %>% ggplot(aes(y = bwt, x = factor(ht))) +  
 geom\_boxplot() +  
 labs(x = "Hypertension")



# Conclusion

Task: Write a short conclusion of whether you think low birthweight of babies can be predicted based on whether the mother smoked, has hypertension or uterine irritability.

In conclusion, the above charts show us that mother has hypertension is most probably to have low birthweight babies compared to mother who smokes or has uterine irritability. Hence, predict low birthweight of babies could possibly be based on hypertension status of mother.

Final Task: “knit” the file as a Word of PDF document and submit it via the relevant link on Blackboard before the deadline.