STA 138: Final Project

Aman Singh & Jennifer Li
ID:915763034
ID:999339769
UC Davis
Instructor: Prabir Burman

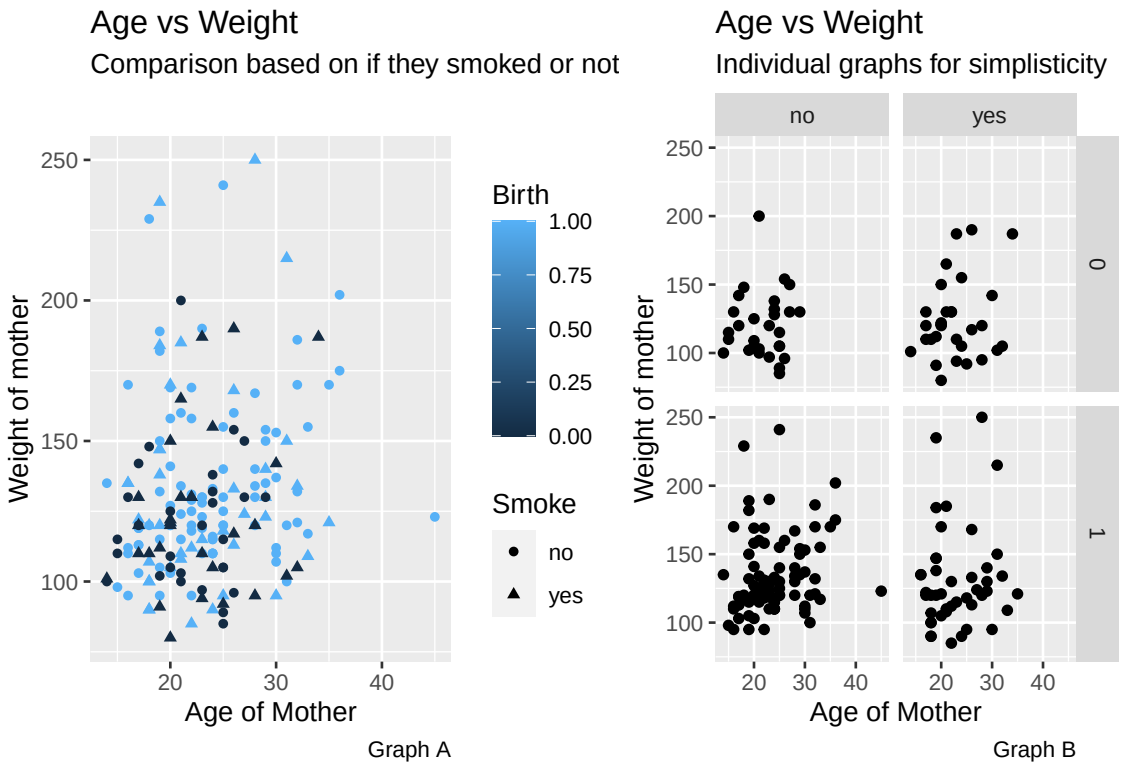
3/09/2021

Question 1: Low Birth Rate

(I) Introduction

Smoking has been an huge issue since the 20^{th} century. There used to be advertisements all the time on the television where they would show as people having a great time and enjoying life. However, decades later, people found out that smoking lead to cancer and to horrible birth defects. Upon learning this, the government banned ads for smoking and started to limit the exposure of smoking to the public. Even when buying an cigarette, you have to be 18 years old as well as there is a huge caution warning on the box stating that it can lead to various defects in the body. The worst defect that we have seen so far has been when women who are pregnant are smoking at the same time. This can lead to some serious defects to the children and can affect them in the long run. While we do know that, what we want to investigate is whether the fact the probability of low birth weight of infant is related to information on mother such as age, weight smoking status. We will be using the dataset called `Baby` provided by Professor Prabir Burman. `Baby` has 7 columns, `age`(age of the mother), `weight`(weight of the mother before pregnancy), `smoke`(smoking status during pregnancy), `pre`(history of pre-mature labor), `hyp`(history of hypertension), `visits`(the number of visits during the first trimester), and `birth`(if the birth weight of the infant was low or not).

(II) Materials and Methods




```
x = glm(birth ~ age + weight + factor(smoke) + factor(pre) + factor(hyp) +
    visits + age:weight + weight:factor(pre) + weight:factor(hyp), data = baby,
    family = binomial)
x
```

```
##
## Call:  glm(formula = birth ~ age + weight + factor(smoke) + factor(pre) +
##     factor(hyp) + visits + age:weight + weight:factor(pre) +
##     weight:factor(hyp), family = binomial, data = baby)
##
## Coefficients:
```

```
##           (Intercept)                   age                weight
##            -3.2742473             0.1165393             0.0258898
##      factor(smoke)yes        factor(pre)yes        factor(hyp)yes
##            -0.5087686            -2.6303649            -1.6168753
##                visits            age:weight weight:factor(pre)yes
##             0.0292054            -0.0004465             0.0064942
## weight:factor(hyp)yes
##            -0.0009711
##
## Degrees of Freedom: 188 Total (i.e. Null);  179 Residual
## Null Deviance:       234.7
## Residual Deviance: 202   AIC: 222
```

(III) Results

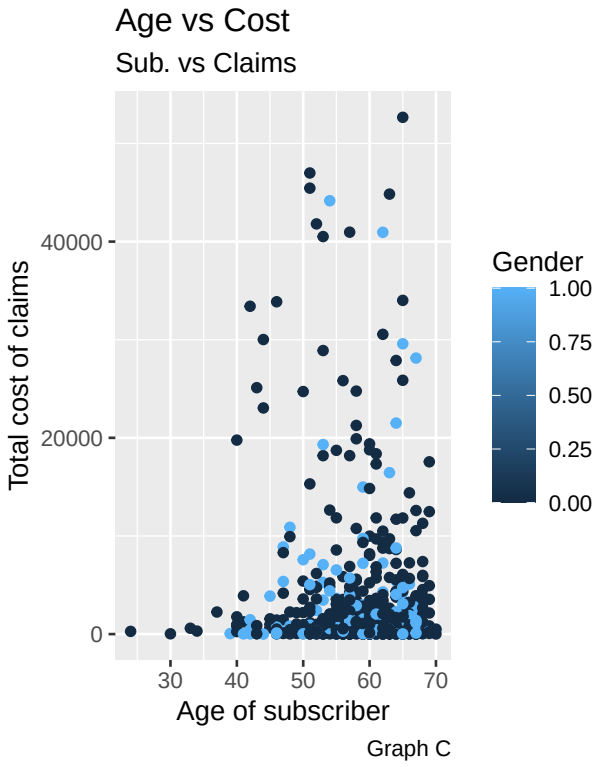
(IV) Conclusion

Question 2: Ischemic heart disease

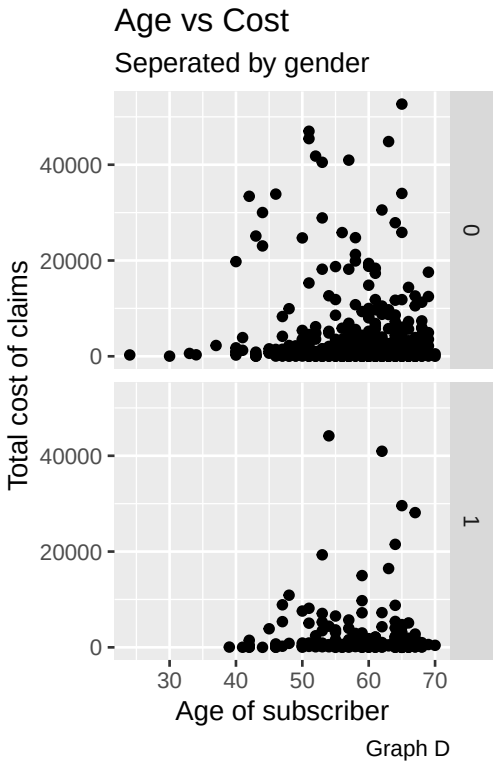
(I) Introduction

Heart disease is one of the most common diseases which can lead to death. It is in fact an very serious disease and is an type of disease which starts off by having an buildup of plaque. This leads to the coronary arteries thus having to narrow, which limits the blood flow to the heart. Some symptoms of coronary artery disease can really range from no symptoms, to chest pain and even an heart attack depending on the person. Some treatment can help but ultimately there is no cure and would have to deal with for the rest of your life. In this paper, we are looking into an dataset called `ischemic` given to us by Professor Prabir Burman which contains 9 columns. The 9 columns are `cost`(the total cost of claims made by the subscriber), `age`(age of the subscriber), `gender`(gender of subscriber),`inter`(total number of interventions or procedures carried out),`drugs`(number of tracked drugs prescribed), `complications`(number of other complications that came from the heart treatment), `comorbidities` (number of other diseases that the subscriber had during the period), `duration` (number of days of duration of treatment condition), and `visits` (number of emergency room visits). We will be using this dataset to perform an poisson regression that will perform an data summary, goodness-of-fit and model selection to model the mean as an function of 8 other variables.

(II) Materials and Methods



Graph C



Graph D

(III) Results

(IV) Conclusion

Ⓥ

Code Appendix

```
# cuttingoffcode
library(knitr)
opts_chunk$set(tidy.opts = list(width.cutoff = 70), tidy = TRUE)
# library & importing data
library(readxl)
library(tidyverse)
library(patchwork)
baby <- read_excel("baby.xls")
# analysis for question 1
p = baby %>% ggplot() + geom_point(mapping = aes(x = age, y = weight, color = birth,
    shape = smoke)) + labs(title = "Age vs Weight", subtitle = "Comparison based on if they smoked or n
    x = "Age of Mother", y = "Weight of mother", color = "Birth", caption = "Graph A",
    shape = "Smoke")

b = baby %>% ggplot(aes(age, weight)) + geom_point() + facet_grid(vars(birth),
    vars(smoke)) + labs(title = "Age vs Weight", subtitle = "Individual graphs for simplisticity",
    x = "Age of Mother", y = "Weight of mother", caption = "Graph B")

p + b
x = glm(birth ~ age + weight + factor(smoke) + factor(pre) + factor(hyp) +
    visits + age:weight + weight:factor(pre) + weight:factor(hyp), data = baby,
    family = binomial)
x
# Question 2
ischemic <- read_excel("ischemic.xlsx")
a = ischemic %>% ggplot() + geom_point(mapping = aes(x = age, y = cost,
    color = gender)) + labs(title = "Age vs Cost", subtitle = "Sub. vs Claims",
    x = "Age of subscriber", y = "Total cost of claims", color = "Gender",
    caption = "Graph C", shape = "Smoke")

b = ischemic %>% ggplot(aes(age, cost)) + geom_point() + facet_grid(vars(gender)) +
    labs(title = "Age vs Cost", subtitle = "Seperated by gender", x = "Age of subscriber",
        y = "Total cost of claims", caption = "Graph D")


a + b
```