HRP 203: Module 3 Project

Valeria Gracia Olvera

GitHub repository

Introduction

Health care costs can be explained by factors such as age, gender, heart problems, and behaviors such as smoking. Estimating the effect of these factors on healthcare costs could help allocate resources in the healthcare system. In addition, it can also help estimate the economic impact of public health campaigns, such as smoking cessation campaigns.

The goal of this project is twofold: (1) estimate the effect of age, sex, heart problems, and smoking on health care costs; (2) evaluate the economic effect of smoking cessation campaigns.

Methods

Data

I used data provided in Assignment 7 which contains 5,000 simulated observations with `smoke`, `sex`, `age`, `cardiac`, and `cost` variables[1] which are defined in Table 1.

Table 1: Variable definition

Variable	Definition
cost	Healthcare cost per year in USD
age	Age in years
female	If =1 female
cardiac	If =1 has heart problems
smoke	If =1 smoker

[1] I assumed the definition and description of the variables by looking at the data, but I have no more information to confirm it.

Model

The relationship between the outcome and the explanatory variables can be written as:

$$cost_i = \beta_0 + \beta_1 age_i + \beta_2 I(female_i) + \beta_3 I(cardiac_i) + \beta_4 I(smoke_i) + \epsilon_i$$

where, i indexes by individual, $I(\cdot)$ represents an indicator variable, and ϵ is the error term.

Results

Descriptive statistics

Descriptive statistics for the two continuous variables and the distribution between smokers or non-smokers (`smoke`), those who have heart problems (`cardiac`) or not, by sex, are presented in Table 2. Note that the proportion of men and women who do not have heart problems is practically the same. On the other hand, smokers and men are more likely to have heart problems and have higher healthcare costs.

Table 2: Descriptive statistics by heart problems

Characteristic	**Yes**, N = 190	**No**, N = 4,810
smoke	77 (41%)	431 (9.0%)
sex		
female	19 (10%)	2,416 (50%)
male	171 (90%)	2,394 (50%)
age	39 (27, 51)	41 (30, 53)
cost	10,156 (9,859, 10,562)	9,651 (9,375, 9,902)

Regression results

I estimated the effect of age, sex, heart problems, and smoking on the cost of healthcare. Table 3 presents the regression results.

Table 3: Regression results

Characteristic	**Beta**	**95% CI**	**p-value**
age	18	18, 19	<0.001
female	-294	-305, -282	<0.001
cardiac	289	259, 319	<0.001
smoke	593	574, 611	<0.001

Figure 1 summarizes the relationship between the cost and the other variables of interest. We can see that the cost increases with age for men who smoke and have heart problems.

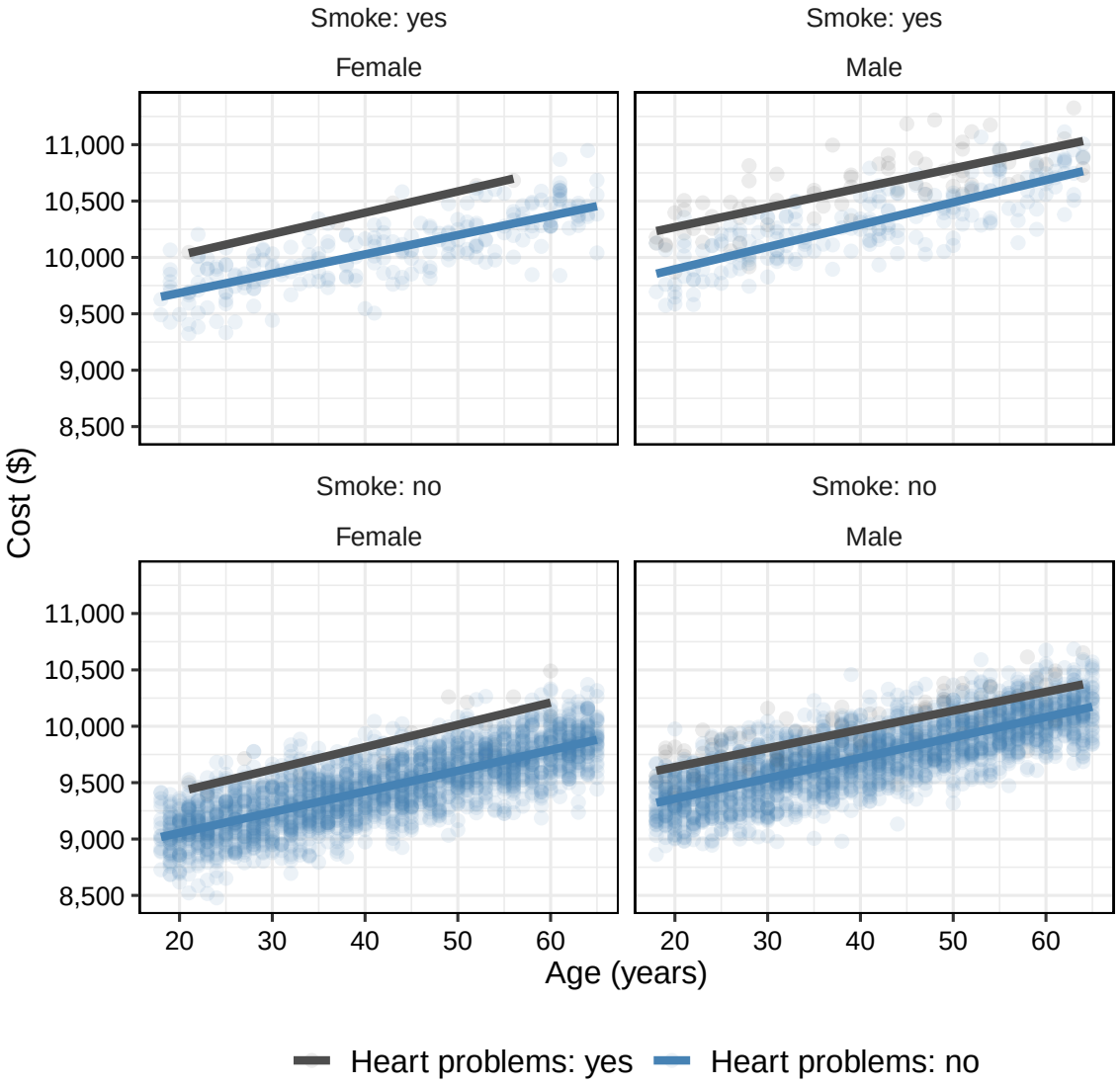

Figure 1: Relationship between cost and all the other variables

Results of a campaign to stop smoking

Finally, I am going to estimate the economic impact of a smoking cessation campaign under different scenarios:

Scenario A The campaign is 60% effective, i.e. 60% of smokers quit smoking.

Scenario B The campaign is 85% effective, i.e. 85% of smokers quit smoking.

I used the previously estimated model to predict the total cost in both scenarios and the status quo. I randomly selected 60% of the smokers in the sample and changed their smoking status to non-smoking for scenario A. I repeated this process now changing the smoking status of 85% of the

smokers in the sample for scenario B. Lastly, I calculated total healthcare costs by scenario and took the difference from the status quo to estimate the campaign's financial savings.

Results are presented in Figure 2. Note that the highest savings comes from Scenario B, where 255 thousand USD are saved with the smoking campaign compared to status quo.

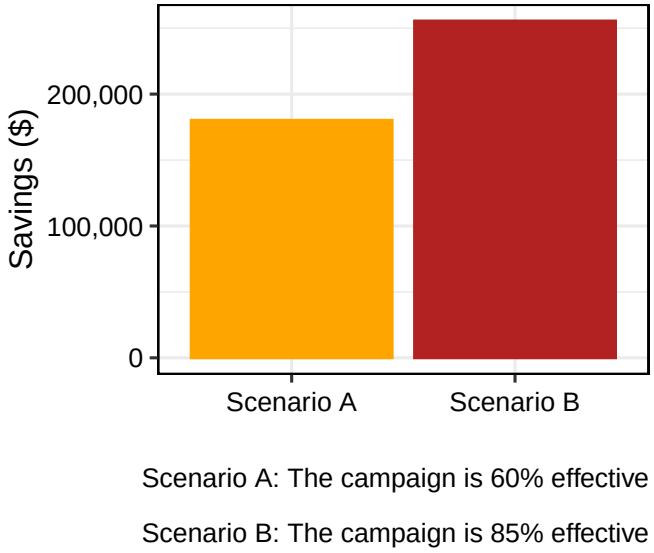

Figure 2: Economic impact of a smoking cessation campaign, savings compared to status quo.