Factors Determining Personal Health Insurance charges.

Zhihao Chen

4/27/2022

ABSTRACT.

The main aim of this paper was to determine relationship between personal health insurance charges and characteristics of the primary beneficiary of the insured. The demographic factors included; age, BMI, number of dependents and smoking status, while the geographic factor was the region of residence of the primary beneficiary. Cross-sectional data were obtained, and data description and visualizations were conducted to create a better insight into the data. The relationship between the variables was assessed using multiple regression techniques to achieve the aim of the study. Understanding the factors determining personal health insurance is vital for improving health and reducing longstanding differences in healthcare. There is no uniform amount of premium the insured is supposed to pay to the insurer, and the existence of these factors causes the variance in the monthly sum paid. It was found that the variables; age, BMI, smoking status and the number of children positively and significantly help in determining the amount of personal health insurance charges. These factors had a positive impact on the amount of payments. As the units of these variables increases, the amount of health insurance charges increases. Therefore, this study helped create a balanced criteria of determining the fair amount for both the insurer and the insured so that the money paid out will not exceed the money paid in. R-Studio software was used for analysis R Core Team (2022).

KEYWORDS.

Insurance, Insurer, Insured, Premiums, Dependent.

INTRODUCTION.

Personal health insurance is the coverage purchased mainly to protect an individual or a family when they are sick. Personal health insurance cover can be obtained through an employer or government programs such as Medicaid or Medicare. On the other hand, private insurance cover refers to that coverage offered by a private entity instead of the federal government Rampal et al. (n.d.). There are social determinants affecting the health of the most vulnerable populations. The more an individual's health is risky, the more they are likely to pay more premiums than an individual whose health is considered stable. There are various factors determining personal health insurance charges, and they include; the age of the primary beneficiary, body mass index (BMI), the number of dependents covered by the insurance, region of the beneficiary such as the U.S, sex of the insurance contractor, if the primary beneficiary is smoking.

There are some considerations an individual ought to advocate for the future premiums to be paid because the insurer takes into account the total premiums paid out incase of the insurance claim. In the United States, there have been efforts to improve the health conditions of its citizens by reconsidering the traditional health care system as the vital driver of the outcomes of health status across all states. This has signified the increased recognition concerning improving and attaining healthy equity that demands border approaches that address economic, social, and environmental factors influencing health. Various researchers argue that health care is vital to health, and most health outcomes are motivated by the underlying factors related to genetics, social environment, and health behaviors of an individual Viganò et al. (2000). Similarly, other researchers argue that some other health behaviors, such as smoking, are the major drivers of health outcomes. In general, those individuals who did not complete their education are more likely to stay in those environments that are hazardous to their health due to staying in substandard houses.

Understanding the factors determining personal health insurance is vital for improving health as well as reducing longstanding differences in healthcare. There is no uniform amount of premium the insured is supposed to pay to the insurer, and the existence of these factors causes the variance in the monthly sum paid. There is a maximum number the insured is supposed to indicate as the beneficiaries or the dependents of the policy, especially for the family but for the company, the number is dependent on the number of employees. Having a personal health insurance cover has numerous advantages, such as government income tax exemption, especially for the health benefits, as well as a wide range of options related to planning. Equally, an individual with personal health insurance cover is free from the high medical cost incurred by those individuals who don't have medical cover. The insured is likely to pay less for the covered health care before meeting the deductible amounts previously paid as premiums. The insured receives free preventive care such as screenings and routine checkups Zhang, Fu, and Fang (2020).

DATA.

```
##    'data.frame':                1338    obs.    of        7    variables:
##    $  age          :  num    19  18  28  33  32  31  46  37  37  60  ...
##     $   sex            :   chr     "female"   "male"   "male"   "male"   ...
##      $   bmi              :   num      27.9   33.8   33   22.7   28.9   ...
##      $   children:   num      0   1   3   0   0   0   1   3   2   0   ...
##       $    smoker       :    chr       "yes"    "no"    "no"    "no"    ...
[i]##   $ region   : chr   "southwest" "southeast" "southeast" "northwest" ...
##  $ charges : num  16885 1726 4449 21984 3867 ...
```

The main aim of this study was to predict the amount of personal health charges based on the primary beneficiary demographic and geographical information. The dataset was obtained from the Kaggle website (https://www.kaggle.com/code/harshinisk/health-insurance-prediction/data). It contained seven variables and 1,338 observations. The individual characteristics contained in the datasheet include; their age, sex, BMI, number of children and smoking status, while the geographical feature was the region of residence of the primary beneficiary.

The outcome variable “charges” and the predictor variables (age, BMI and number of children) were continuous, while the variables (sex, region and smoking status) were categorical. Based on these characteristics, an exploratory analysis was done. Finally, a multiple regression model was built to determine the relationship between the factors and predict the amount of health insurance charges.

The dataset consisted of 49.5% females and 50.5% males. This was illustrated in the pie chart below;

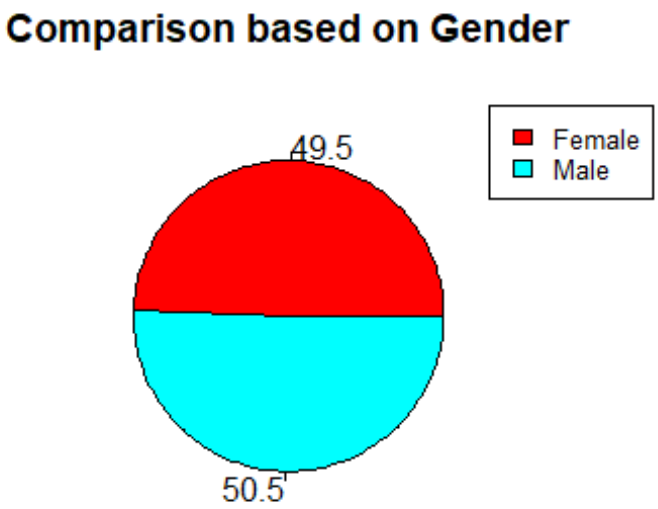

Figure 1: Data distribution per sex.

The youngest respondent was 18 years of age, while the oldest was 64 years old.

For the body mass index of the participants varied between 15.96 kg/m2 and 53.13 kg/m2.

Most of the respondents came from the south east while the participants from the other regions were equally sampled as indicated in the plot below;

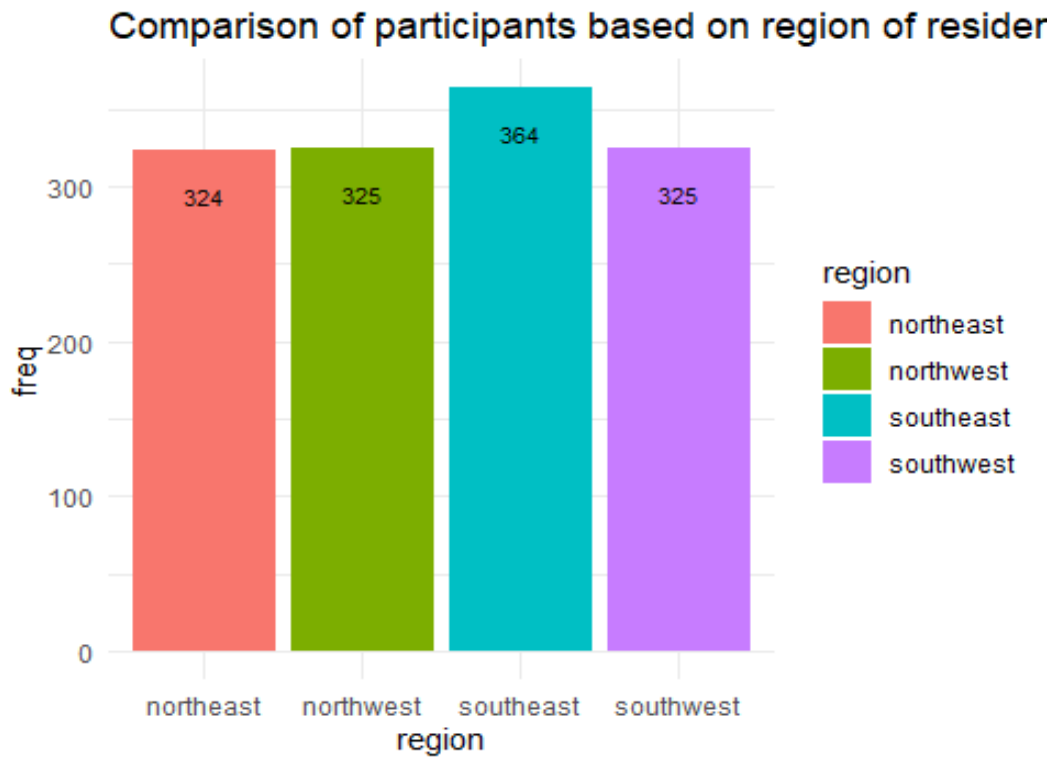

Figure 2: Distribution of respondents based of region of residence.

The dataset consisted of 79.5% non-smokers with only 20.5% of the respondents were smokers. The illustration was as below;

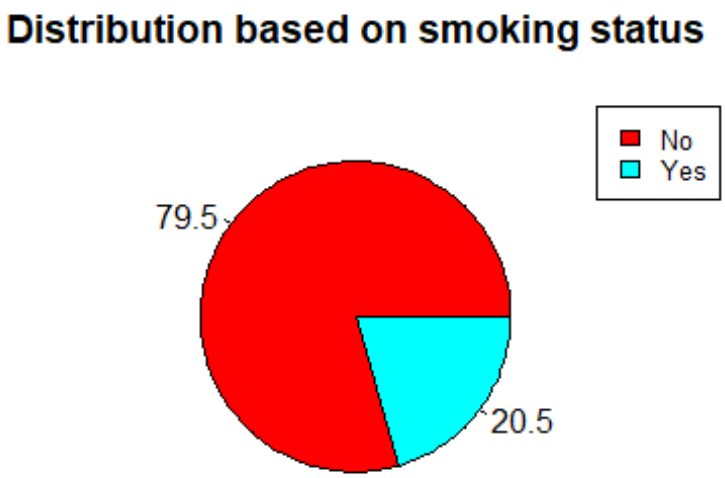

Figure 3: Smoking status among the participants.

From the bar chart below, most of the participants had no children.

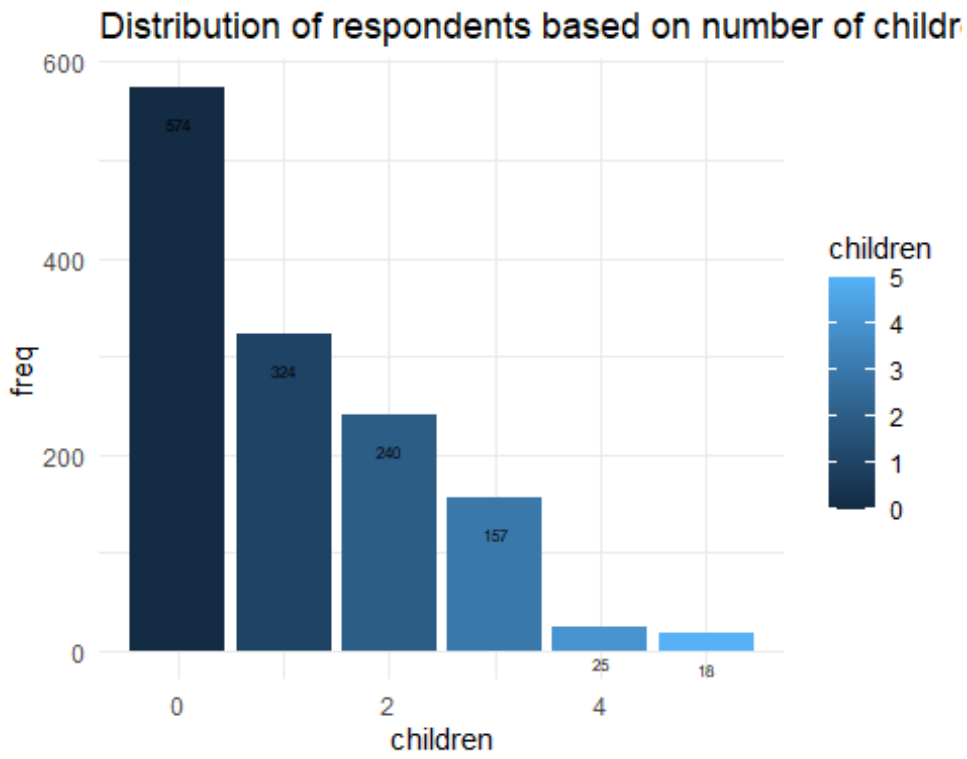

Figure 4: Distribution of number of dependents.

MODEL

Exploratory data analysis was conducted initially involving the investigations of the data to discover the patterns and checking for anomalies. These were achieved with the help of summary statistics and graphical representations. This process helps unearth the main characteristics of the data.

The dataset was first split into train and test sets. The training set was used to fit the model while the testing set was used to evaluate the model performance and make predictions. A multiple linear regression was considered as it analyzes the relationship between a single dependent variable and several independent variables.This aligned with this study varibales as there was one response variable and six independent variables.

The multiple regression model was formulated as follows; $Yi=\beta 0+\beta iXi+\varepsilon$ Where;

Yi; dependent variable.

Xi; explanatory variables (for i=1,…, 6).

β0; Y-intercept (constant term).

βi; Slope coefficient for explanatory variables.

ε; residual term.

RESULTS

Exploratory data analysis.

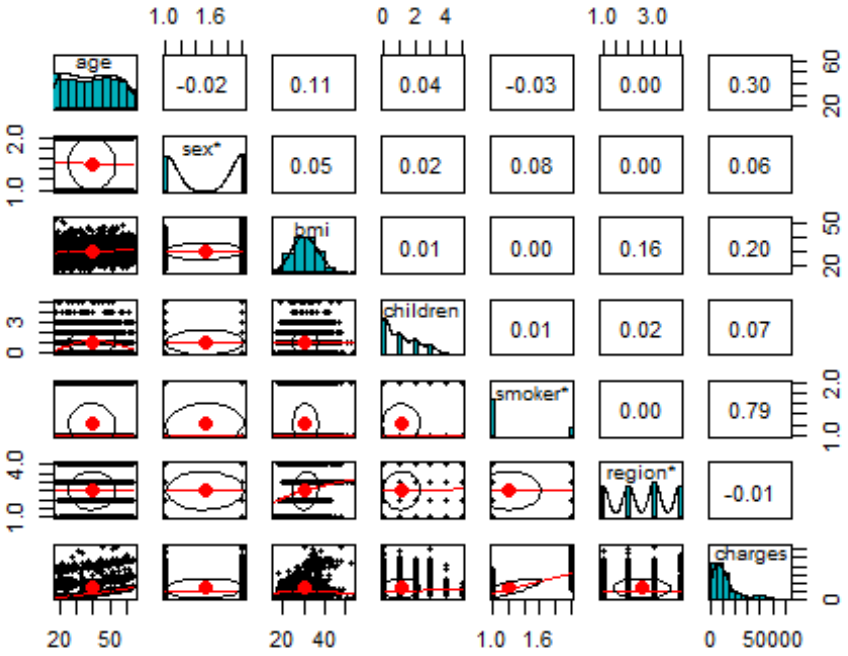

Figure 5: Scatter matrix for the study variables.

The above plot was used to access the association between the predictors and the outcome. As age and body mass index increase, the amount of health insurance charges also increases. Based on the correlation coefficients, there seemed to be a negligible relationship between the charges and the explanatory variables; sex, number of children and region residence.

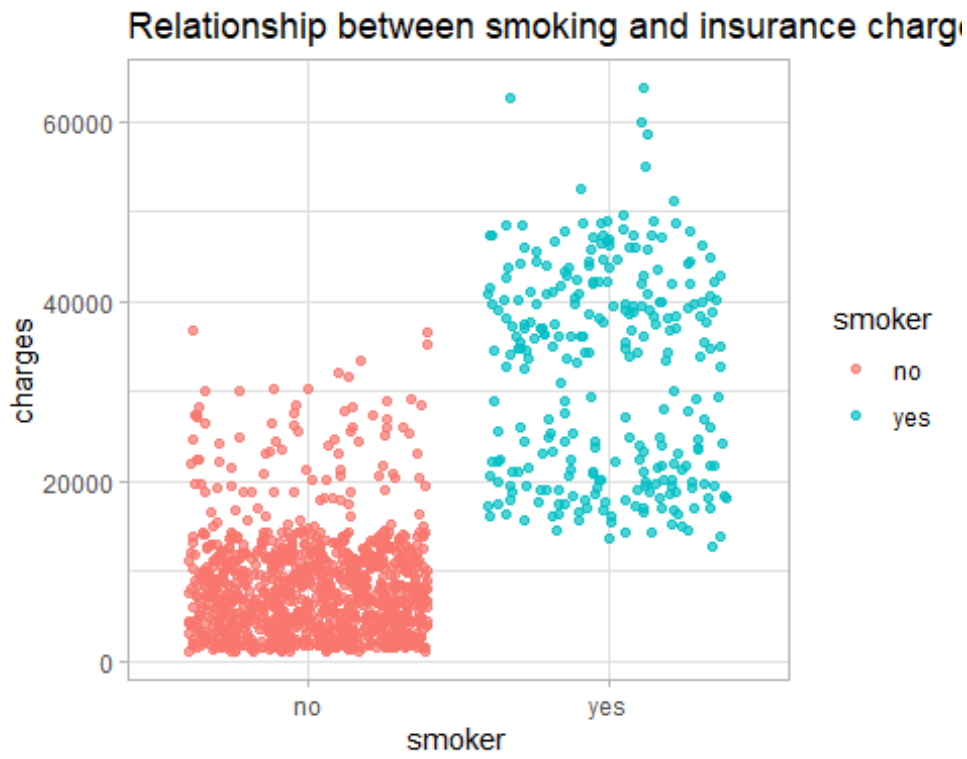

Figure 6: Correlation between smoking status and insurance charges.

Fitting Multiple regression model

From the initial model, the variable sex and region were found to be statistically insignificant. Therefore, the best model could be built by dropping the insignificant variables as conducted below;

```
##
##            Final            Model            Regression            Results
##                                    ======================================
##                                                      Dependent   variable:
##                                                        ------------------
##                                                                    charges
##                                    --------------------------------------
##   age                                                               267.875
##                                                                    (13.411)
##
##   bmi                                                               337.493
##                                                                    (30.553)
##
##   children                                                          613.235
##                                                                   (154.532)
##
##   smokeryes                                                      23,813.290
##                                                                   (460.364)
##
##   Constant                                                      -13,034.930
##                                                                 (1,047.882)
##
##                                    --------------------------------------
##   Observations                                                      1,070
##   R2                                                                0.752
##   Adjusted   R2                                                     0.751
##     Residual     Std.     Error                                   6,109.661
##   F   Statistic                                                     805.888
##                                    ======================================
## Note:                                 NA
```

The final model fitted above, had all the variables significant. The results indicate that for every additional year increase in the age of an individual, the health insurance charge increase by about 254.05 dollars. The insurance charges increases by approximately 295.17 dollars with a 1kg/m^2 increase in the individual's BMI. Smoking increases the amount of health insurance fee by about $24,203.62. Finally, for an additional child as a dependent increases the health insurance charge by 540.34 dollars.

DISCUSSION.

Based on the results above, it was found that the variables; age, BMI, smoking status and the number of children significantly help in determining the amount of personal health insurance charges. These factors had a positive impact on the amount of payments. The study notes that the health insurance charges increase with an additional year in the age of a primary beneficiary of health insurance. This could be because as a person become older, they get closer to their life expectancy and the risk taken by the insurer also increases, thus an increase in the charges. As the body mass index increases, it results in higher insurance charges. The main reason is that BMI helps indicate an individual's health status. A person with a higher BMI stands at a higher risk for diseases; therefore has to make regular visits to the hospitals for weight-related issues.

Personal health insurance charges was much higher for individuals who smoke than for non-smoking beneficiaries. This could be attributed to the belief that smokers are more likely to develop health issues, thus resulting in more risk for the insurer. Furthermore, a higher number of dependents (children) leads to higher health insurance charges because the number of individuals to be covered increases hence greater risk for the insurance company. This study helped create balanced criteria for determining the fair amount for both the insurer and the insured. This is because when individuals are charged based on the risk helps ensure that the money paid out will not exceed the money paid in.

REFERENCES.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rampal, Jubeena, Prabhdeep Singh, Rajbir Kaur, and Kirandeep Singh. n.d. "An Ensemble Model to Predict Health Insurance Premium Using Machine Learning."

Viganò, Antonio, Marlene Dorgan, Jeanette Buckingham, Eduardo Bruera, and Maria E Suarez-Almazor. 2000. "Survival Prediction in Terminal Cancer Patients: A Systematic Review of the Medical Literature." *Palliative Medicine* 14 (5): 363–74.

Zhang, Liangwen, Sijia Fu, and Ya Fang. 2020. "Prediction the Contribution Rate of Long-Term Care Insurance for the Aged in China Based on the Balance of Supply and Demand." *Sustainability* 12 (8): 3144.

---

[i] The data and codes can be found in the GitHub link:
https://github.com/ansonc1437/sta304-final-project.git