

**MIDDLE EAST TECHNICAL UNIVERSITY**

**GRADUATE SCHOOL OF INFORMATICS**

**COGS516 RESEARCH PROPOSAL**

**Health Insurances: Gender Rating Effect on Insurance  
Prices**

***By: Abdullah Burkan BEREKETOĞLU – 2355170***

***Submitted to: Barbaros Yet***

***Spring – April 28, 2022***

# Table of Contents

<b>Table of Contents</b>	<b>2</b>
<b>Introduction</b>	<b>3</b>
<b>Approach</b>	<b>3</b>
<b>Review</b>	<b>4</b>
<b>Bibliography</b>	<b>5</b>

- Approach is the proposal part which can be thought of as the one-page proposal, rest is a literature review that is held before the work.

# Introduction

After the industrial revolution, with the increase of big cities and fast population growth, civilizations passed on a new generation of humane issues to be upheld<sup>[1]</sup>; some of these particular issues are held by private entrepreneurs with the rise of the United States. These entrepreneurs started insurance companies for certain humane matters that the government should provide by following specific government regulations and procedures that protect the consumer from the abuse of the private company. Later on, in the EU/EEA, new, more humane insurance policies and systems models were initiated<sup>[1]</sup>. Insurance, specifically health insurance, after the mid-20th century, initialized with its new model for newly emerging economic areas<sup>[1]</sup>. Countries such as Switzerland had a 100-year programmed system that continuously updated to the new age of the 21<sup>st</sup> century, and it is suggested in the Orlu et al.<sup>[1]</sup> to be adopted by Turkey due to this research to be concluded, which is gender-based insurance policies.

American Center for Disease Control and Prevention Center (CDC) suggests that particular disability and risk factors can be used like the following price policy and coverage procedure for insurances, and they can be named; Alcohol usage, Illicit Drug usage, body measurements (height, size, etc.), mandatory diet, Certain disabilities, and physical or mental function issues, Exercise or physical activity capacity, Obesity/BMI, Smoking<sup>[2]</sup>. Even to date, many private and social health insurance companies still use other parameters such as age to calculate a price. They include different premiums for different age groups specially tailored to the consumer group of the specific age groups<sup>[3]</sup>. One of which of these factors is *gender*, even today by many countries except the EU and certain states of the United States of America (Montana is the first to prohibit the usage of gender as an insurance price factor.)<sup>[4]</sup>.

In this project, the direct gender effect will be excerpted to understand the importance and whether it is needed to be a significant factor in health insurance policies. For the analysis, a data set called *Medical Cost Personal Dataset - Health Insurance Cost Prediction (Insurance) data* by Amy Aguirre is used with ~ 1300 samples<sup>[5]</sup>.

## Approach

In the project, it is planned to use the following approach to attack the problem of whether gender is a crucial indicator that should be used in the determination of price by making a directed acyclic graph of dependencies and pulling out the gender as the independent factor to affect the cost of insurance.

In the *Medical Cost Personal Dataset - Health Insurance Cost Prediction (Insurance) data*<sup>[5]</sup>, there exist precisely 1338 data with a slightly imbalanced gender ratio of consumers given with a balance of 51:49 (male: female)<sup>[5]</sup>. The features of the data-set can be provided as; age, sex(gender), BMI, children count of the person, whether the person smokes or not, their region, and the price for the total of 6 features given<sup>[5]</sup>.

Data will be separated into different regions with separate data to eliminate the region dependency. For the small or imbalanced data, simple monte Carlo methods or oversampling methods with weights taken into account would be added to the data's preprocessing. Later on, it is planned to continue with creating the directed acyclic graph (DAG) to show the relations between the variables as the prior belief. For now, it seems as but will be analyzed with regular statistical analysis to see the correlation between gender and BMI, age and BMI, and a mixture of the three. Also,

find the correlation between male-female differences and smoking, smoking, and BMI. It is also will be noted that children can also be one of the factors to be taken into account in the correlation of BMI for gender regardless of male or female for the beginning; the correlation limit in this experiment will be determined by the famous economic principle called Pareto Principle which indicates that eighty-percent of the cases will be driven by twenty-percent of the indicators, so if the correlation is below twenty-percent correlation will be neglected and reported in the analysis. A DAG will be drawn after the pre-processing and primary statistical analyses finish to have a prior belief. With the DAG relation, a Bayesian data analysis model to make the direct effect analysis of the gender to price or the so-called name here charges.

In the model, the Bayesian multiple linear regression method will be used. For more intricate systems for some steps other than multiple linear regression, splines can be used after the behavior of the data in regular statistical analyses is determined. If the charge is assumed that behave logarithmically, exponentially, or in a polynomial shape, etc., a different system rather than linear regression without separating data into different sections would be more desirable. This behavior will be reported in the study if any such behavior exists in the analyses in the pre-process.

If the behavior is linear for the features for the charge, then the multiple linear regression method is to be followed. Else non-linear approaches such as splines are to be followed.

The charge predictions for gender are to be visualized with summary statistics, and the highest density interval for 91 percent interval is given. 91 is picked because it is below 10 percent from tails with a slight difference. For the last step, it is to be concluded with the difference between males and females to be simulated the difference statistics as the main result to be reported after specific interventions to make the gender charge independently from the rest.

## Review

In the literature, it is seen that age and gender play a significant role in the determination of the prices of health insurance <sup>[3]</sup>. Still, many contrary regulations started to be integrated into the governmental systems against the latter, gender, playing a role in the health insurance prices and premiums. A private institution with the closing gap in premiums between the government-issued health insurance and the gender gap is now over in the EU faces a drastic decline in people who buy and continue private insurance coverages. Also, with the increasing costs, many exits the entire healthcare insurance system<sup>[6][7]</sup>.

This price gap mentioned in the system is according to one of the insurance companies run in India named, Gender Rating, and still in use in India<sup>[7]</sup>. Turkey also still has this gender gap, but women after a certain age have the risk of breast cancer and other reproductive system problems<sup>[1]</sup>. By Merzel<sup>[8]</sup>, it is also discussed that in low-income zones such as Central Harlem, NYC, people with low-income most of the time cannot afford private coverage, but females have the opportunity to have it covered when they are in the workforce, with that being said, females do not get

affected by low-income or socioeconomic factors. On the other hand, male counterparts are heavily influenced by fewer available opportunities even in the workforce due to companies not offering them insurance making them less covered within the year 2000<sup>[8]</sup>. Literature states that certain states in the United States of America and after Obamacare now have better coverage for people from different socioeconomic backgrounds<sup>[4]</sup>. It is stated in the 2008 Los Angeles Times article that being female increases the insurance price rates drastically<sup>[9]</sup>. It is still for every one in ten women affects women between 19 to 64 age, which consists of nearly 98 million people in the United States<sup>[10]</sup>. In Taiwan, elderly or mid-aged women have different cancer insurance policies with lower claim rates for dread diseases which shows that the system is biased towards gender/age<sup>[11]</sup>. In the IFFCO-TOKIO factors article, it is stated that ten factors affect the health insurance premiums in most insurance policies. These are the recent day's age, medical history, occupation, policy duration, BMI, smoking, and location; as stated in Orlu, prices are determined by gender. Still, the premiums are not as given in the IFFCO-TOKIO article<sup>[1][12]</sup>. Lastly, Huang and Salm mentioned that the ban on gender-based pricing increased the number of people who buy health insurance, even though the new prices are slightly lower, but not drastically<sup>[13]</sup>.

## Bibliography

- [1] C. ORLU and A. KÖSE, “Türkiye ve isviçre Genel Sağlık SİGORTASI SİSTEMLERİNİN Kiyaslanması ve türkiye için YENİ BİR model önerisi,” *Finansal Araştırmalar ve Çalışmalar Dergisi*, pp. 319–345, 2018.
- [2] CDC, “Disability and risk factors,” *Centers for Disease Control and Prevention*, 12-May-2016. [Online]. Available: <https://www.cdc.gov/nchs/fastats/disability-and-risk-factors.htm>. [Accessed: 27-Apr-2022].
- [3] D. H. Yamamoto, “Health care costs from birth to death,” *Age-Curve-Study*, Jun-2013. [Online]. Available: [https://healthcostinstitute.org/images/pdfs/Age-Curve-Study\\_0.pdf](https://healthcostinstitute.org/images/pdfs/Age-Curve-Study_0.pdf). [Accessed: 26-Apr-2022].
- [4]A. Fontinelle, “Gender and insurance costs,” *Investopedia*, 26-Apr-2022. [Online]. Available: <https://www.investopedia.com/gender-and-insurance-costs-5114126>. [Accessed: 27-Apr-2022].
- [5] A. Aguirre, “Medical Cost Personal Dataset - Health Insurance Cost Prediction (Insurance) data.” *Kaggle* [Online]. Available: <https://www.kaggle.com/code/annetxu/health-insurance-cost-prediction/data>. [Accessed: 27-Apr-2022].
- [6] M. Chernew, D. M. Cutler, and P. S. Keenan, “Increasing health insurance costs and the decline in insurance coverage,” *Health Services Research*, vol. 40, no. 4, pp. 1021–1039, 2005.

- [7] Future Generali, “How gender impacts health insurance premium ,” *Future Generali India LifeInsurance*, 11-Feb-2022. [Online]. Available: <https://life.futuregenerali.in/life-insurance-made-simple/life-insurance/how-health-insurance-premium-varies-by-gender/#:~:text=Gender%20Rating&text=%27%20Insurers%20operate%20on%20risk%2Dbased,to%20bill%20that%20insurance%20policy>. [Accessed: 27-Apr-2022].
- [8] C. Merzel, “Gender differences in health care access indicators in an urban, low-income community,” *American Journal of Public Health*, vol. 90, no. 6, pp. 909–916, 2000.
- [9] D. Lazarus, “Gender can cost you in individual health insurance,” *Los Angeles Times: Gender can cost you in ... - csun.edu*, 22-Jun-2008. [Online]. Available: <https://www.csun.edu/pubrels/clips/June08/06-23-08Y.pdf>. [Accessed: 27-Apr-2022].
- [10] KFF, “Women's health insurance coverage,” KFF, 08-Nov-2021. [Online]. Available: <https://www.kff.org/other/fact-sheet/womens-health-insurance-coverage/>. [Accessed: 27-Apr-2022].
- [11] C.-S. Li, C.-J. Hung, S.-C. Peng, and Y.-L. Ho, “Impact of gender and age on claim rates of Dread Disease and cancer insurance policies in Taiwan,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 1, p. 216, 2021.
- [12] “10 factors that affect your health insurance premium costs,” *Motor, Health, Travel, and HomeInsuranceOnline*. [Online]. Available: <https://www.iffcotokio.co.in/health-insurance/10-factors-that-affect-your-health-insurance-premium-costs>. [Accessed: 27-Apr-2022].
- [13] S. Huang and M. Salm, “The effect of a ban on gender-based pricing on risk selection in the German health insurance market,” *Health Economics*, vol. 29, no. 1, pp. 3–17, 2019.