INFO 5082: FINAL PROJECT

STATISTICAL ANALYSIS OF HEALTH QUALITY AND EXPENDITURE

By,

Gollapally Shashank Reddy

Under the guidance of

Junhua Ding

Introduction

A safe country is a prosperous nation, as we all know from the proverb "Health is Wealth." The standard of a country's healthcare system is regarded as one of the most important factors that determine its state, as well as a significant contributor to the economy. Statistical modeling aids us in determining where our research should be directed and identifying the variables that influence healthcare quality and costs. In addition, there are several problems to address when maintaining a country's health quality. One of these issues is the cost of health care. Our research and analysis are primarily focused on determining how to reduce healthcare costs while preserving or improving health quality.

The data collection and data source are crucial in the research. Since analysis is based on data, the data used must be reliable; otherwise, the analysis would be useless. To avoid such an occurrence in this paper, we created our own dataset using data from reputable sources such as the OECD and WHO, as well as literature review.

After gathering and analyzing data, we hope to gain a better understanding of the factors that influence life expectancy and health spending. We also want to make recommendations for lowering health-care costs and improving the country's health-care quality. The statistical data modeling entailed creating multiple models and determining the most unbiased model that generalizes and provides feasible suggestions for lowering health-care costs and improving health-care quality.

The main takeaway from our models is that rising health services such as hospitals and medical graduates will increase a country's life expectancy. If a larger percentage of the population is covered by public insurance, the country's healthcare costs will be reduced to the greatest extent possible.

Problem definition & significance

We would like to present our findings and recommendations to OECD policymakers and senior health officials. Even though all of the countries under consideration are industrialized, they have struggled to develop good strategic health policies, resulting in high expenditure over time with little change in quality.

Our model advises these countries on the factors they should consider when formulating policies to improve the quality of health care while keeping costs down. This is an intriguing issue that could explain why the United States spends about $10,000 per citizen on average, compared to $3000 in other OECD countries.

Prior Literature

Health quality is a significant indicator of a country's overall health. Health quality can be calculated on a variety of scales, with the most significant metrics being infant, child, and adult mortality, as well as life expectancy. In contrast to Adult or Child, which only accounts for a portion of the life cycle, Life Expectancy encompasses mortality during the life cycle.

According to historical statistics, life expectancy has increased dramatically in the twentieth century. According to studies, technological advances, rising income levels, and countries' healthcare spending are all major determinants of the increase. Increased demand and need for healthcare accessibility resulted in an increase in healthcare jobs and better health outcomes. Education reforms improved sanitation and personal health. Deaths from various causes such as diseases, on the other hand, reduced the average number of years in a person's life. To access Life Expectancy, we aggregated the above factors in the available format in this model. Aside from efficiency, health expenditure is an important factor in determining health-care policies and allocating resources.

One of the main indicators affecting a country's health expenditures is the availability of health services. Medical practices and supplies, according to Health Economics, are the major determinants of health expenditure development. The age structure of a country's population is often used to discern its health characteristics. According to a WHO study of health funding, high-income countries have a 0.477 positive impact on expenditures. The use of primary care gatekeepers seemed to reduce health-care spending. The provision of health care by the government was linked to lower health spending.

Data Preparation

The information we used came from the OECD [5] and WHO [6] websites. From 2010 to 2015, data was obtained for 20 different variables in 39 different countries. We conducted a literature review to identify factors that explain variation in health expenditures and expenses. For the purposes of assessing health quality and expenditure, we used a country's life expectancy as an indicator of health quality and health expenditure in USD as our dependent variables.

The dataset includes various variables such as number of hospitals, hospital employment, medical graduates, nurse graduates, total equipment, social factors such as mean schooling years, the fatalities which includes deaths due to respiratory, circulatory diseases, cancer and accidents and economic factor such as expenditure per capita of the country to measure the Life expectancy i.e., health quality model. To analyze the health expenses model, independent variables such as insurance type, social factors such as population structure (over 65 years), medical procedures, and health facilities available such as total equipment, hospitals, and hospital jobs are considered.

We imputed null values with the countries that have equivalent expenditure and by searching for specific country websites after collecting the data. Our key assumption is that most countries with similar expenditures would have similar values for the other variables. All of the data has been scaled to the population of the country in question. The variables and their scales are listed below.

Variable	Units
Diagnostic_Exams	per 1000 population
ExpenditurePerCapita	$ / capita
Hospitals	per million population
Life_Expectancy_at_Birth	years
Percpopulationulationabo	percent above 65
Physicians	per 1000 population
Private_Insurance	% of total population
Public_Insurance	% of total population
death_by_cancer	per 1000 population
death_by_circular	per 1000 population
death_by_accident	per 1000 population
Mean_Schooling_Years	years
NationalIncome	$ / capita
hospital_employment	per 1000 population
tot_equipment	per 1000 population
medical_grads	per 100,000
nurse_grads	per 100,000

The data was cleaned in Microsoft Excel and R, and the data from various sources was merged in R studio using various libraries. R and Power BI were used to create the visualizations.

Variable Selection and Hypothesis

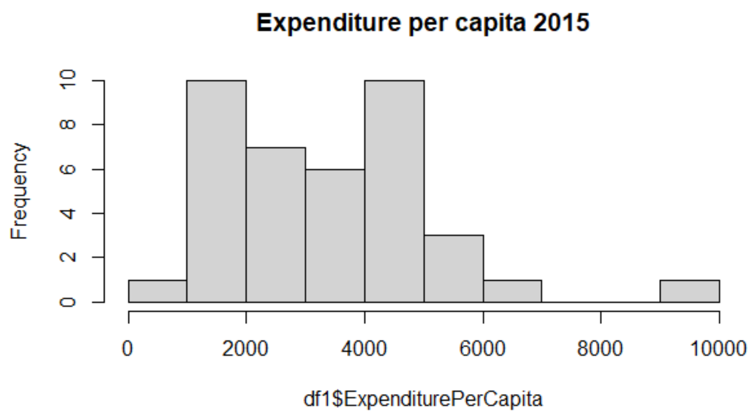
In this section, we will look at the factors that affect the target variables, such as "Life Expectancy," which is a statistical measure of how long an organism is expected to live, and "Health Expense," which is the cost of a country's medical expenses per capita, and how these factors affect the target variables.

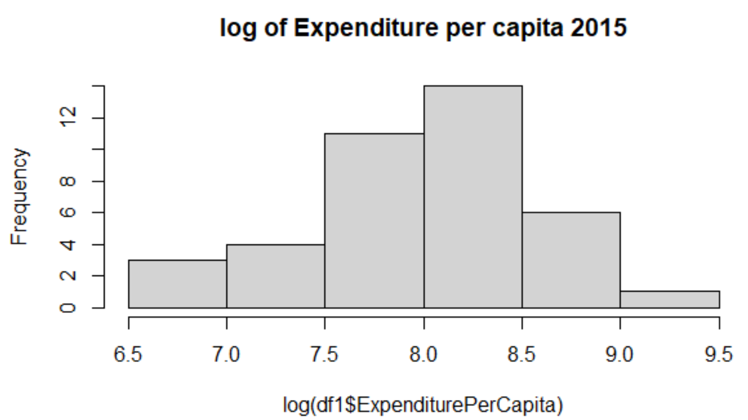
When developing the hypothesis, we want to concentrate on the actionable variables. The actionable variables are those of which the government has control and can make strategic decisions to increase life expectancy or lower health-care costs. Based on our domain awareness and literature review, we discovered that the following variables, “Life Expectancy” and “Healthcare Expense,” can be used to predict our target variables.

Variables		Hypothetical Description
1. Health Resources	a. Hospitals	With more hospitals in a country, more patients can be treated with less waiting time, and therefore a country's life expectancy can improve. The cost of health care may rise initially as a country's hospital population grows, but it may decrease in the long run.
	b. Hospital Employment	The greater number of employees in hospitals, the better and instant care to the patients, hence it might increase the life expectancy. The greater number of employees, their payment will increase and hence the health expense might increase.
	c. Total Equipment	With more equipment, the people can be diagnosed or treated with less waiting time and hence might increase the life expectancy. The greater number of equipment, might increase the health expense, as need to pay the initial cost and the maintenance cost for those.
	d. Medical Graduates	The increase in medical graduates, will increase the upcoming medical force of a country, so better and instant assistance to the people and hence it might increase the life expectancy.
	e. Nurse Graduates	The increase in nurse graduates, will increase the upcoming medical care force of a country, so people can get better and instant care and hence it might increase the life expectancy.
2. Health Quality	a. Death by respiratory disease	Now a days, most of the people are suffering from respiratory diseases, because of increase in smoking practices & pollution, hence the more death count by these diseases can result in decrease in life expectancy of a country.
	b. Death by circulatory disease	Many people are suffering from circulatory diseases because of unhealthy food habits & lack of agile practices, hence the more death count by circulatory diseases, can result in decrease in life expectancy of a country.
	c. Death by Cancer	As the death count increases by cancer, because of unhealthy food practices and decrease in immune systems, hence the life expectancy might decrease.
	d. Death by accidents	As the usage of vehicles are increasing, the deaths by accidents is also increasing in each country and hence it can result in decrease of life expectancy of a country.
3. Social Factors	a. Mean Schooling years	With the increase in literate people, the hygienic practices and primary care increases and hence the life expectancy might increase.
	b. Population Structure (Age>65)	With the higher number of aging populations, the health resources should increase, as they need more assistance and frequent care and hence it might increase the health care expenditure.
4. Economic Factors	a. Expenditure per capita	With increase of health care expenses, life expectancy might increase, as the health resources per capita increases, but excessive spending can result in wastage of money and might not help in improving life expectancy.
5. Medical Procedures	a. Diagnostic Exams	With the increase in number of diagnostic exams, the payment for these exams increases and hence the health expenses might also increase.
6. Insurance	a. Public Insurance	The premium cost of public insurance is comparatively low and hence increase in public insurances, might decrease healthcare expenses.
	b. Private Insurance	The premium cost of private insurance is very high and hence increase in number of private insurance holders, might increase healthcare expenses.

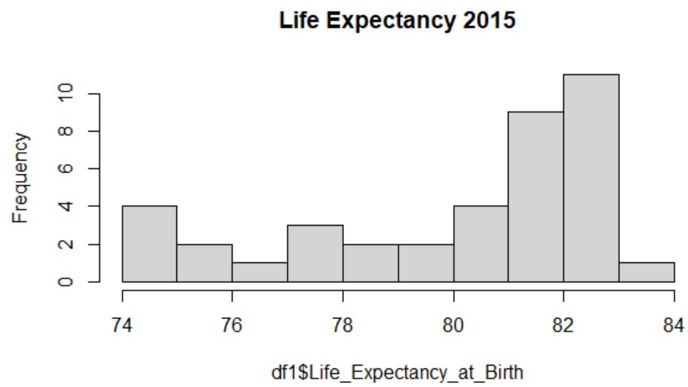
Exploratory Data Analysis

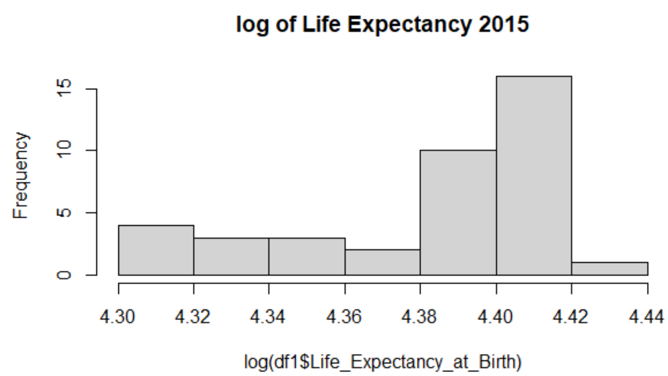
For 2015, the relationship between dependent variables (Life Expectancy and Expenditure per capita) is depicted in the graph above. In comparison to other countries, the United States has high costs ($10,000) and a Life Expectancy of 80 years.



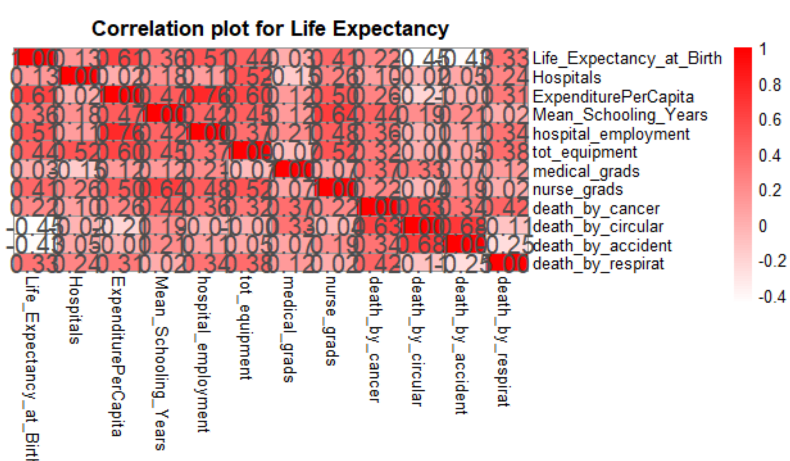


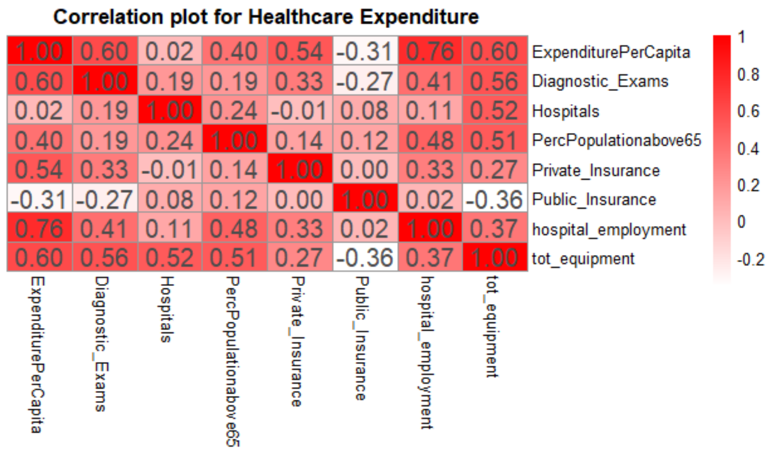
Since our dependent variable "Expenditure per capita" is skewed to the right in the graph above, we applied Log transformation to it, which not only removed skewness but also improved the interpretability of the variable in the model.





We can see that the data is left skewed in the above map. The use of transformations will not improve the skewness of the data.





Model 1 and Model 2 have a strong correlation between their dependent variables. We found no significant correlation (0.8) between the variables under consideration, ruling out the possibility of dependent variable multicollinearity.

Models and Results

We designed the models in R, after cleaning and merging the data with the different columns.

Life Expectancy:

Panel statistical models will fit appropriate because we have data from 2010 to 2015 (time variant) and for different countries (cross sectional). So, using life expectancy as the dependent variable and 11 predictor variables, we created three panel models: model 1 is a simple OLS model/pooling, model 2 is a fixed effects panel model, and model 3 is a random effects panel model.

Pooling Model, Model 1:

```
pooling_LE<-plm(Life_Expectancy_at_Birth~Hospitals+ExpenditurePerCapita +
    Mean_Schooling_Years+ hospital_employment+tot_equipment+medical_grads
    +nurse_grads+death_by_cancer+death_by_circular+death_by_accident+death_by_respi
rat,data=dle,model = "pooling")
```

Fixed Model , Model 2:

```
pooling_LE_Fixed<-plm(Life_Expectancy_at_Birth~Hospitals+ExpenditurePerCapita +
    Mean_Schooling_Years+ hospital_employment+tot_equipment+medical_grads
    +nurse_grads+death_by_cancer+death_by_circular+death_by_accident+death_by_respi
rat,data=dle,model = "within")
```

Random Model, Model 3:

pooling_LE_Random<-plm(Life_Expectancy_at_Birth~Hospitals+ExpenditurePerCapita + Mean_Schooling_Years+ hospital_employment+tot_equipment+medical_grads +nurse_grads+death_by_cancer+death_by_circular+death_by_accident+death_by_respirat,data=dle,model = "random")

Summary of Models:

The coefficients of the fixed effect and random effects models are constant, and the R2 value is relatively good compared to the other two models. Fixed effects models explain the exact nature of the relationship to our initial hypothesis, and the Hausman test indicates Fixed model is good.

```
> stargazer(pooling_LE,pooling_LE_Fixed,pooling_LE_Random,type ='text')

==============================================================================
                                      Dependent variable:
                     ---------------------------------------------------------
                                    Life_Expectancy_at_Birth
                           (1)                  (2)                 (3)
------------------------------------------------------------------------------
Hospitals                 -0.012              0.067**              0.002
                         (0.013)             (0.029)              (0.022)

ExpenditurePerCapita      -0.0001             0.0001               0.0002
                         (0.0001)            (0.0001)             (0.0001)

Mean_Schooling_Years      -0.002              0.265**              0.267***
                         (0.069)             (0.107)              (0.094)

hospital_employment       0.164***            0.288***             0.162***
                         (0.033)             (0.069)              (0.049)

tot_equipment             0.019***            0.100***             0.012
                         (0.005)             (0.022)              (0.008)

medical_grads             0.018               0.042*               0.064***
                         (0.033)             (0.024)              (0.024)

nurse_grads               0.003               0.009**              0.008*
                         (0.006)             (0.004)              (0.005)
```

death_by_cancer	1.327***	0.663***	0.606***
	(0.108)	(0.223)	(0.160)
death_by_circular	-0.750***	-0.200*	-0.477***
	(0.072)	(0.109)	(0.082)
death_by_accident	-4.546***	-1.266*	-1.869***
	(0.808)	(0.709)	(0.717)
death_by_respirat	-2.108***	-0.295	0.256
	(0.308)	(0.348)	(0.315)
Constant	78.042***		71.997***
	(0.786)		(1.197)
Observations	234	234	234
R2	0.787	0.541	0.509
Adjusted R2	0.776	0.419	0.485
F Statistic	74.552*** (df = 11; 222)	19.727*** (df = 11; 184)	230.248***
Note:			*p<0.1; **p<0.05; ***p<0.01

We can infer the effects of these variables on life expectancy as follows from the fixed effect model.

1	Variables	Impact
2	Hospitals	1 hospital per ten thousand persons,life Expectancy will increase by 6.7 years.
3	Expenditure per capita	increase in 10000$ expenditure per capita, Life Expectancy will increase by 1 years.
4	Mean_Schooling_Years	increase in 1 years of mean schooling per person, life expectancy will increase by 0.265 years
5	hospital_employment	incorporating 10 more employees in health sector per 1000 people. We can increase the life expectancy by 2.8
6	tot_equipment	increase of 100 equipment's per 1000 population, we can increase the life expectancy by 10
7	medical_grads	1 medical graduate per 1000 population, can increase the life expectancy by 4.2 years
8	nurse_grads	1 nurse graduate per 100 population, can increase the life expectancy by 9 years
9	death_by_cancer	increase of 10 deaths by cancer per 1000 population, life expectancy will decrease by 6.6 years
10	death_by_circular	increase of 10 deaths by circular per 1000 population, life expectancy will decrease by 2 years
11	death_by_accident	increase of 1 death by accidents per 1000 population, life expectancy will decrease by 1.2 years
12	death_by_respirate	increase of 1 death by respiratory per 1000 population, life expectancy will decrease by 0.2 years

According to the fixed effect model's interpretation, an increase in the number of hospitals, medical graduates, and nurse graduates have the highest impact on life expectancy, while deaths due to any illness have a severe negative impact on life expectancy.

In addition, for the Fixed level model on Life expectancy model, here are the fixed effect coefficients for different countries.

```
> fixef(pooling_LE_Fixed)
      Australia         Austria         Belgium          Brazil          Canada 
         54.955          60.575          59.111          65.917          61.710 
          Chile        Colombia      Costa Rica  Czech Republic         Denmark 
         69.052          64.498          69.824          61.781          55.263 
        Estonia         Finland          France         Germany          Greece 
         61.648          58.483          62.181          57.472          57.212 
        Hungary         Iceland         Ireland          Israel           Italy 
         60.484          59.454          64.119          69.116          60.538 
          Japan           Korea          Latvia       Lithuania      Luxembourg 
         47.907          55.938          56.349          58.031          64.924 
         Mexico     Netherlands     New Zealand          Norway          Poland 
         66.486          61.672          63.526          60.622          63.521 
       Portugal Slovak Republic        Slovenia           Spain          Sweden 
         64.227          60.928          64.477          66.345          59.999 
    Switzerland          Turkey  United Kingdom   United States 
         57.907          65.407          63.374          45.803 
```

Health Expense Model:

We used an exponential model for health spending, and the log(Health Expense perCapita) has a normal distribution, as shown in figure 3 of the visualization section. We created a panel model for health expenses, similar to the life expectancy model, since the data is available from 2010 to 2015 (time variant) and for different countries (Cross-sectional).

Pooling Model, Model 1 :

```
pooling_HE<-plm(log(ExpenditurePerCapita)~Diagnostic_Exams+Hospitals+
        PercPopulationabove65+ Private_Insurance+ Public_Insurance +
        hospital_employment+tot_equipment ,data=d,model = "pooling")
```

Fixed Model, Model 2 :

```
pooling_HE_Fixed<-plm(log(ExpenditurePerCapita)~Diagnostic_Exams+Hospitals+
        PercPopulationabove65+ Private_Insurance+ Public_Insurance +
        hospital_employment+tot_equipment,data=d,model = "within")
```

Random Model, Model 3 :

```
pooling_HE_Random<-plm(log(ExpenditurePerCapita)~Diagnostic_Exams+Hospitals+
        PercPopulationabove65+ Private_Insurance+ Public_Insurance +
        hospital_employment+tot_equipment,data=d,model = "random")
```

Summary of the Models:

```
> stargazer(pooling_HE,pooling_HE_Fixed,pooling_HE_Random,type="text")

=================================================================================
                                       Dependent variable:
                      -----------------------------------------------------------
                                    log(ExpenditurePerCapita)
                               (1)                        (2)              (3)
---------------------------------------------------------------------------------
Diagnostic_Exams             0.0004                    0.001***          0.001***
                            (0.0003)                   (0.0003)          (0.0002)

Hospitals                   -0.014***                   -0.006          -0.009***
                            (0.002)                    (0.005)           (0.003)

PercPopulationabove65        0.015***                   0.047***          0.045***
                            (0.005)                    (0.010)           (0.007)

Private_Insurance            0.005***                   0.001             0.004***
                            (0.001)                    (0.002)           (0.001)

Public_Insurance             0.002                      -0.010**          -0.005*
                            (0.002)                    (0.004)           (0.003)

hospital_employment          0.058***                   0.044***          0.046***
                            (0.004)                    (0.010)           (0.007)

tot_equipment                0.005***                   0.009***          0.003**
                            (0.001)                    (0.003)           (0.001)

Constant                     6.335***                                     6.716***
                            (0.176)                                      (0.310)

---------------------------------------------------------------------------------
Observations                  234                        234               234
R2                           0.820                      0.603             0.641
Adjusted R2                  0.814                      0.509             0.630
F Statistic         146.752*** (df = 7; 226) 40.876*** (df = 7; 188) 402.883***
=================================================================================
Note:                                             *p<0.1; **p<0.05; ***p<0.01
```

We can see that the coefficients of the fixed effect and random effects models are constant, and that the fixed effect model describes the precise nature of the relationship to our initial hypothesis, and that the Hausman test indicates that the fixed model is good.

We can infer the impact of these variables on life expectancy as follows from the fixed effect model.

	VARIABLES	IMPACT ON HEALTH EXPENSES
2	Private_Insurance	with increase in 10% percentage of population under private insurance coverage. The expenses increases by 1%.
3	Public_Insurance	with increase in 1% percentage of population under public insurance coverage. The expenditure decreases by 1%.
4	Hospitals	increasing hospitals by 1 per 1000 person, The healthcare expenses decrease by 6%.
5	Diagnostic_Exams	with increase in 1 diagnostic exams per 1000 population, the healthcare expenses increases by 1%.
6	PercPopulationabove65	with increase in 1% percentage of population above 65 years of age. The cost increases by 4.7%.
7	hospital_employment	with increase in 1 employee in health sector per 1000 population. The healthcare expenses increases by 4.4%.
8	tot_equipment	with increase in 10 equipment's per 1000 population. The healthcare expenses increases by 0.9%.

According to the above interpretation of the fixed effect model, increasing the percentage of people with public insurance and hospitals has the greatest impact on lowering health expenditure, while promoting the percentage of people with private insurance has the greatest impact on increasing health expenditure.

Quality Checks:

Life Expectancy Model:

```
> plmtest(pooling_LE)

        Lagrange Multiplier Test - (Honda) for balanced panels

data:  Life_Expectancy_at_Birth ~ Hospitals + ExpenditurePerCapita +  ...
normal = 20.382, p-value < 2.2e-16
alternative hypothesis: significant effects
```

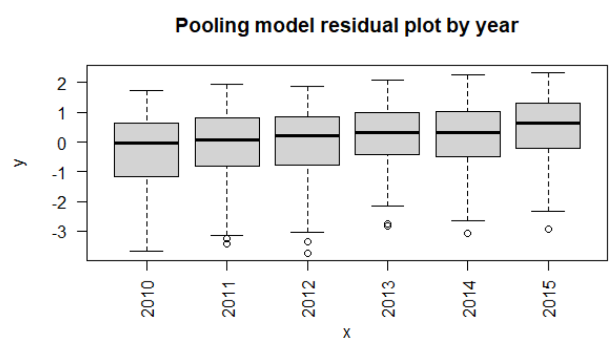
LMTEST is usually used to see if the data is panel. The null hypothesis is rejected with a p-value of 2.2e-16, indicating that the data is panel.

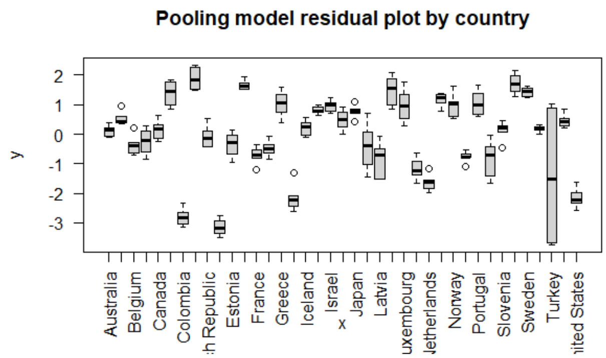
```
> phtest(pooling_LE_Fixed,pooling_LE_Random)

        Hausman Test

data:  Life_Expectancy_at_Birth ~ Hospitals + ExpenditurePerCapita +  ...
chisq = 27.091, df = 11, p-value = 0.004452
alternative hypothesis: one model is inconsistent
```

To compare Fixed and Random effects models, we used the Hausman Test. With a p-value of 0.0044, the Hausman test indicated that the fixed effect model is more consistent than the random effects model.

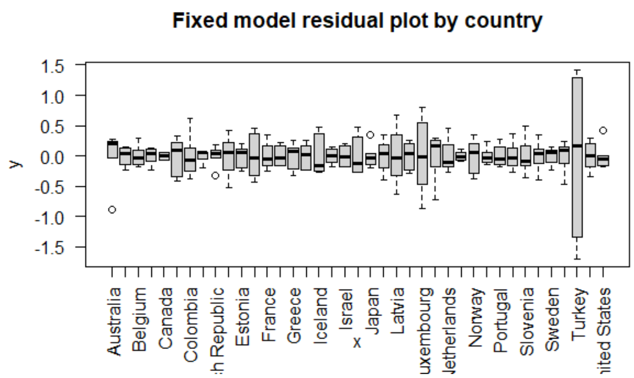
Residual Graphs:

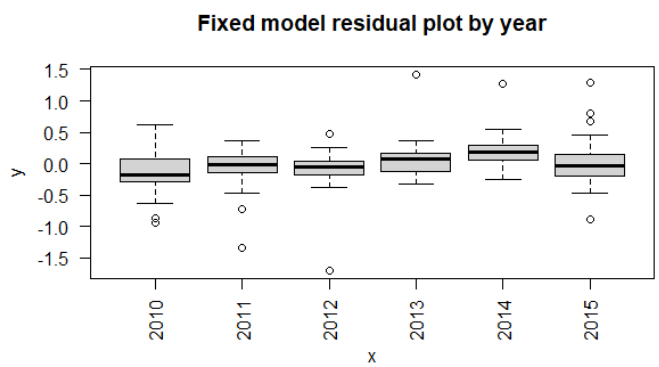
Pooling model:



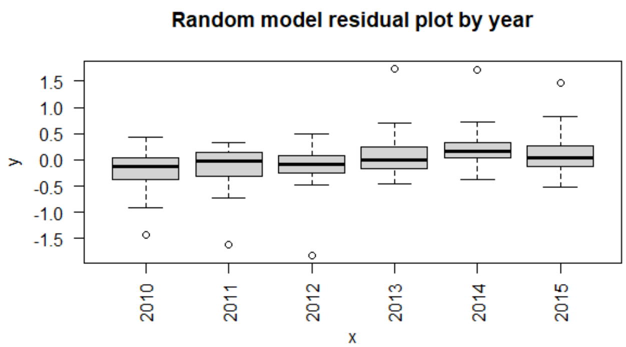


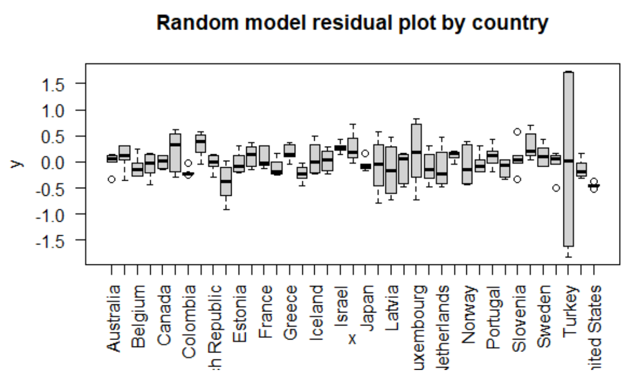
Fixed Effects Model:





Random Effects Model:





Fixed effects model residual graphs by time and country are more stable and around the mean zero than random effects and pooling model residual graphs.

Health Expenses model:

```
> plmtest(pooling_HE)

        Lagrange Multiplier Test - (Honda) for balanced panels

data:  log(ExpenditurePerCapita) ~ Diagnostic_Exams + Hospitals + PercPopulationabove65 +  ...
normal = 22.666, p-value < 2.2e-16
alternative hypothesis: significant effects
```

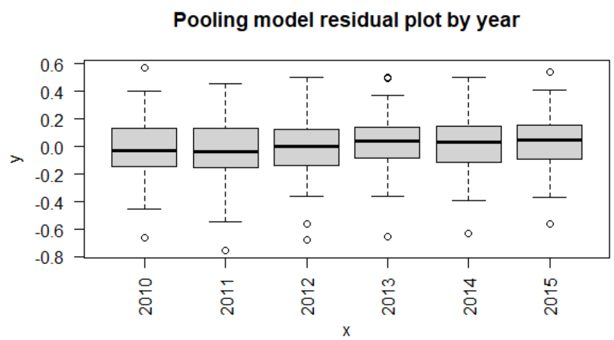
LMTEST is usually used to see if the data is panel. The null hypothesis that the data is not panel is rejected with a p-value of 2.2e-16, indicating that the data is panel.

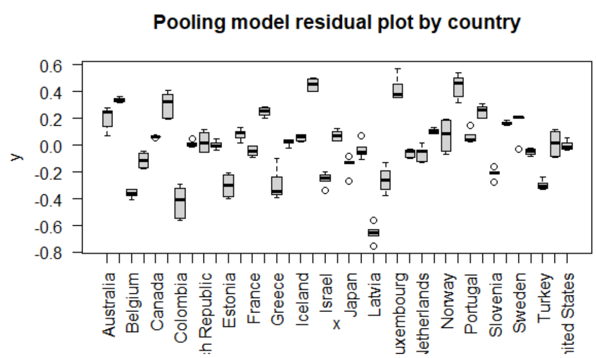
```
> phtest(pooling_HE_Fixed,pooling_HE_Random)

        Hausman Test

data:  log(ExpenditurePerCapita) ~ Diagnostic_Exams + Hospitals + PercPopulationabove65 +  ...
chisq = 18.896, df = 7, p-value = 0.008521
alternative hypothesis: one model is inconsistent
```

To compare Fixed and Random effects models, we used the Hausman Test. At a p-value of 0.0085, the Hausman test indicated that the fixed effect model is more consistent than the random effect model.

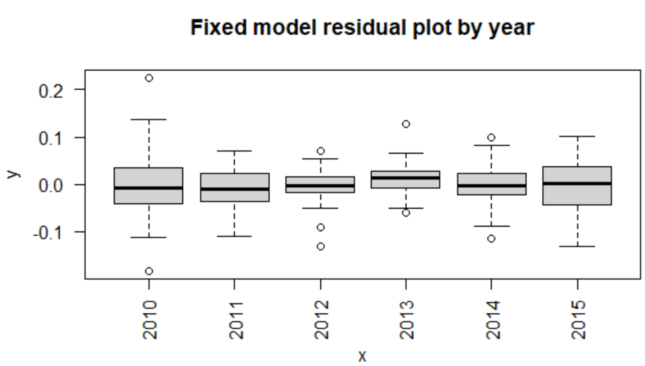
Residual Graphs:

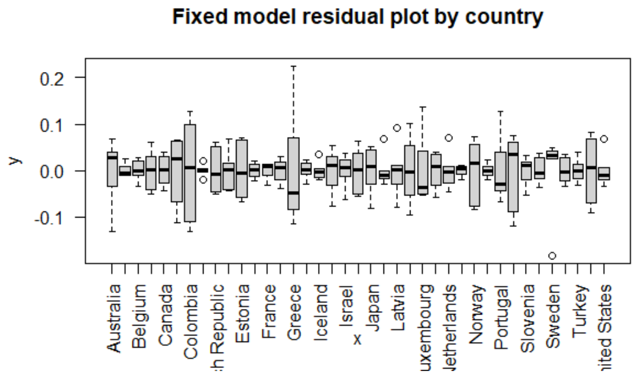
Pooling Model:



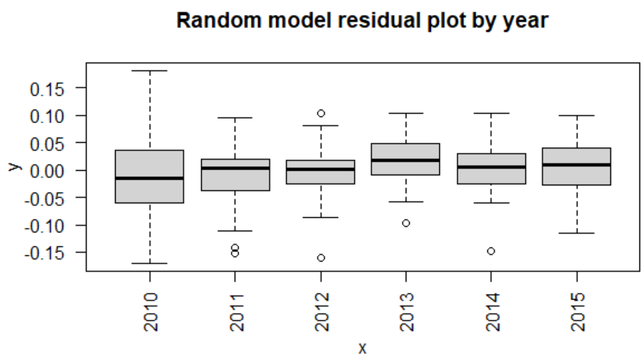


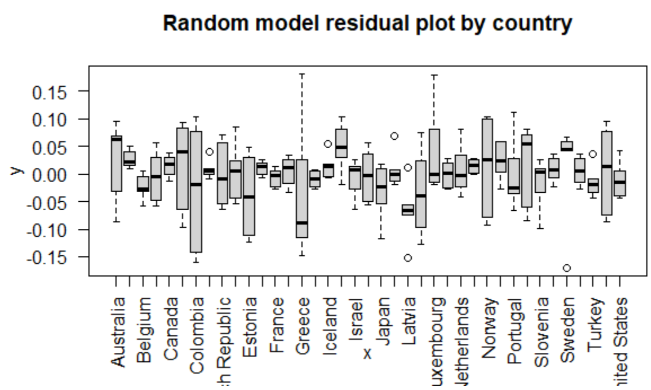
Fixed Effects Model:





Random Effects Model:





Fixed effects model residual graphs by time and country are more stable and around the mean zero than random effects and pooling model residual graphs.

Conclusion

- To raise life expectancy by 6.7 years, a quantitative measure for making this strategic decision is to build 10 hospitals per 1 million people.
- To better control health-care costs, the government can work to raise the percentage of the population covered by public insurance.
- A quantitative measure for making this strategic decision would be to compare a 1% reduction in costs per individual with a 1% increase in the population covered by public insurance.
- Also, we noticed that the unusual case of high health expenditures in the United States is mostly attributable to a low number of citizens covered by public insurance, i.e., if this percentage is raised, health costs will be reduced, and the recent presidential election 2020 campaign is all around “Medicare for All.” I hope this model clarifies Campaigners' point.

References

- Dataset, OECD : https://stats.oecd.org/Index.aspx
- Dataset, WHO : https://www.who.int/data/gho/
- Panel Data Modelling - Panel Data Analysis Fixed & Random Effects - Oscar Torres-Reyna
- Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data- Andrew Bell, Kelvyn Jones
- Dormont, B., Grignon, M. & Huber, H., 2006. Health expenditure growth: reassessing the threat of ageing. Health Economics, 15(9), pp.947-963.
- Dreger, C. & Reimers, H.E., 2005. Health care expenditures in OECD countries: a panel unit root and cointegration analysis.
- https://www.who.int/health_financing/documents/report_en_11_deter-he.pdf