Student Name: Krongpong Monpengpinij

Student ID: 6129003

This document contains the student extracurriculum activities.

Projectss/Topics:

1) Heptic VR Carido Vascular Access Training simulator
2) Thai National Foundation Projects
3) Report on LVEF.
4) Report on Elderly.

Heptic VR Carido Vascular Access Training simulator.

Author: Krongpong Monpengpinij

Student ID: 6129003

Send to:

Background:

I had the chance to collaborates with the faculty of ICT, Mahidol University to help them in one of their projects. In this project I am required to stay at Germany for about three months. This project is aim toward medical/dental students. Students of this field are required to practice on patient during their studies. Due to the student inexperience, they can accidentally harm the patient in the process of their treatment. We are exposing the patient to unnecessary risk of harm. The project is designed to reduce these risks. With the team from faculty of ICT, we are trying to develop a simulator in which the student can use to practice some of the common medical procedure. We have chosen catheterization procedure to be simulated. As this is a new project, my role is to identify all the important components which are to be included in the simulation. This involved me to do many researches. For this project, meeting is being held every two weeks. I am also required to present the progress to this meeting to the team. In addition, using various software, I also create a patient model that can be use in the simulation. We have submitted this project to the international symposium on ICT in Medicine and Public Health: Perspective from AI and Cognitive Science, we are awarded the best presentation award. Below are some details of this symposium:

Duration: 3 months.

Faculty of ICT, Mahidol University
Summer Intern 2019
Project Presentation Symposium

BREMEN SPATIAL COGNITION CENTER

Best Group Project Award

to

Project: **Haptic VR Cardio Vascular Access Training Simulator**

By: Krongpong Monpengpinit, Johannes Bunk

Prof. Dr. Peter Haddawy

Prof. Dr. Christian Freksa

Mahidol University

Bremen, August 2019

Universität Bremen*

Thai National Foundation Projects

Author: Krongpong Monpengpinij

Student ID: 6129003

Send to: Thai National Foundation

Background:

I had the chance to collaborates with the team at Thai National Foundation on their projects. My role here are data analyst and interpretation of the results. Their projects are based on Thailand education system. Scholarships for student are provided by governments, but the process of selecting a suitable
candidates and granting them scholarships are not well implemented. This is because the citeria filtering the application are not well designed, which as a result can introduce personal biases from the examiner in the process of granting students with scholarships. As this is an initial stage of the project, first we want to identify variables that has the most impact on the examiners decision on whom are suitable to be granted with scholarships. Through various processes, at the end we have identify various variebles, and interpretation has been made on why we think these variables are likely to influence the examiner's decision. This project has been selected as a presentation at eh 11th Asian Conference on Education, at Tokyo, Japan.

Duration: 2 months.

Date: 15 / 10 / 2019

Report- hours and school size

1 Introduction and Hypothesis

In this report we will explore the score on “read think write” and the school size. We would to see if the school size have any effects on the score. If so, what is the variables that are different in each school size. The school size are divided into three size according. The variable we will be looking at for each school size is the hours the student spent in class at school.

hypothesis;

1. Different school size yield significantly different outcome of “read think write” test. Where bigger school size will yield a better outcome.
2. The hours spent in class in each school size will effect the outcome.

The statistic will we be using is the ANOVA (one-way) test.

2 Descriptive

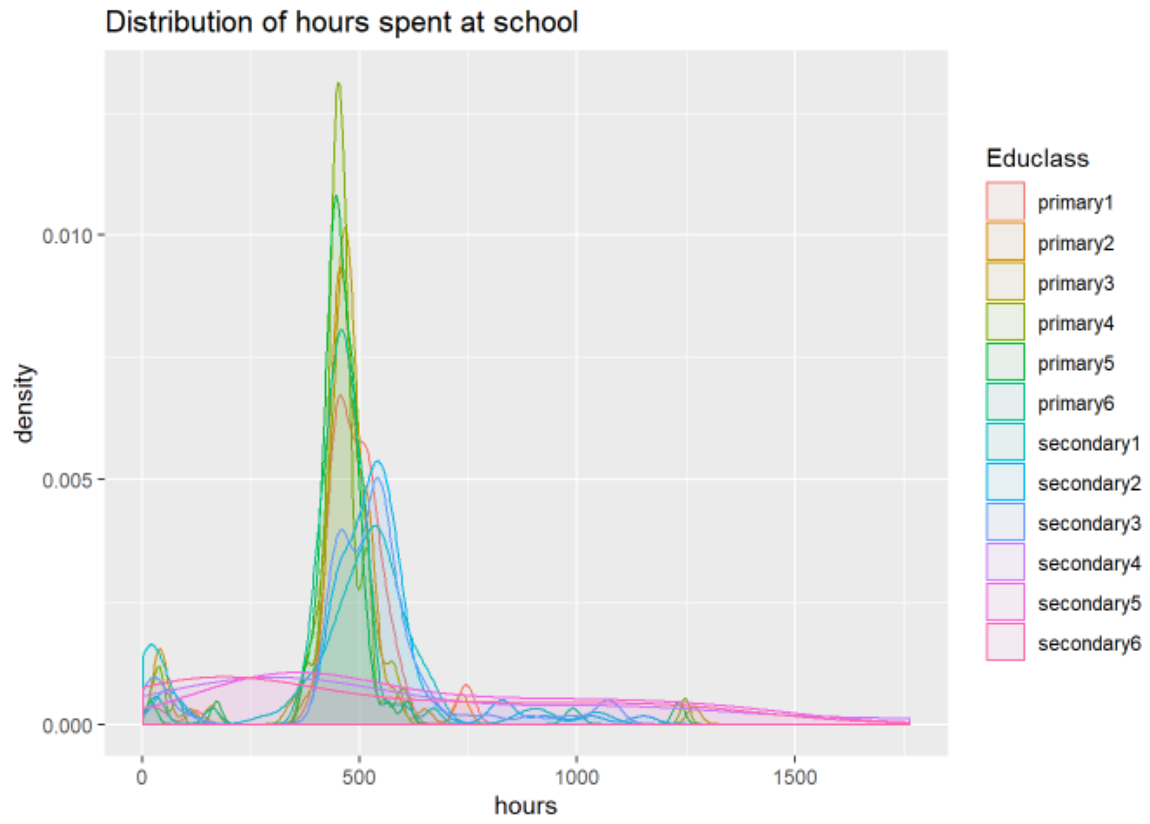
Mean of “read write think” score for different school size

```
## # A tibble: 3 x 7
##   School.Size   count rtw1_mean rtw2_mean rtw3_mean rtw4_mean rtw5_mean
##   <fct>         <int>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
## 1 School Size 1  7445      2.31      2.23      2.21      2.24      2.21
## 2 School Size 2  2810      2.30      2.23      2.19      2.17      2.15
## 3 School Size 3 10057      2.44      2.39      2.37      2.37      2.37
```

mean of hours spent for different school size

```
## # A tibble: 3 x 3
##   School.Size   count hours_mean
##   <fct>         <int>      <dbl>
## 1 School Size 1   356       458.
## 2 School Size 2    83       486.
## 3 School Size 3   279       511.
```

Distribution of hours spent at school



3 Statistic test

First we will test our first hypotheis - different school size yield different outcome as reflected on the score of “read write think”.

null hypothesis: The scores of “read write think” are not significantly different for each school size.

ANOVA-test

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## School.Size     2     85   42.50   187.1 <2e-16 ***
## Residuals   20309   4614    0.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## School.Size     2    129   64.25   290.3 <2e-16 ***
## Residuals   20309   4495    0.22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## School.Size     2    139   69.69   330.1 <2e-16 ***
## Residuals   20309   4288    0.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## School.Size     2    115   57.43   271.2 <2e-16 ***
## Residuals   20309   4301    0.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## School.Size     2    160   80.03   381.1 <2e-16 ***
## Residuals   20309   4264    0.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-values for each test are all less than 0.05. Therefore, we will reject null hypothesis.

Now testing our second hypothesis.

Null hypotheis: there is no significant hours spent in class for each school size.

ANOVA-test

```
##              Df   Sum Sq Mean Sq F value  Pr(>F)
## School.Size   2   444189  222095   6.194 0.00215 **
## Residuals   715 25635312   35854
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As our p-value is less than 0.05, we will reject our null hypothesis. This mean that there is a significant different in hours spent in class for each school size.

Post-hoc

```
##              Df   Sum Sq Mean Sq F value  Pr(>F)
## School.Size   2   444189  222095   6.194 0.00215 **
## Residuals   715 25635312   35854
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Only school size 1 vs school size 3 are significant from each other.

4 Conclusion

We established that the score of “read write think” are significantly different in each school size. Now if we want to factor in the hours hypothesis we must consider school size 1 vs school size 3 only, as it is the only that is significantly different from each other in term of hours spent. Follows from this statement, school size 3 students spent about 53 hours more in class than school size 1. The score of every “read write think” in school size 3 is about 0.14 higher than school size 1.

Report- province, school year and score.

1 Introduction and Hypothesis

In this report we will explore province and school year and its effect on the “read write think” score. We think that students in each province (surin and burirum) will have different “read write think” score. In addition, we will consider the effect of school year when combined with province too in this report.

hypothersis:

1. There are significant different in scores for each provinces.
2. School year also have effects on the score received.

Statistic use:

t-test

ANOVA (two-way (additive))

2 Statistical test

First we will test our first hypotheis - There is a significant different in scores for each provinces.

null hypothesis: The scores of “read write think” are not significantly different for the two provinces.

t-test

```
##
##  Welch Two Sample t-test
##
## data:  raw_data$rtw1 by raw_data$province
## t = -3.3925, df = 5504.5, p-value = 0.0006974
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.08098489 -0.02166673
## sample estimates:
## mean in group chonburi    mean in group surin
##               2.304693               2.356019
```

```
##
##  Welch Two Sample t-test
##
## data:  raw_data$rtw2 by raw_data$province
## t = -5.713, df = 5511, p-value = 1.168e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.11534210 -0.05640675
## sample estimates:
## mean in group chonburi    mean in group surin
##               2.203697               2.289571
```

```
##
##  Welch Two Sample t-test
##
## data:  raw_data$rtw3 by raw_data$province
## t = -10.077, df = 5618.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1755276 -0.1183570
## sample estimates:
## mean in group chonburi    mean in group surin
##               2.134318               2.281261
```

```
##
##  Welch Two Sample t-test
##
## data:  raw_data$rtw4 by raw_data$province
## t = -9.3466, df = 5545.1, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670871 -0.1091486
## sample estimates:
## mean in group chonburi    mean in group surin
##               2.160662               2.298779
```

```
##
##  Welch Two Sample t-test
##
## data:  raw_data$rtw5 by raw_data$province
## t = -9.1889, df = 5537.2, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1646256 -0.1067328
## sample estimates:
## mean in group chonburi    mean in group surin
##               2.131417               2.267096
```

The p-values for each test are all less than 0.05. Therefore, we will reject null hypothesis.

Now to test our second hypothesis.

Null hypotheis: School year does not have a significant effects on the score received.

ANOVA-tests (two-way (additive))

The p-values for each test are all less than 0.05. Therefore, we will reject null hypothesis.

Now to test our second hypothesis.

Null hypotheis: School year does not have a significant effects on the score received.

ANOVA-tests (two-way (additive))

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## province      1    4.5   4.469  12.143 0.000496 ***
## Educlass     11   33.5   3.044   8.271 1.35e-14 ***
## Residuals  7117 2619.1   0.368
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## province      1   12.5  12.509   34.50 4.45e-09 ***
## Educlass     11   40.0   3.633   10.02  < 2e-16 ***
## Residuals  7117 2580.2   0.363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## province      1   36.6   36.63   106.8 <2e-16 ***
## Educlass     11   51.3    4.67    13.6 <2e-16 ***
## Residuals  7117 2441.5    0.34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## province      1   32.4   32.36   93.64 <2e-16 ***
## Educlass     11   81.5    7.41   21.44 <2e-16 ***
## Residuals  7117 2459.6    0.35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## province      1   31.2  31.227   89.95 <2e-16 ***
## Educlass     11   64.2   5.839   16.82 <2e-16 ***
## Residuals  7117 2470.8   0.347
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

All of the p-values from the test above are less than 0.05, therefore we will reject our null hypothesis.

3 Conclusion

Firstly, we find out that each provinces have significantly different “read write think” score in all aspect. On average students in Surin will score more than students in Burirum by around 0.15 in all aspect. The result of our second hypothesis convey that each school year also plays a significant role in “read write think” score. However, we cannot conclude which variables (school year vs provinces) have more effects on the scores.

Report-pYing

1 Introduction

In this report we will explores different factors that might determine which student will be granted with a scholarship from the goverment. There are some criteria in which the goverment are looking for in candidates. However, observation seen from candidate who received the scholarship are not quite fit the criteria. In this report we will look at external factors such as family income and housing which might play an important role for someone to be granted a schorlarship.

Hypothesis: External factors and receiving a schorlarship are dependent to each other.

factors/ variables of interest:

1.GPA

2. Family
3. Household income
4. House ownership
5. Residence type
6. Housing

Statistical test:

chi-square test.

wilcox test.

2 Statistical Test

Wilcox test

GPA

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  gpa_data$GPA by gpa_data$Scholarship
## W = 14550, p-value = 0.7812
## alternative hypothesis: true location shift is not equal to 0
```

From the p-value, gpa between candidate who are granted with a scholarship and who dont are not significantly different.

Family income

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  income_data$Household_Income by income_data$Scholarship
## W = 1563, p-value = 0.001462
## alternative hypothesis: true location shift is not equal to 0
```

From the p-value, income for candidates who are granted with the scholarship differ significantly from candidates who does not get it.

Chi-square

Family

```
## Warning in chisq.test(family_data$Scholarship, family_data$Family): Chi-
## squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  family_data$Scholarship and family_data$Family
## X-squared = 48.536, df = 7, p-value = 2.797e-08
```

Houseowner ship

```
##
##  Pearson's Chi-squared test
##
## data:  ownership_data$Scholarship and ownership_data$House_ownership
## X-squared = 7.8834, df = 2, p-value = 0.01942
```

Residence type

```
## Warning in chisq.test(residence_data$Scholarship,
## residence_data$Residence_type): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  residence_data$Scholarship and residence_data$Residence_type
## X-squared = 8.8999, df = 3, p-value = 0.03065
```

Housing

```
## Warning in chisq.test(housing_data$Scholarship, housing_data$Housing): Chi-
## squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  housing_data$Scholarship and housing_data$Housing
## X-squared = 25.791, df = 5, p-value = 9.798e-05
```

From all the p-values, all of this external factors plays a significant role in determining who are most likely to get a scholarship.

3 Conclusion

Upon looking at the result we can see the government might not focus on the candidates performance at school, but their individual well being. That being said, people who got the scholarship have about 0.01 gpa score higher. Individual well being might reflect in other factors other than the gpa. Lets look at the income for example. People that gets the scholarship their family incomes are half of that people that did not get it. This might express some personal bias in determining whos getting the scholarship.

Report on LVEF

Author: Krongpong Monpengpinij

Student ID: 6129003

Send to: Kongkiat Kespechara (CEO of Bangkok Hospital Eastern group at Bangkok Hospital)

Background:

I had the chance to be a part of a project with a team of physicians from Bangkok Hospital Pattaya. In this project we want to find factors that influence left ventricular ejection fraction, an indicator uses to assess the condition of the patient heart, after they have cardiac arrest. Once we can find an impactful factor, we can then create a strategies or improve existing one in order to improve patient's left ventricular ejection fraction. My role in this project is data analyst, interpretation of the results and report to the team. This project is entitled: Clinical outcomes of strategy of reduction in door to balloon time.

Duration: 2 months.

Date: 13 / 05 / 2019

Report on LVEF

Krongpong

13/05/2019

1 Part1.

1.1 Objective (Part 1)

- Ojective 1. The objective of this part is to determine which variables (Door to balloon, Result, DM, HT, DLD, KILLIP, and Age) are suitable to use/ contribute more when predicting the LVEF values/group. We will look at each variable specifically and see which one are necessary to predict the LVEF value accurately. To know which variables contributed most when predicting LVEF is crucial since we can focus on that variable and can change it (value) if necessary to help improve LVEF outcome. Ultimately it help us to determine which factors/variables we need to focused on if we would like to improve the LVEF value of the patient.
- Objective 2. We also need to see the trend of these variables and its effect on LVEF value. For instance, if there is an improvement in LVEF values (e.g. “Normal”) with a decrease in Door to balloon time value, then we should try to achieve a decrease Door to balloon time as our goal. In this report we will look on the trend of a variable that make the LVEF worst. The first objective tell us which variables we should focus on, and objective 2 tell us what we need to do (course of actions) with this variables.

1.2 Method

- To achieve objective 1. A model is required to determine which variable mentioned above is most important when predicting the value of LVEF. The model I choose to use is the random forest model. This model will incorporate the variables Door to balloon, Result, DM, HT, DLD, KILLIP, and Age, which this model it will try to use it to predict the value of LVEF. To determine which one of these variables is most important in predicting accurate LVEF value, plot of graph of variables importance will be done.
- To achieve objective 2. Once the model is done and we determine which variables are most important in predicting the value of LVEF, we will then determine the effects of these variables on the LVEF value. This will be done by plotting the partial dependence graph.

1.5 Determining the variables that are suitable/most important in determining LVEF (Objective 1)

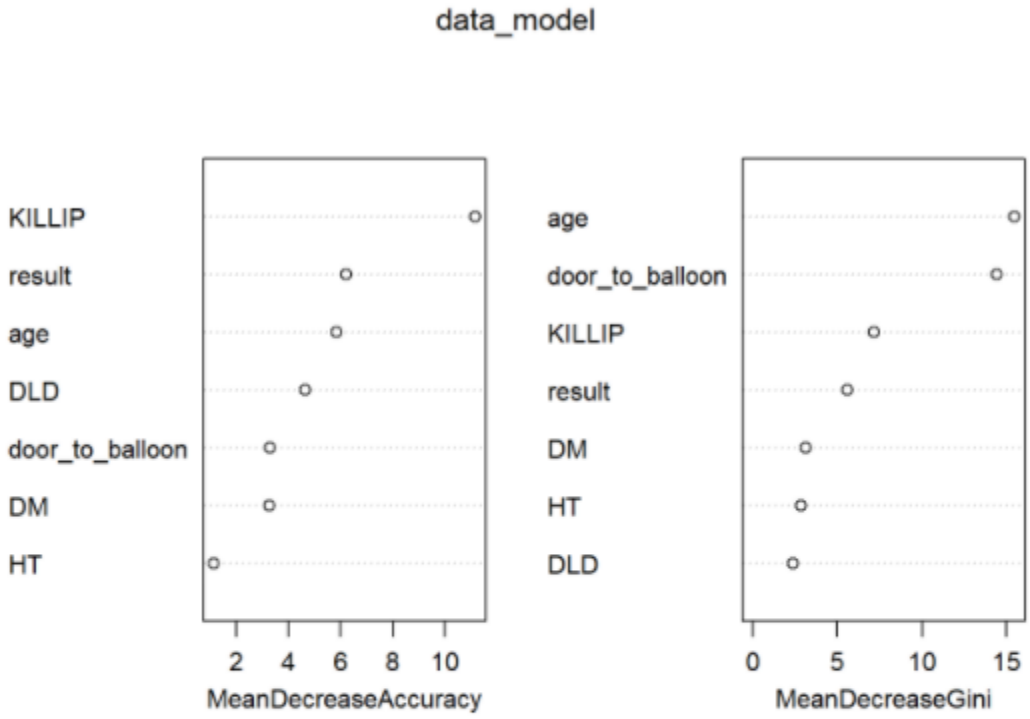
In order to achieve this, I have create a model (random forest model) using variables as mentioned before to predict LVEF. Here is the result below:

Model

```
##
## Call:
##  randomForest(formula = LVEF_group ~ ., data = train_data, importance
= T,      proximity = T)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 50.51%
## Confusion matrix:
##                          Moderately below normal normal
## Moderately below normal                       0      3
## normal                                        0     30
## Severely below normal                         0      9
## Slightly below normal                         0     21
##                          Severely below normal Slightly below normal
## Moderately below normal                     1                      0
## normal                                      1                     11
## Severely below normal                       7                      3
## Slightly below normal                       1                     12
##                          class.error
## Moderately below normal   1.0000000
## normal                    0.2857143
## Severely below normal     0.6315789
## Slightly below normal     0.6470588
```

Please note that this model is not that accurate in predicting LVEF, as a consequence might not reflect reality

Importance of Variables



- This is not align with our hypothesis, in fact Door to balloon time is not quite that important in predicting the LVEF value. This mean that we should focus on other variable if we want to make more impact on LVEF (value.)

We will focus on the left graph, this graph rank each variables of our interest on its importance on predict LVEF values. For instance, KILLIP is the most important variable and if we does not this variable in our model, the model will significantly decrease its accuracy on predicting the LVEF value. The top three most important variables are KILLIP, result, and age.

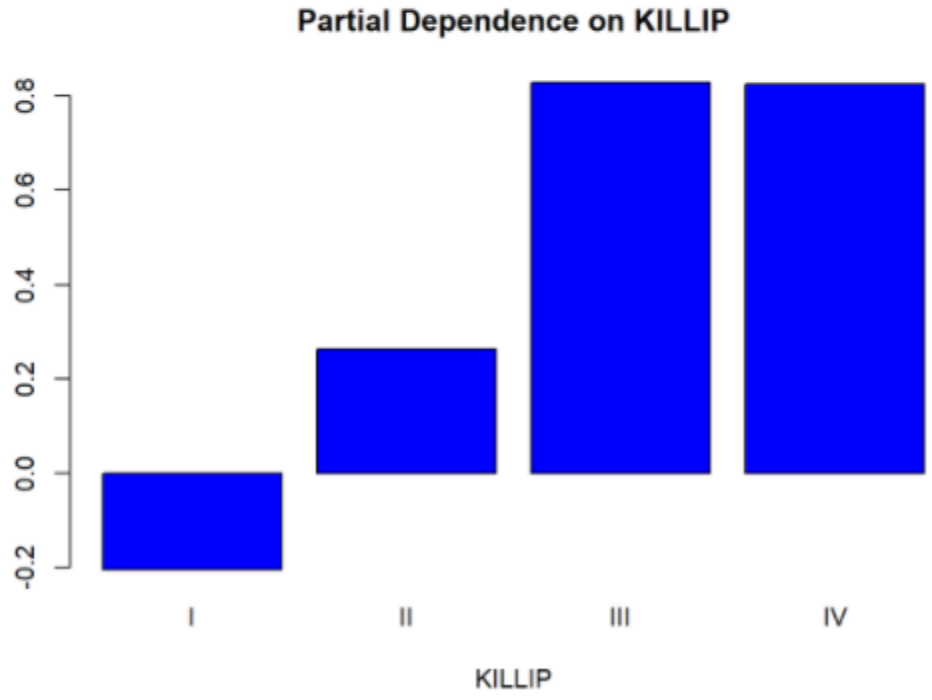
1.6 trend of some variables and its effect on LVEF value. (Objective 2)

Partial dependence plot will tell us how a given variable will affect the outcome of LVEF. In this specific report we will look at the trend and how it is like when the patient have worst LVEF variable (“severely below normal” group.)

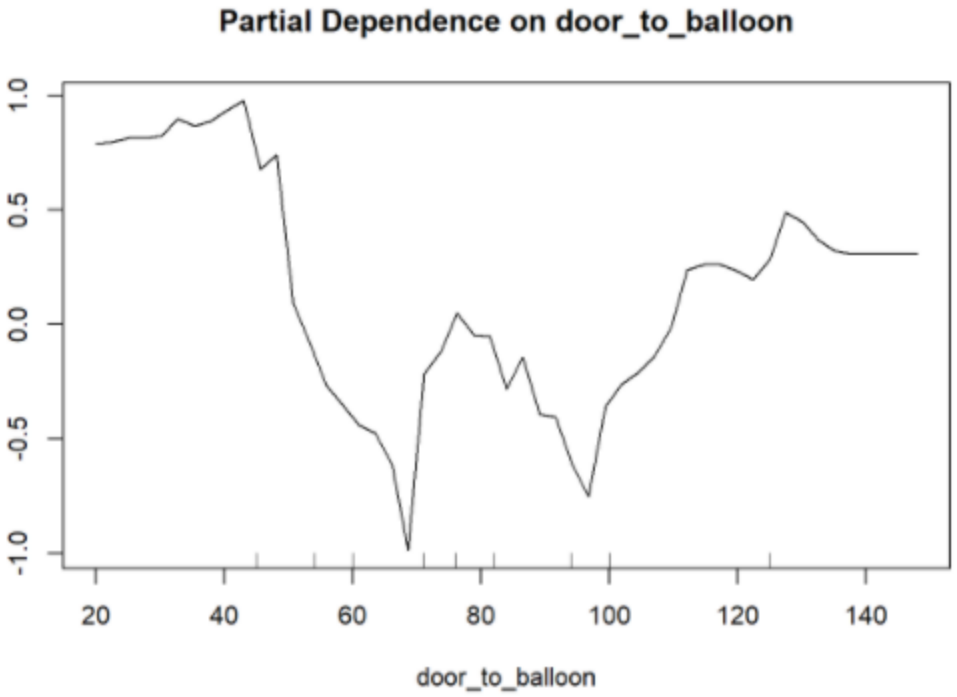
Let me explain the graph.

The y-axis is like probability, where the more positive the more likely a given outcome is going to happen, in this case LVEF is “severely below normal”. For instance, for KILLIP graph, group III and IV is more positive than other group this mean that people within this group will be morely likely to have “severely below normal” as their LVEF valule.

KILLIP



Door to Balloon



I would like to elaborate more on the graph above. The graph have this “W” shape which maybe very confusing. Therefore, I will break it into parts and explain it thoroughly.

Why this partial dependence graph is a line graph?

You can clearly see that this partial dependence graph is different for other partial dependence graph in that it is a line graph. This is because door to balloon time is a continous data type, it cannot be categorised like, for example on previous partial dependence graph on KILLIP. Therefore a line representation is more suitable.

Axis of the graph (Interpretation):

Even though it is represented by a line, however, the axis of this graph is no different from other partial dependance graphs. The intepretation is the same. the x-axis represent the value of the door to balloon time, which is in minutes. The y-axis tell us how likely a certain event is going to occurs. In this case is tell us how likely a person with a certain value of door to balloon time is going to have a “severely below normal” LVEF value. The higher the value of y-axis the more likely the event is going to occur, and in contrast, the lower the value of y-axis the less likely and event is going to occur. To give an example, let say person A have a y-axis value of -0.5 and x-axis (door to balloon time) of 70 minutes, and person B have a y-axis value of 0.0 and x-axis of 50 minutes. Then we would say that person B have a more likely hood of having “severely below normal” LVEF value, which we can then also associate this with the door to balloon time. In other words, people with 50 minutes door to balloon time are more likely to have a “severely below normal” LVEF value.

Scale of y-Axis of the graph:

Now you might be wondering does y-axis of 0.0 value mean that an event is not going to occurs? The answer is no, the y-axis is not probability but are similar. Everything is relative to one another, for instance, one point of door to balloon time is relative to another point of door to balloon time. This is why the scale of y-axis can be negative or zero. 20 - 40 minutes:

The line graph is going up, this indicate that from time 20 min to 40 min you will see more people with “severely below normal.”

40 - 70 minutes:

The line graph is going down, this indicate that from time 40 min to 70 min you will see less people with “severely below normal.”

70 - 75 minutes:

The line graph is going up, this indicate that and from time 70 min to 75 min you will see more people with “severely below normal.”

75 - 95 minutes:

The line graph is going down, this indicate that and from time 75 min to 95 min you will see less people with “severely below normal.”

more 95 minutes:

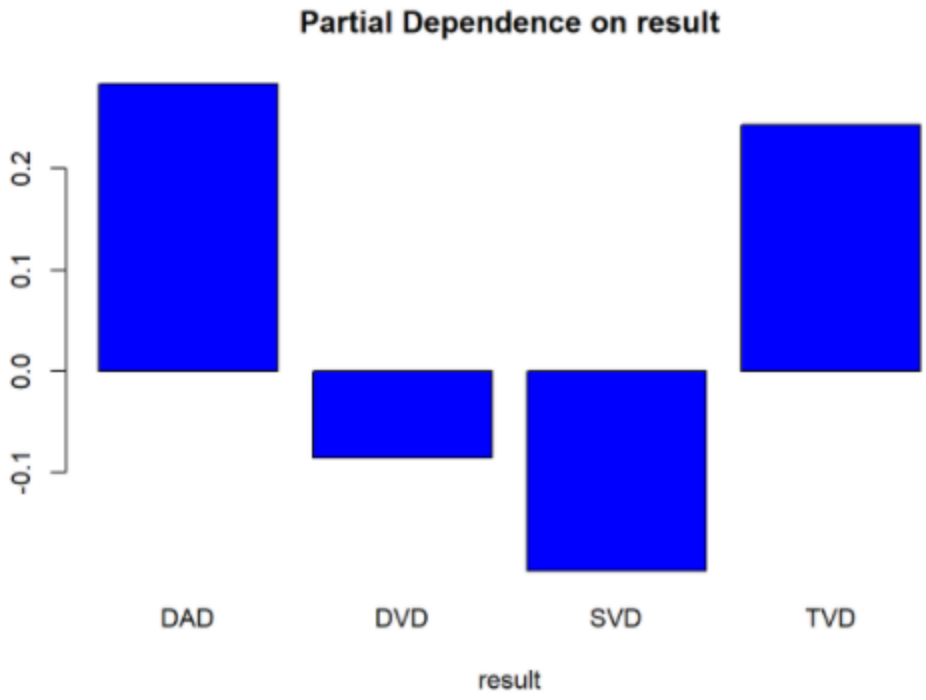
The line graph is going up this indicate that from 95 min onward you will see more people with “severely below normal.”

The trend is not consistance, is hard to draw conclusion upon it. However, I would recommend that keeping the door to balloon time lower than 95 minutes might be best for the patients. Since we see a steady rise in the line graph from this point onward.

All of this is what our model has learned. One weakness is that this mean that all of this depends on the accuracy of our model in predicting this event. Unfornately, our model is not accurate in predicting. To improve our model we might need massive amount of data.

reference: https://christophm.github.io/interpretable-ml-book/pdp.html

Result



Two of the result group, “DAD” and “TVD” seem to be prevalance in a person with “severely below normal”" value of LVEF.

2 Part2

2.1 Objective (Part 2)

- We have established which variable is the most important in predicting LVEF and its trend when the LVEF is “severely below normal” in part 1. This part I would like to look at those people that are likely to be “severely below normal”-high risk group (people who have trend of variable established in part 1) and see if there are someone that are not within this high risk group. I would like to see what did they do differently (How each of the variables differ) from on people that actual have “severely below normal” within this high risk group.

2.2 Method

- Perform various statistical test (Chi-square, t-test) on variables between “Normal” and “Severly below normal”. This is to see if there are actually significant difference between these two group for our variables of interest.

2.3 Summary

I extracted the people that are at more likely to be in the “severly below normal” for LVEF with the criteria below.

- KILLIP group: III and IV AND
- result group: DAD and TVD AND
- door to balloon time: > 100 min

Summary

```
##  door_to_balloon  result     DM       HT      DLD     KILLIP
##  Min.   : 33.00   DAD: 0   no :14   no :11   no :17   I  :12
##  1st Qu.: 58.00   DVD: 3   yes: 7   yes:10   yes: 4   II : 2
##  Median : 75.00   SVD: 1                               III: 5
##  Mean   : 69.95   TVD:17                               IV : 2
##  3rd Qu.: 81.00
##  Max.   :111.00
##       age                          LVEF_group
##  Min.   :28.00   Moderately below normal:1
##  1st Qu.:57.00   normal                 :8
##  Median :58.00   Severely below normal  :7
##  Mean   :60.95   Slightly below normal  :5
##  3rd Qu.:69.00
##  Max.   :82.00
```

2.4 Perform Statistical Test

Age

```
##
##  Welch Two Sample t-test
##
## data:  data_normal_severe$age by data_normal_severe$LVEF_group
## t = -0.79894, df = 7.0931, p-value = 0.4502
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -22.22904  10.97904
## sample estimates:
##                mean in group normal mean in group Severely below norm
al
##                                57.375                                63.0
00
```

DM

```
## Warning in chisq.test(balloon_tbl_DM): Chi-squared approximation may
be
## incorrect
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  balloon_tbl_DM
## X-squared = 0.033482, df = 1, p-value = 0.8548
```

HT

```
## Warning in chisq.test(balloon_tbl_HT): Chi-squared approximation may
be
## incorrect
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  balloon_tbl_HT
## X-squared = 0.058594, df = 1, p-value = 0.8087
```

DLD

```
## Warning in chisq.test(balloon_tbl_DLD): Chi-squared approximation may
be
## incorrect
```

KILLIP

```
## Warning in chisq.test(balloon_tbl_KIL): Chi-squared approximation may
be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  balloon_tbl_KIL
## X-squared = 8.471, df = 3, p-value = 0.03722
```

Result

```
## Warning in chisq.test(balloon_tbl_re): Chi-squared approximation may
be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  balloon_tbl_re
## X-squared = 4.7727, df = 2, p-value = 0.09196
```

As you can see that KILLIP is the only one that is significant different for “normal” and “severly below normal”. Which coincide with our model last time, where KILLIP is the most important in determine the group of LVEF.

In other word, the outcome of LVEF (“normal” or “severely below normal”) depend on the group of KILLIP

I want to determine too if Door to balloon time will determine the group of LVEF in this high group or not. In other words, is LVEF depend on Door to balloon time. I have group door to balloon time into two group one being “>40 min & <= 60” and another being “>60”

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  balloon_time_data
## X-squared = 8.2295e-31, df = 1, p-value = 1
```

As you can see that KILLIP is the only one that is significant different for “normal” and “severly below normal”. Which coincide with our model last time, where KILLIP is the most important in determine the group of LVEF.

In other word, the outcome of LVEF (“normal” or “severely below normal”) depend on the group of KILLIP

I want to determine too if Door to balloon time will determine the group of LVEF in this high group or not. In other words, is LVEF depend on Door to balloon time. I have group door to balloon time into two group one being “>40 min & <= 60” and another being “>60”

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  balloon_time_data
## X-squared = 8.2295e-31, df = 1, p-value = 1
```

These two variable are independent from one another. The Door to balloon time does say anything about LVEF. This indicate that lower balloon time may not actually decrease the risk of someone getting a “severely below normal”

3 Conclusion

Part1

- Door to balloon time is not the most important value in determining the value of LVEF.
- The most important variable in determining LVEF is KILLIP.
- There seem to be a decrease in chance of a person of having “severely below normal”" value of LVEF (<35%) from door to balloon time of around 40 minutes to 70 minutes. But after 70 minutes there seem to be a fluctuation.

Part2

- LVEF(“normal” or “severly below normal”) is dependent on the group of KILLIP. KILLIP group CAN say something about LVEF.
- LVEF(“normal” or “severly below normal”) is independent on Door to balloon time. Door to balloon time CANNOT say anything about LVEF.

Report on Elderly.

Author: Krongpong Monpengpinij

Student ID: 6129003

Send to: Kongkiat Kespechara (CEO of Bangkok Hospital Eastern group at Bangkok Hospital)

Background:

As we are now in the era of advance technology and medical knowledge, people life expectancy increases drastically over the past few decades. This also include Thailand; we are expecting to enter the aging population country in less than ten years. Good quality of life is essential for this population to lead a healthy and fulfilling life. Incidences such as fall can deteriorate elderly population quality of life, since it can lead to conditions such as paralysis. I had chance to be a part of a project that investigate the incidence of fall in elderly population in Thailand. We are trying to predict which characteristics of elderly are at high risk of fall, therefore we can try to intervene. This is an ongoing project.

Duration: 1 month.

Date: 03 / 02 / 2020

elder_report

1 Introduction

The dataset given contain data associated with elderly in Thailand. In total there are 200 variables and 233889 observation. This large amount of data are not suitable to perform analysis on. Therefore, the data are subsetted into smaller data. The subsetting method use is called stratified sampling. The subsetting is based on the incidence of falling of the eldering (variable name = “F64”). The incidence of fall can be divided into 9 groups. The group are as followed:

group 1 = ลื่น

group 2 = สะดุดสิ่งกีดขวาง

group 3 = พื้นต่างระดับ

group 4 = ตกบันได

group 5 = หน้ามืด/วิงเวียน

group 6 = อื่นๆ (ระบุ)

group 9 = ไม่ทราบ

By using stratified sampling method, 70 observations are taken from each group stated above. However, group 9 does not have 70 observations, this group is omitted from the analysis part. In the end, the dataset that will be further analyse contains 420 observation and 200 variables.

2 Objective

- To determine variables that are most influential in elderly live.
- Focusing on the incidence of falls, are the individuals in each group characteristic differ from each other.

3 Statistical method

The most suitable statistical model to achieve our objective is the Multiple correspondence analysis (MCA). This model is great for summarizing and visualizing data containing more than two caetgorical variable, which in fact our dataset variables are all categorical and in large volume. This model is similar to the principal component analysis (common use in microbiota analysis) but the variables are in form of categorical.

4 Results

Figure 1: Contribution of each variables.

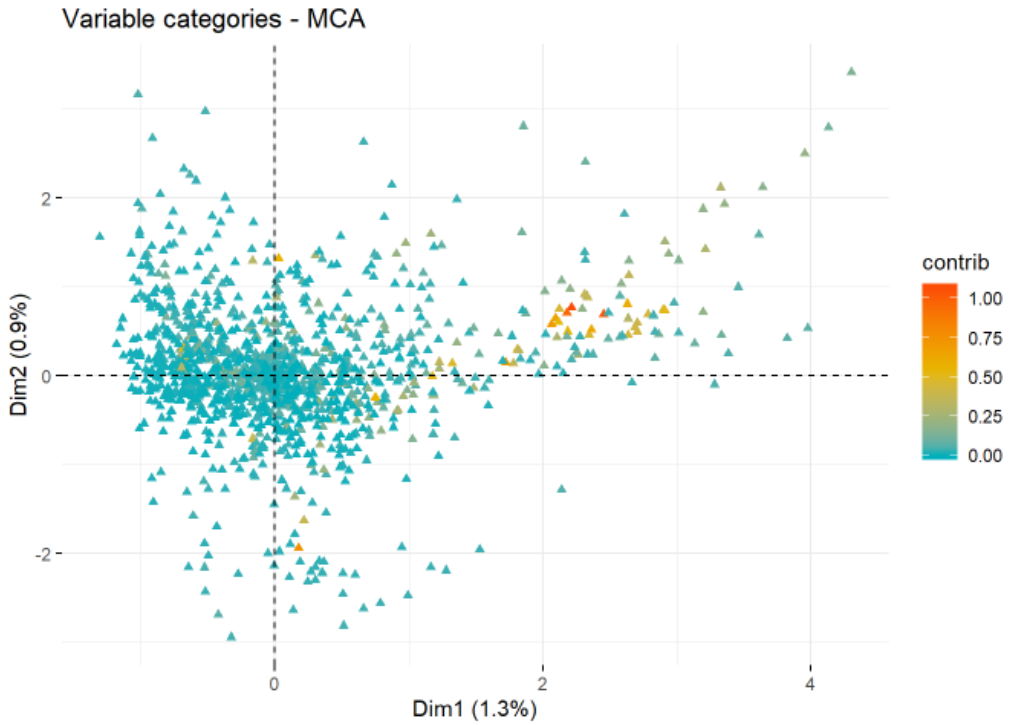


Figure 2: Contribution of top 15 variables to dim 1.

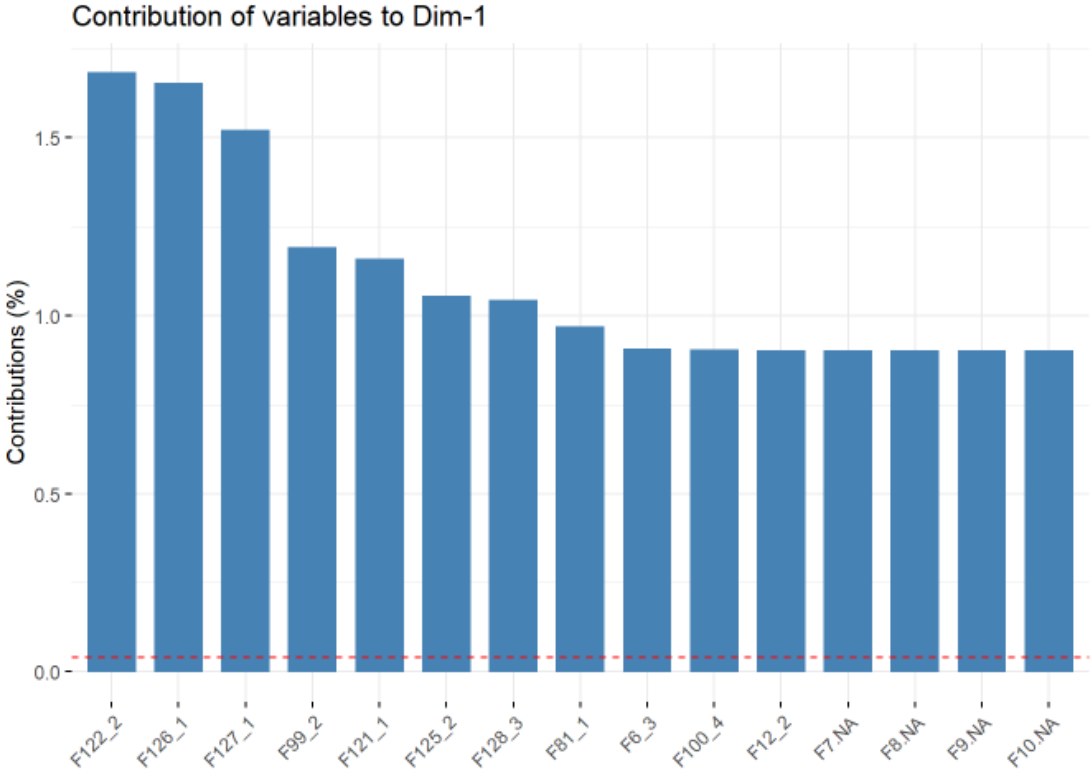
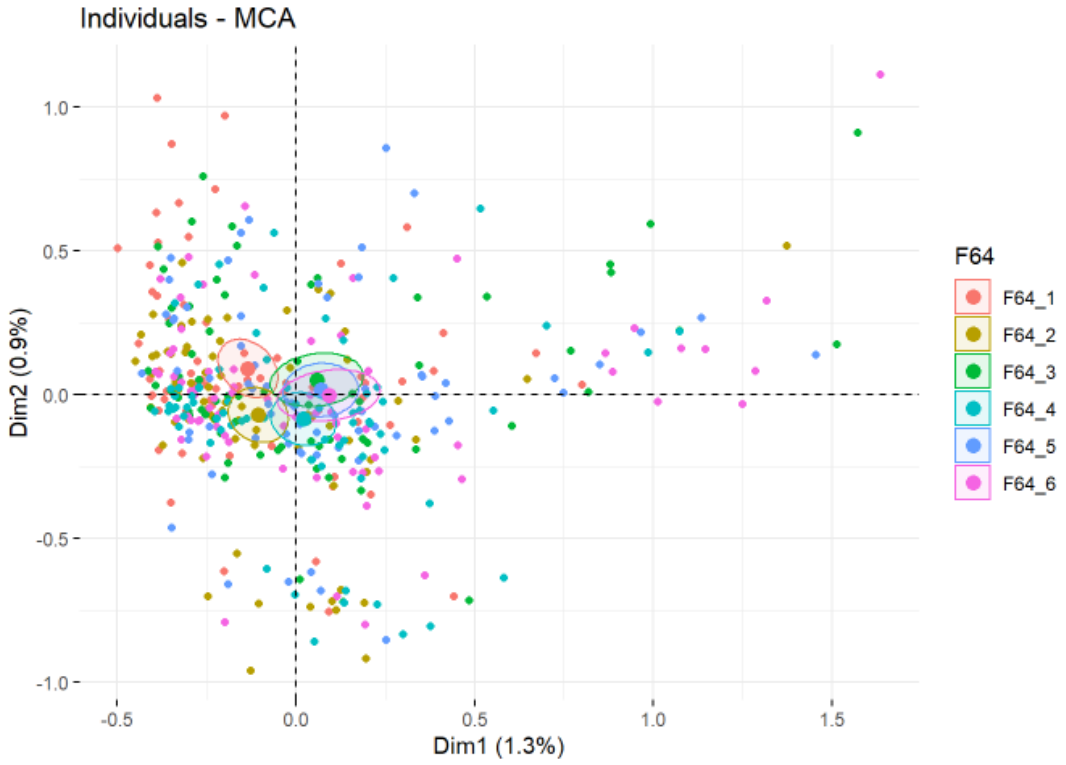


Figure 3: Comparing falling incidences for each group.



5 Discussion

firstly, let address our first objective, to find influential variables in elderly. This step is crucial since there are too many variables to handle, by shrinking it down to a very impactful variables we can focus on how these variables impact the live of the elderly. This objective can be achieve by looking at figure 1 and figure 2. Figure 1 give a global view on every variables and its contribution. Figure 2 is more important, this graph essentially tell us which variables are most influential. From the graph its seem that these variables:

F122 = เพศของผู้ดูแล

F126 = สถานที่อยู่อาศัยของผู้ดูแล

F127 = การเคยได้รับการฝึกอบรม/ดูแลผู้สูงอายุ

F99 = ความต้องการผู้ดูแลปรนนิบัติ

F121 = ผลการสัมภาษณ์ผู้ที่เป็นผู้ดูแลหลัก

F125 = ระดับการศึกษาสูงสุดที่สำเร็จของผู้ดูแล

F128 = อาหารที่ป้องกันอาการท้องผูกในผู้สูงอายุ

F81 = ความสามารถในการทำกิจกรรมต่างๆ ด้วยตนเอง

are most influential in elderly live.

Secondly, we would like to know if there are any different characteristic of individual in different group of fall incidences. let us look at figure 3. The clustering may represent individuals in each fall incidences group. If there are different in characteristic of the different group, the positioning of the cluster should be in different quadrant. its seem that group 1, group 2, group 5 and group 3+4+6 are in different quadrant, which mean that individuals in these groups have different characteristics.

6 Conclusion

- The most influential variables in elderly live are:

F122 = เพศของผู้ดูแล

F126 = สถานที่อยู่อาศัยของผู้ดูแล

F127 = การเคยได้รับการฝึกอบรม/ดูแลผู้สูงอายุ

F99 = ความต้องการผู้ดูแลปรนนิบัติ

F121 = ผลการสัมภาษณ์ผู้ที่เป็นผู้ดูแลหลัก

F125 = ระดับการศึกษาสูงสุดที่สำเร็จของผู้ดูแล

F128 = อาหารที่ป้องกันอาการท้องผูกในผู้สูงอายุ

F81 = ความสามารถในการทำกิจกรรมต่างๆ ด้วยตนเอง

- individuals in fall incidences group 1, group 2, group 5 and group 3+4+6 might have different characteristics.

7 Further plan

The next statistical test will be dependency test on the most influential variables and fall incidences groups.