Analysis of Public Health Data: Diabetes among citizens of the United States, Care and Treatment

Jung Eun Park

Faculty Advisor: Engin A. Sungur

Second Reader: Peter Dolan

University of Minnesota, Morris

Spring 2017

ABSTRACT

In this project, I analyzed data collected by National Health and Nutrition Examination Survey by NHANES and reported on September 2013 to determine what factors are significant in causing the diabetes, and how the various factors are related to lead having diabetes. In this research, I explore the underlying distribution for diabetes for all survey participants between health status I am comparing: Having high blood pressure, taking insulin, Current SBP and DBP, and go further having retinopathy. Because of various factors of data predictors, analyzing which factors would be the most readily available to analyze the effect of diabetes regarding those predictors, furthermore, it is much easier to predict or prevent having serious diabetes.

In my thesis, I will use the two-dimensional graphs and tables in order to find out how two variables are related to diabetes, and go further using multivariable at the same time with using R studio. I will use Binomial and Logistic Regression method with a different set of health care data and focus on various factors influencing causing diabetes. Additionally, odds ratios of regression models are provided on this project, and deviance is used in order to comparing regression models to

observe which variables could have higher probability in terms of causing diabetes disease. I will provide other methods to visualizing my results in consideration of understanding this project no matter who has lack of knowledge on this part.

I. INTRODUCTION

The latest statistics from the Centers for Disease Control and Prevention (CDC) reveal that 29.1 million Americans are now living with diabetes – 21 million have been diagnosed and 8.1 million have diabetes but they haven't yet been diagnosed (American Diabetes Association, 2016). Diabetes is a metabolic disease in which blood sugar is high because the pancreas fails to produce sufficient insulin or the body cells fails to respond to insulin produced by the pancreas (Shoback, David, Gardner & Dolores, 2011). It is possible to lead cause of non-traumatic blindness, heart disease and loss control of many other human body hormones (United State Preventive Service Task Force, 2008). Thus, the U.S. government spends more than $1 billion a year on diabetes research these days, for 75 years, the American Diabetes Association (ADA) has dedicated its resources to research, educate, and fight for progress and against discrimination (American Diabetes Association, 2016). Though it is commoner in developed countries, the greatest increase in prevalence is expected to occur in Asia and Africa because of rapid urbanization, lifestyle changes and western diets adopted. According to the ADA research, the rate of diagnosed diabetes by race/ ethnic background are: 15.9% of American Indians/ Natives, 13.2% of Black, 12.8% of Hispanics, 9% of Asian Americans and 7.6% of White (American Diabetes Association, 2016). It notices that Asian and Black people have higher rate of diagnosed diabetes compare to White. The risk factors for diabetes are various such as age, genetics(race), family history (high blood pressure, high blood sugar) or poor diet.

II. DATA ANALYSIS

In this project, the Diabetes data is used which is provided by National Health and Nutrition Examination Survey (NHANES), first published in 2013, based on 2011-2012. This data documentation is originally composed of XPT file (DIQ_G.XPT), therefore, it is required to convert this file into CSV file in order to use in R studio.

Diabetes by Age

First, 'age' is considered as one of the significant variables might have a close relationship with diagnosing diabetes. From original data, there are 9364 participants who are both males and females age from 1 year to 150 years are involved, however, only 708 participants were willing to answer on this variable. To handle big data, 'age' values are classified into 11 different groups: Less than 1 year, 2 to 9 years, 10 to 19 years, 20 to 29 years, 30 to 39 years, 40 to 49 years, 50 to 59 years, 60 to 69 years, 70 to 79 years, 80 years or older, and Don't know, otherwise, counted as missing values(See Table 1).

From Figure 1, 'ggplot2' package was used in terms of creating bar plot regarding on variable 'age'. Put briefly, this package provides how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details. The frequency of diagnosing diabetes by doctors age between 50 to 59 years is the highest, age 40 – 49 years is the second highest, and the next highest is 60 – 69 years (See Figure 1). The result is founded that the age between 40 to 60 years is the highest risk age when doctors or other health professional first told you have diabetes.

Influence of SBP and DBP in Diabetes

There is a lot of awareness on diabetes because it is a chronic disease with severe complications if untreated. Complication of diabetes include; high Systolic Blood Pressure (SBP), high Diastolic Blood Pressure (DBP), Type 1 diabetes, previously called insulin dependent diabetes, and diabetic retinopathy (reduced vision). Systolic Blood Pressure indicates how much pressure your blood is exerting against your artery walls when the heart beats, and Diastolic Blood Pressure indicates how much pressure your blood is exerting against your artery wall while the heart is resting between beats (American Heart Association, 2016). The normal blood pressure range of SBP is less than 120 mm Hg and less than 80 mm Hg for DBP (American Heart Association, 2016). According to recent studies, the risk of death from ischemic heart disease and stroke doubles with every 20 mm Hg systolic or 10 mm Hg diastolic increase among people from age 40 to 89 (American Heart Association, 2016). Since SBP and DBP are significant factors on causing diabetes and heart diseases, The SBP values of participants were discriminated into 8 different levels based on diagnosed diabetes and described in a simple table (See Table 2). From Table 2, the SBP range between 121 and 140 has the highest frequency of participants, the second highest is range between 100 and 120.

DBP is comparatively similar to SBP, for instance, both are significant factors related to diabetes and heart disease. However, as mentioned before, the normal blood pressure of DBP and SBP are different, and it leads that some people are might be confused between SBP and DBP. The DBP values of participants were discriminated into 8 different levels based on diagnosed diabetes and described in a simple table (See Table 3). The DBP range between 061 and 080 has the highest frequency of participants, the second highest is range between 081 and 100, otherwise, other range groups have relatively lower frequencies (See Table 3).

According to the Table 2 and Table 3, the highest range of SBP from participants is range between 121-140, and the highest range of DBP from participants is range between 61-80. That means

current SBP values of participants are slightly higher than normal range of SBP values suggested by doctors. On the other hands, DBP values of participants are reasonably on the normal range, which is below 80. Overall, participants who diagnosed diabetes have slightly higher SBP rather than DBP. ß

To be more accurate, SBP range between 121 and 140 has 43.78% which has the highest frequency of participants and it is out of normal range, therefore, it concludes that more than half of participants have higher SBP (See Figure 3). In DBP case, the range between DBP 61 and 80 has 64.85% which has the highest frequency of participants. The normal range of DBP is below 80, therefore, it concludes that more than half of participants have normal DBP values (See Figure 4).

Effects of Insulin

Insulin is a hormone that your pancreas makes to allow cells to use glucose. When your body isn't making or using insulin correctly, you can take artificial insulin to help control your blood sugar. In fact, insulin cannot be taken as a pill because it would be broken down during digestion. It must be injected into the fat under your skin for it to get into your blood directly (American Diabetes Association). Insulin shots help the body if it is in need of using or storing the blood glucose. It regulates level of glucose in the bloodstream within a normal range and stores excess glucose in the liver in the form of glycogen. This keeps blood sugar levels within a narrow range. There are plenty of remarkable dangers of not taking insulin. The short-term dangers of not taking prescribed medication are symptoms dry mouth, nausea, urination, and drop in blood pressure. A slow erosion of health is the silent, insidious danger of not taking insulin as necessary for diabetes patients. For long-term dangers of not taking insulin could be causing vision problems, cardiovascular disease, and kidney disease. It is always important to detect any disease in its early stages and prevent to be developing the disease.

It would be an interesting thought that whether how taking insulin affects to diabetes, thus, a variable called 'Insulin' from the data is used this analysis in order to observe the relationship between taking insulin and diabetes. Before exploring of Insulin deeply, it is important to notice that there are two types of diabetes are existed. For the type I diabetes, it is generally diagnosed in children and young adults, caused by an autoimmune response against insulin-producing beta cells. Treatments of type I diabetes must include insulin, as the body no longer produces it. For the type II diabetes, it is usually diagnosed in adults, cased in unknown, however, it has higher relationships with weight, age, inactivity, and genetics. Treatments for type II diabetes usually include some combination of meditations, diet, and enough exercise.

From the Table 4, participants who have diagnosed diabetes are currently taking insulin are about 29.94% of total participants based on having diabetes, and close to 70.06% are not taking insulin even if they are in needs of taking insulin to control glucose and blood sugar level or not required to taking insulin. The table result shows that participants with diabetes are considerably not taking insulin, and it is assumed that type 1 diabetes. No matter who have type 1 diabetes or type 2 diabetes, both bodies need assisting of insulin's action in terms of regulating level of glucose in bloodstream. If diabetes is left untreated, high blood glucose can lead to complications such as blindness, and other organs damage.

Retinopathy

At this point, diabetic retinopathy will be considered as a significant factors similar to insulin. Diabetic retinopathy affects blood vessels in the light-sensitive tissue called the retina that lines the back of the eye (National Eye Institute, 2015). These blood vessels are able to swell; sometimes abnormal new blood vessels grow on the retina. It is the most common cause of vision loss among

people with diabetes and the leading cause of blindness among middle-age adults. If it is left untreated, it will increase sudden changes in vision or blurred vision, having eye floaters and spots, and eye pain. Diabetic retinopathy is considered as a new variable excerpt from the data whether it is influenced by taking insulin since blindness is a significant factor by diabetes. A two-dimensional table is generated with Retinopathy for vertical axis and taking insulin for horizontal axis (See Table 5).

Furthermore, Chi-squared test is done by with retinopathy and insulin (See Table 6). The null hypothesis is set as whether diabetic retinopathy and insulin are independent, on the other hand, the alternative hypothesis is set as they are dependent. From the Table 6, the Chi-square value is 44.127, degree of freedom is 1, and the p-value is extremely small. As the p-value is small enough, which is smaller than the 0.05 significance level, we reject the null hypothesis that taking insulin is independent of causing Retinopathy. In other words, taking insulin have a higher proportion with having a chance of retinopathy.

Logistic Regression

The logistic regression approach is used in this project in terms of analyze the Diabetes data with several factors which are close related to having diabetes. Logistic regression is the statistical method for analyzing a dataset to conduct when the dependent variable is binary. Logistic regression is used to describe data and explain the relationship between one dependent binary variable and one or more variables. The goal of logistic regression is to find the best fitting model to describe the relationship between the binary characteristics of response variable and a set of independent predictor variables. Logistic regression generates the coefficients of a formula to predict a log transformation of the probability of presence of the response variable (See Equation (1)).

In simple logistic regression, there is one nominal variable with two values such as yes/no and one measurement variable. One goal is to see whether the probability of getting a particular value of

the nominal variable is associated with the measurement variable, the other goal is to predict the probability of getting a specific value of the nominal variable with the given measurement variable. R or R studio make very easy to fit a logistic regression model. The function to be called is 'glm'. There is a simple regression model is created with one predictor variable (See Table 7). Fitting this model to the data yields estimates of $\hat{\beta}_1$ = 1.2825 according to Equation 2. So increasing the chance of having diabetic retinopathy level by 10% increases the probability of taking insulin by 12.83 on the logit scale, that means taking insulin has positive effect on having retinopathy. The confidence interval for insulin coefficient does not contain zero, there is evidence to suggest that taking insulin is related to chance of having diabetic retinopathy (See table 8).

With multiple predictors variables, the model output is somewhat different from that of an ordinary simple model. For another model with two predictor variables, the result of this model yields estimates of $\hat{\beta}_1$ = 1.2078 and $\hat{\beta}_2$ = -0.2844. So increasing the chance of having retinopathy level by 10% increases the probability of taking insulin by 12.08 on the logit scale if the taking diabetic pill to reduce the blood sugar level fixed, and taking diabetic pill decreases the probability of having retinopathy by 0.28 if taking insulin is fixed (See table 9). The confidence interval for taking diabetic pill coefficient contains zero, there is no evidence to suggest that taking diabetic pill to lower blood sugar is related to chance of having retinopathy (See table 10).

Odds Ratio

At the center of the logistic regression analysis is the task the log odds of an event. An odds ratio is a measure of association between an exposure and outcome. Odds are the probability of an event occurring divided by the probability of the event not occurring. An odds ratio is used to compare the odds for two groups; calculated by dividing the odds in group 1(success) by the odds in group

2(failure). If the odds are greater than one, then the event is more likely to happen than not. If the odds are less than one, then the event is less likely to happen than not. The equation of odds would be odds of success of variable y(response) given that variable x(predictor) condition is that exponent raised to the power of estimated coefficient of variable y plus taking variable x estimated coefficient (See equation (3)). The odds ratio of success of retinopathy with taking insulin is e^{β_1} since the nominator's e^{β_0} value is divided by denominator's e^{β_0} (See equation (4)). Additionally, the odds increases as variable x increases, if β_1 is greater than zero, on the other hand, the odds decreases as variable x decreases, if $\beta_1$is smaller than zero. There is no relationship between the variable x and variable y when β_1 equals to zero.

For instance, using the first regression model, the estimated odds ratio is $\widehat{OR}$ = 3.606 (See Table 11). That means taking insulin has 3.606 times the odds of taking no insulin being diagnosed retinopathy. The confidence interval for the odds ratio does not include 1.0, so again there is statistical evidence of difference in risk of diagnosing retinopathy by insulin parameter (See Table 12). For the second regression model, it is required to explicitly adjust for the other predictor, the odds ratio for variable x_1 when variable x_2 is fixed. In this case, variable $x_1$is defined as taking insulin and variable x_2 is defined as taking diabetic pill. The estimated odds ratio is $\widehat{OR}$ = 3.346 (See Table 13). This means that taking insulin has 3.346 times the odds of taking no insulin being diagnosed retinopathy with taking diabetic pill constant. The confidence interval for the odds ratio does not include 1.0, so again there is statistical evidence of a difference in risk of diagnosing retinopathy by insulin parameter with taking diabetic pill constant (See Table 14).

The results of both two models notice that the estimated odds ratio and confidence interval for the insulin effect is slightly modified by the inclusion of diabetic pill to lower blood sugar adjustment

in the model. The estimated odds ratio for the second model has slightly smaller than the first simple model, that means insulin has higher chance of having retinopathy compare to adding diabetic pill constant.

Deviance

The deviance test is a measure of goodness of fit between model and data. This approach is based on estimating two models, M_1 and M_2. It is assumed that M_1 excludes the effects hypothesized to be null, while these effects are included in M_2. For each model, a deviance statistic, equals to -2 in L (likelihood ration test) for that model, is computed (See Theorem 1). In general, good models will have smaller and positive values of deviance, on the contrast, the larger the deviance, the poorer the fit to the data.

By doing likelihood ratio testing, and comparing, $\Delta G^2 = G^2$ for $M_1 - G^2$ for M_2 where $\hat{y} = m_i\hat{\theta}(x_i)$, are the estimated number of events in m_i trials. (See Equation (5)). In this project, M_1 is defined as $\theta(x) = m(\beta_0 + \beta_1 insulin)$, and M_2 is defined as $\theta(x) = m(\beta_0 + \beta_1 insulin + \beta_2 takePill)$. The anova test helps to observe the analysis of deviance. The test statistic value from anova test with the first and second regression models is $G^2_{M_1} - G^2_{M_2}$ = 658.19 – 656.42 = 1.77, and the p-value is $P(X_1^2 > 1.77) = 0.1828$ (See Table 13). The p-value is quite large and means that the data observed are consistent with the M_1, and that means taking diabetic pill to reduce blood sugar level is not associated with having retinopathy after adjusting insulin.

Care and Treatment

Many Americans have prediabetes or diabetes. Prediabetes is now recognized as a reversible condition that increases an individual's risk for development of diabetes. Lifestyle risk factors for prediabetes include overweight and physical inactivity. Lifestyle intervention program effectively prevents

prediabetes from developing into type II diabetes (Tuso, 2014). This program suggests prediabetes patients to do physical activity and dietary changes. Additionally, it suggests them to reduce intake saturated fat less than 10% of energy intake and increase intake fiber greater than or equal to 15g per 1000 kcal. Lifestyle intervention may decrease the percentage of prediabetes patients in whom diabetes develops to 20% (Kirkman, 2012). There is another good method to treat prediabetes, that is pharmacotherapy. Many evidences of potential benefits from pharmacotherapy to prevent diabetes in patients with prediabetes were reported by many researchers (Tuso, 2014). Metformin, which has a good safety profile, has beneficial effects on BMI, and lipid concentrations and 45% risk reduction for development of type II diabetes was reported (Tuso, 2014).

Treatment of Diabetes patients is similar to prediabetes, however, it desires that diabetic patients are more in needs of cares rather than prediabetes patients. Prevention or delay of aggravating diabetes is significant fact for diabetes patients. Lifestyle interventions are also beneficial to care of diabetes (George, 2013). In addition to lifestyle intervention, many patients require glucose lowering medication in order to achieve normalization of blood glucose and AIC levels (George, 2013). Lipid lowering and blood pressure control should be considered as well. Blood pressure control helps to lowering SBP levels to moderate targets, reducing cardiovascular risk in especially older adults (Kirkman, 2012). The most important treatment for moth prediabetes and diabetes patient is screening for chronic diabetes complications. It is good on cost-effective part, increases long life expectancy, and preventing latter risk of severe complications (Kirkman, 2012).

III. FIGURES AND TABLES

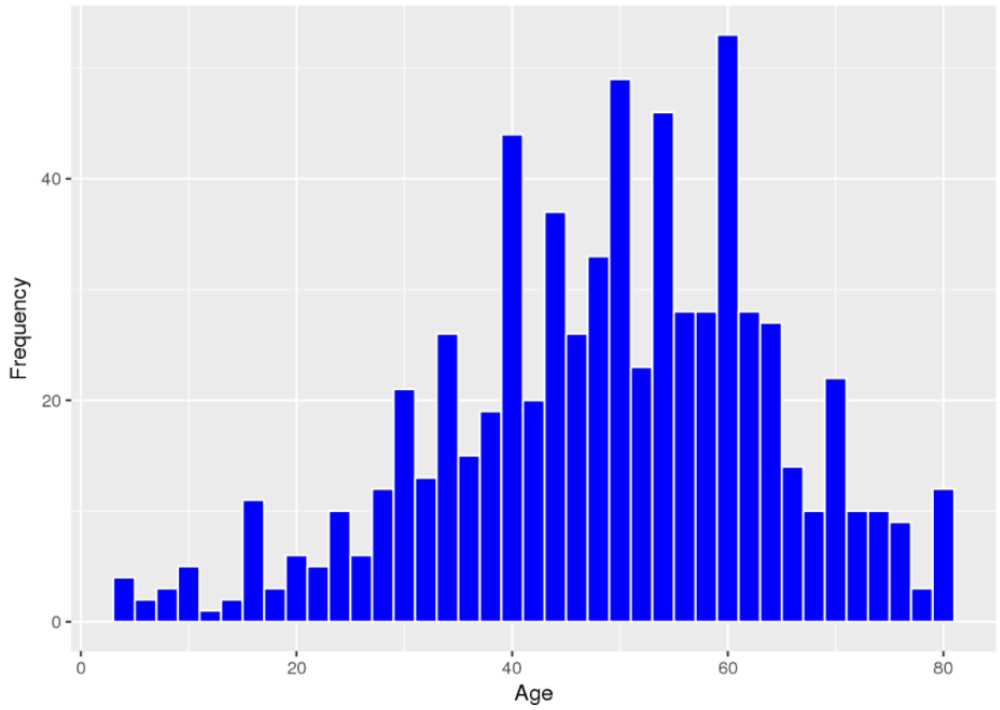

Figure 1. *Histogram of age by diagnosed diabetes range in 1 to 150 years (Over 80 years count as 80).*

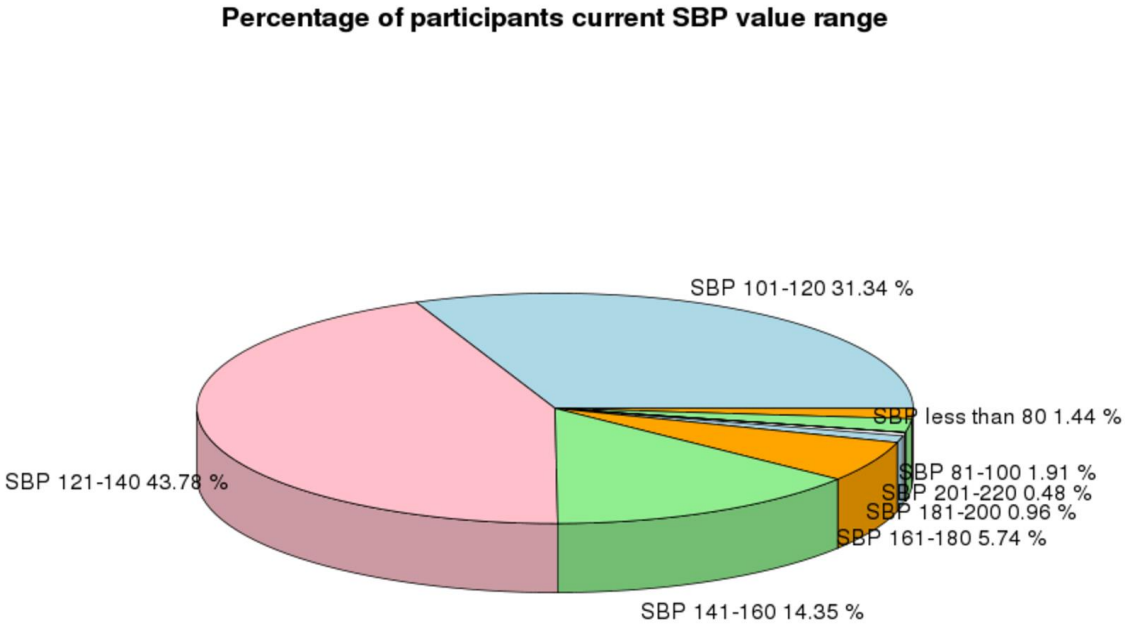

Figure 2. *Pie plot of percentage of participants current SBP value by diagnosed diabetes range in 63 to 220 mm Hg.*

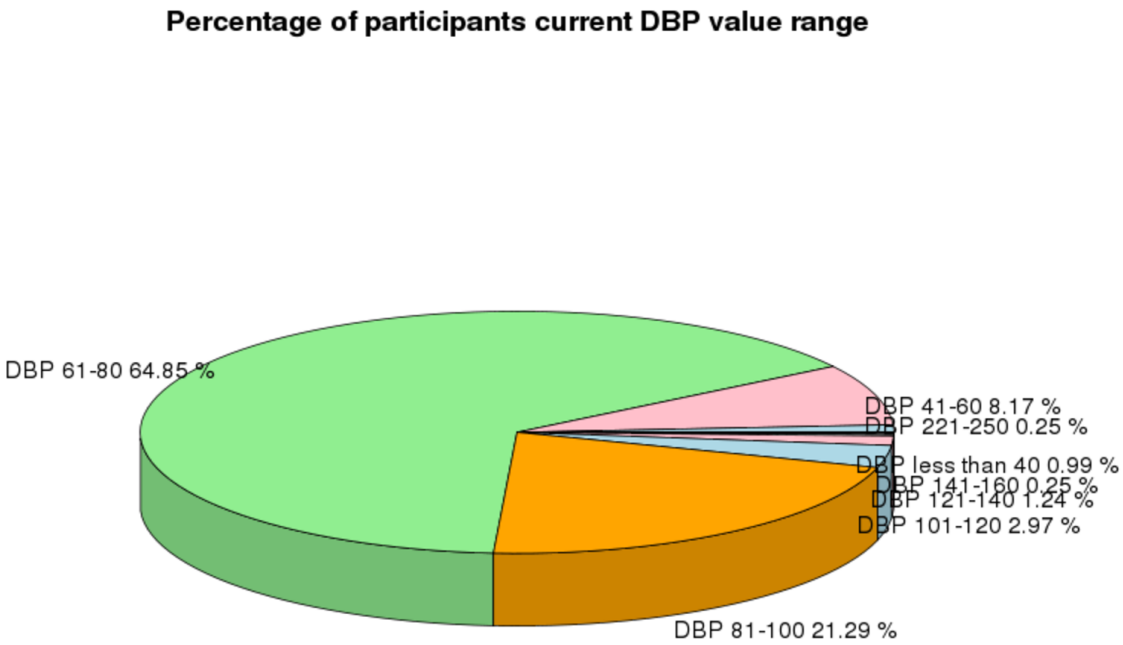

Figure 3. *Pie plot of percentage of participants current DBP value by diagnosed diabetes range in 25 to 250 mm Hg.*

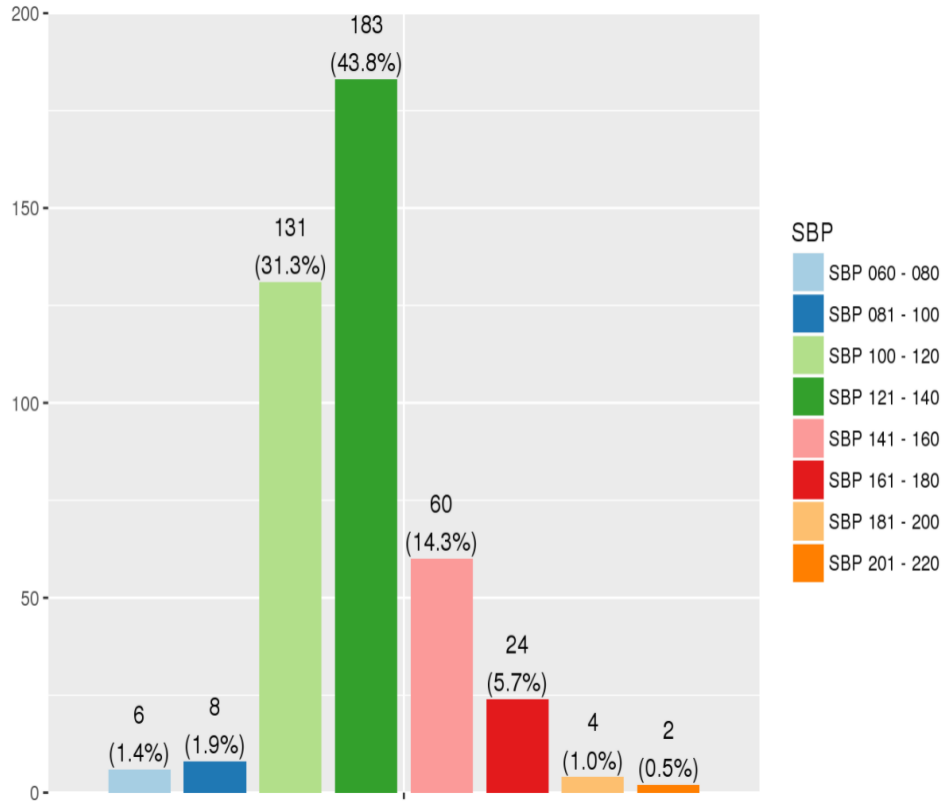

Figure 4. *bar plot of participants SBP values with diabetes divided by 8 levels.*

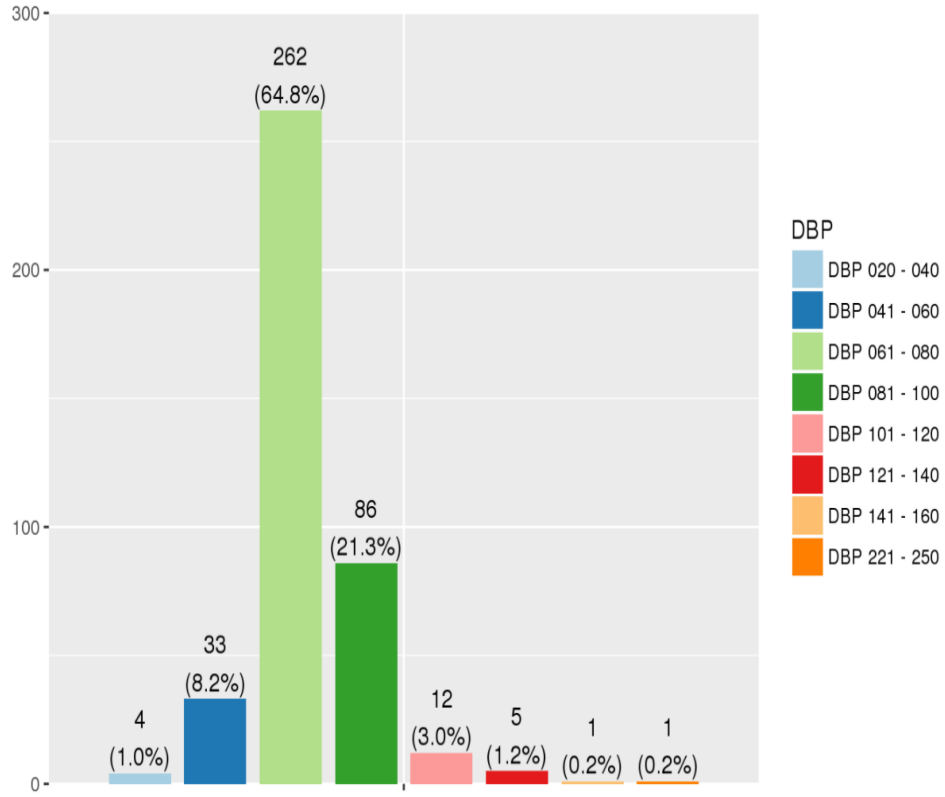

Figure 5. *bar plot of participants DBP values with diabetes divided by 8 levels.*

```
## age
##  Age 01 or less     Age 02 - 09     Age 10 - 19     Age 20 - 29
##               5               9              22              39
##     Age 30 - 39     Age 40 - 49     Age 50 - 59     Age 60 - 69
##              94             160             174             132
##     Age 70 - 79 Age 80 or older      Don't know
##              54              12               7
```

Table 1. *Classify values of age into 11 different groups based on diagnosed diabetes: Age 1 or less, Age 02-09, Age 10-19, Age 20-29, Age 30-39, Age 40-49,. Age 50-59, Age 60-69, Age 70-79, Age 90 or older, and Don't know.*

```
## SBP
## SBP 060 - 080 SBP 081 - 100 SBP 100 - 120 SBP 121 - 140 SBP 141 - 160
##             6             8           131           183            60
## SBP 161 - 180 SBP 181 - 200 SBP 201 - 220
##            24             4             2
```

Table 2. *Classify values of SBP into 8 different groups based on diagnosed diabetes: SBP range 60-80, SBP range 81-100, SBP range 101-120, SBP range 121-140, SBP range 141-160 SBP range 161-180, SBP range 181-200, SBP range 201-220.*

```
## DBP
## DBP 020 - 040 DBP 041 - 060 DBP 061 - 080 DBP 081 - 100 DBP 101 - 120
##             4            33           262            86            12
## DBP 121 - 140 DBP 141 - 160 DBP 221 - 250
##             5             1             1
```

Table 3. *Classify values of age into 8 different levels based on diagnosed diabetes: DBP range 20-40, DBP range 41-60, DBP range 61-80, DBP range 81-100, DBP range 101-120, DBP range 121-140, DBP range 141-160, DBP range 161-180, DBP range 181-200, DBP range 201-220, DBP range 221-250.*

Diabetes	*insulin*		***Total***
	Yes	No	
Yes	212	496	708
No	1	8523	8524
Borderline	0	125	125
Total	213	9144	9357

χ^2=2635.662 · df=2 · Cramer's V=0.531 · Fisher's p=0.000

Table 4. *A simple two-dimensional table of taking insulin (Yes or No) with diabetes status (Yes, No, or Borde rline). Simple Chi-squared Test is provided.*

insulin	*Retinopathy*		***Total***
	Yes	No	
Yes	75	137	212
No	65	426	491
Total	140	563	703

$\chi^2=44.127 \cdot df=1 \cdot \phi=0.254 \cdot p=0.000$

Table 5. *A simple two-way table of taking insulin (Yes or No) with having Retinopathy (Yes or No).*

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  t2
## X-squared = 44.127, df = 1, p-value = 3.077e-11
```

Table 6. *Chi-Squared Test of Independence between Taking insulin and Retinopathy*

```
rmpm1 <- glm(retinopathy ~ insulin, family = binomial, data = newdata)
summary(rmpm1)
```

```
##
## Call:
## glm(formula = retinopathy ~ insulin, family = binomial, data = newdata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9372  -0.5335  -0.5335  -0.5335   2.0100
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.8777     0.1332 -14.099  < 2e-16 ***
## insulinYes    1.2825     0.1960   6.543 6.03e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 701.01  on 700  degrees of freedom
## Residual deviance: 658.19  on 699  degrees of freedom
##   (374 observations deleted due to missingness)
## AIC: 662.19
##
## Number of Fisher Scoring iterations: 4
```

Table 7. *The first regression model; retinopathy as response, insulin as one predictor are used.*

```
summary(rmpm1)$coefficients
```

```
##              Estimate Std. Error    z value     Pr(>|z|)
## (Intercept) -1.877702  0.1331823 -14.098738 3.866207e-45
## insulinYes   1.282535  0.1960199   6.542883 6.034386e-11
```

```
confint(rmpm1)
```

```
## Waiting for profiling to be done...
```

```
##                 2.5 %    97.5 %
## (Intercept) -2.1475371 -1.624597
## insulinYes   0.8996133  1.669041
```

Table 8. *Summary of first regression model with confidence interval*

```
rmpm2 <- glm(retinopathy ~ insulin + takePill, family = binomial, data = newdata)
summary(rmpm2)
```

```
##
## Call:
## glm(formula = retinopathy ~ insulin + takePill, family = binomial,
##     data = newdata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9938  -0.5910  -0.5181  -0.5181   2.0371
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.6562     0.2092  -7.917 2.44e-15 ***
## insulinYes    1.2078     0.2037   5.931 3.02e-09 ***
## takePillYes  -0.2844     0.2119  -1.342     0.18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 701.01  on 700  degrees of freedom
## Residual deviance: 656.42  on 698  degrees of freedom
##   (374 observations deleted due to missingness)
## AIC: 662.42
##
## Number of Fisher Scoring iterations: 4
```

Table 9. *The second regression model; retinopathy as response, insulin and diabetic pill as predictor variables*

```
summary(rmpm2)$coefficients
```

```
##               Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) -1.6562451  0.2092043 -7.916880 2.435442e-15
## insulinYes   1.2078118  0.2036517  5.930773 3.015123e-09
## takePillYes -0.2844396  0.2119142 -1.342239 1.795184e-01
```

```
confint(rmpm2)
```

```
## Waiting for profiling to be done...
```

```
##                  2.5 %     97.5 %
## (Intercept) -2.0782617 -1.2570502
## insulinYes   0.8096918  1.6091215
## takePillYes -0.6961901  0.1356919
```

Table 10. *Summary of second regression model with confidence interval*

```
exp(coef(rmpm1))[2]
```

```
## insulinYes
##   3.605769
```

```
exp(confint(rmpm1))[2, ]
```

```
## Waiting for profiling to be done...
```

```
##    2.5 %   97.5 %
## 2.458652 5.307075
```

Table 11. *The Odds Ratio value of the first regression model with confidence interval of odds*

```
exp(coef(rmpm2))[2]
```

```
## insulinYes
##   3.346155
```

```
exp(confint(rmpm2))[2, ]
```

```
## Waiting for profiling to be done...
```

```
##    2.5 %   97.5 %
## 2.247215 4.998418
```

Table 12. *The Odds Ratio value of the second regression model with confidence interval of odds*

```r
library(stats)
library(car)
anova(rmpm1, rmpm2, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: retinopathy ~ insulin
## Model 2: retinopathy ~ insulin + takePill
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       699     658.19
## 2       698     656.42  1   1.7749   0.1828
```

Table 13. *Analysis of deviance with two regression models; anova test is used.*

IV. CONCLUSION

SBP and DBP are regarding as significant factors which are highly related to causing diabetes. From the data analysis in this project, about 35% of participants with diabetes are in normal range of SBP, however, approximate 65% are are outside the normal range. For DBP, about 73% of participants with diabetes are in normal range, otherwise are in out of the normal. This fact shows that participants who have diagnosed diabetes have higher SBP values and lower DBP values. It seems that they are in need of lowering SBP value to moderate range. Moreover, it is founded that there is a relationship between retinopathy and taking insulin – those taking insulin have a higher proportion with retinopathy. There are many researches are reported that insulin helps curing diabetic diseases, however, I'd like to point out that people who must require to take insulin are already in serious level and many other factors could possibly have affected on diabetes so that on the face of it, taking insulin is related to having a chance of retinopathy. Lastly, taking diabetic pill to reduce blood sugar level is not associated with having retinopathy after adjusting for taking insulin. Diabetic pill exerts influence on reducing blood sugar level, therefore, it might be able to reduce causing retinopathy.

There are considerable treatment and care are existed and suggested by doctors or health professionals. Even if these cares are beneficial, it is always important to detect any disease in its early stages and prevent to be developing the disease. Periodical examination of health by doctors or health professionals increases to avoid having chronic diseases.

APPENDIX A. FORMATTING DETAILS

A.1. Formatting equations

$$\text{logit(p)} = ln\left(\frac{p}{1-p}\right) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k \quad (1)$$

$$P(Y = 1\,|X) = \frac{\exp(B_0+B_1X)}{1+\exp(B_0+B_1X)} = \frac{e^{B_0+B_1X}}{1+e^{B_0+B_1X}} \quad (2)$$

$$odds(Y = 1|X) = \exp(\beta_0 + \beta_1X) = e^{\beta_0+\beta_1X} \quad (3)$$

$$OR = \frac{odds(Y|x=1)}{odds(Y|x=0)} = \frac{\exp(\beta_0+\beta_1)}{\exp(\beta_0)} = \exp(\beta_1) = e^{\beta_1} \quad (4)$$

$$G^2 = 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (m_i - y_i)\log\left(\frac{m_i-y_i}{m_i-\hat{y}_i}\right)\right] \quad (5)$$

A.2. Formatting definitions and theorems

Definition 1. ***The Likelihood Ratio Test*** *is a hypothesis test that compares the goodness-of-fit of two models, an unconstrained model with all parameters free, and its corresponding model constrained by the null hypothesis to fewer parameters, to determine which offers a better fit for your sample data.*

Theorem 1. ***The Likelihood Ratio Test***

Suppose two alternative models are under consideration, one model is simpler or shorter than the other.

Another common situation is to consider nested models, where one is obtained from the other one by putting some of the parameters to be zero. Suppose now we test

$H_0: odds = \exp(\beta_0 + \beta_1x_1 + \ldots + \beta_mx_m)$ vs. $H_a: odds = \exp(\beta_0 + \beta_1x_1 + \ldots \beta_mx_m + \ldots + \beta_kx_k)$

The likelihood ratio statistic is

$$-2\log(\Lambda) = -2[I(H_0) - I(H_a) = -2I(H_0) - -2I(H_a)]$$

This test examines the benefit of x_{m+1} up to x_k in explaining the response after adjusting for the first m predictor variables. For the degree of freedom, the number of observations (n) – the number of beta parameters.

ACKNOWLEDGMENTS

This work was supported by **Engin A. Sungur**, Statistics professor from University of Minnesota-Morris. I sincerely thank to my advisor **Dr. Engin Sungur** for his guidance and immense encouragement in carrying out this project work.

I am extremely grateful to **Dr. Jon Anderson** for giving suggestions and correctness on this project. To the all faculties at the University of Minnesota-Morris Division of Science & Math/Statistics, greatly thank you for always listening and giving me words of encouragement.

I thank the University of Minnesota-Morris Division of Science & Math/Statistics for providing a chance to work on this project and explore how to integrate Statistics into our daily lives.

REFERENCES

American Diabetes Association. (2016). Insulin Basics. Retrieved from http://www.diabetes.org/living-with-diabetes/treatment-and-care/medication/insulin/insulin-basics.html

American Diabetes Association. (2016). Statistics About Diabetes. Retrieved from http://www.diabetes.org/diabetes-basics/statistics/

American Heart Association. (2016). Understanding Blood Pressure Readings. Retrieved from http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/GettheFactsAboutHighBloodPressure/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp#.WNcz_xLyub8

Bansal, N. (2015, March 15). Prediabetes Diagnosis and Treatment: A Review. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4360422/

Centers for Disease Control and Prevention (CDC). (2013, September). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Retrieved from https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/DIQ_G.htm

George, M. M., & Copeland, K. C. (2013). Current Treatment Options for Type 2 Diabetes Mellitus in Youth: Today's Realities and Lessons from the TODAY Study. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3545061/

Kezerle, L, Shalev, L., & Barski, L. (2014, August 28). Treating the Elderly Diabetic Patient: Special Considerations. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4154881/

Kirkman, M. S, Briscoe, V. J., Clark, N., Florez, H., Haas, L. B., Halter, J. B., Huang, E. S., Korytkowski, M. T., Munshi, M. N., Odegard, P. S., Pratley, R. E., and Swift, C. S. (2012, December). Diabetes in Older Adults: A Consensus Report. *Diabetes Care*. American Diabetes Association, Retrieved from http://care.diabetesjournals.org/content/35/12/2650

Marshall, J. (2015, March 30). Dangers of Not Taking Insulin or Other Diabetes Medication. Retrieved from http://www.informationaboutdiabetes.com/lifestyle/lifestyle/dangers-of-not-taking-insulin-or-other-diabetes-medication

Mayo Clinic (2016, April 21). Diabetes Symptoms: When Diabetes Symptoms Are a Concern. Retrieved from http://www.mayoclinic.org/diseases-conditions/diabetes/in-depth/diabetes-symptoms/ART-20044248

Schervish, L. B., & Warshaw, H. S. (n.d.). 16 Hot Topic in Diabetes Research Today. Retrieved from http://www.diabeticlivingonline.com/monitoring/doctor-visits/16-hot-topics-diabetes-research-today?page=0

Shoback, David, G., & Gardner, D. (2011). *Greenpan's basic & clinical endocrinology (9th ed.)*. New York: McGraw-Hill ISBN 0-07-162243-8.

The National Eye Institute. (2015, September). Facts About Diabetic Eye Disease. Retrieved from https://nei.nih.gov/health/diabetic/retinopathy

Tuso, P. (2014). Prediabetes and Lifestyle Modification: Time to Prevent a Preventionable Disease. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4116271/

United States Preventive Services Task Force (USPSTF). (2008). Screening for type 2 diabetes in Adults. Retrieved from 1855048697206727
http:// www.uspreventiveservicestaskforce.org/uspstf08/type2/type2rs.htm