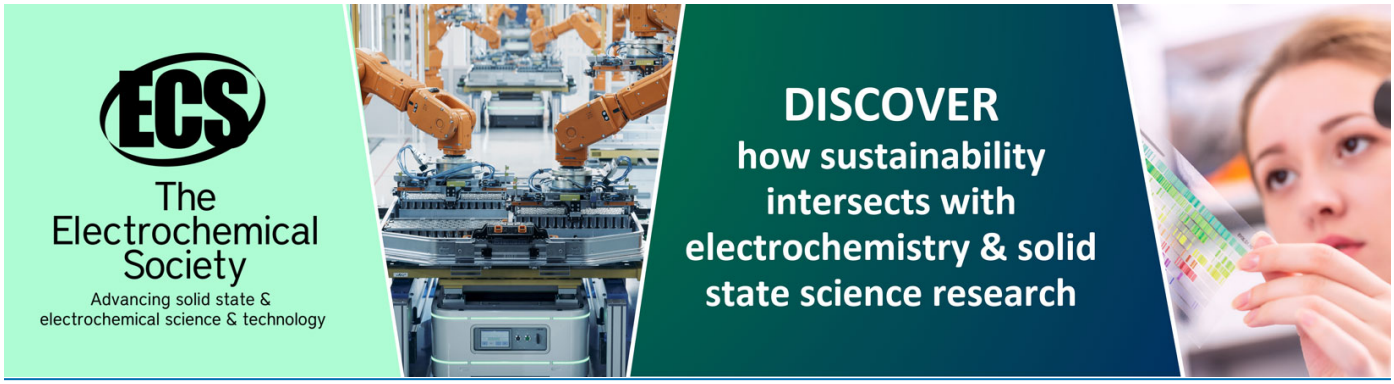
Probability of caused factors stroke disease use link and reliability functions



Probability of caused factors stroke disease use link and reliability functions

Sudarno, T Widiharih, M A Mukid

Department of Statistics, Faculty Sciences and Mathematics, Diponegoro University, Semarang, Indonesia

Corresponding author: dsghani@gmail.com

Abstract. Diabetes mellitus disease is disease which abnormal metabolism for a long time, because pancreas can not be able to produce insulin hormone be enough, or because body can not be able to use insulin hormone has been produced by effective. A stroke occurs if the flow of oxygen-rich blood to a portion of the brain is blocked. Without oxygen, brain cells start to die after a few minutes. Sudden bleeding in the brain also can cause a stroke if it damages brain cells. These objectives are finding significant factors which cause diabetes mellitus disease and determine ordinal regression model. Ordinal regression model is used to look for probability and reliability functions of a patient has stroke disease. The method used to three link functions, that are logit link function, normit link function, and cloglog link function. Testing of homogeneity prediction result of link functions uses linear hypothesis test. Factors caused diabetes mellitus are body mass index, high density lipoprotein, and albuminuria. These factors cause to diabetes mellitus and stroke could be used to prevent diseases, in order to all persons are healthy and happy. The result that probability of a patient with macroalbuminuria has stroke greater than microalbuminuria and a patient with microalbuminuria has stroke greater than normal. Probability of patient with macroalbuminuria by logit, normit, and clogloc link functions is decrease, respectively. Probability of patient with microalbuminuria by logit, normit, and cloglog link functions is increase, respectively. Reliability of a patient with macroalbuminuria, normal, and microalbuminuria have stroke, respectively, is decrease. Reliability of patient with macroalbuminuria by logit, normit, and clogloc link functions, respectively, is increase. Reliability of patient with microalbuminuria by logit, normit, and clogloc link functions, respectively, is zero. All of link function methods yield estimation probability value is the same. AIC value of logit link function, normit link function, and cloglog link function are, respectively, 167.6826, 168.3965, and 169.6107. These results are same by the result of linear hypothesis analysis that AIC values are not different meanwhile their AIC values are not equal. Therefore, logit model, normit model and cloglog model could be used to predict probability with result almost same.

1. Introduction

Diabetes mellitus disease is disease which abnormal metabolism for a long time, because insulin pancreas can not be able to produce insulin hormone be enough, or the body can not be able to use insulin hormone has been produced by effective. Insulin hormone is hormone which function for regulate equilibrium level of blood sugar in body. If pancreas function is bad and insulin hormone is a few, then the body will have high level of blood sugar (hyperglycemia). This condition can cause abnormal to many organs of patient body. The body is a set of many organs which have functions respectively. If the level of blood sugar is not controlled by well, it can cause many various

complications in body organs. The diseases often appear at patient with high level of blood sugar are coronary heart, stroke, high blood pressure, and kidney. Generally, a person could get this disease, because generated from parents, human life to eat and drink, and bad hobby (WHO, 2010).

There are some symptoms or way to detection whatever a person has diabetes mellitus. Diabetes disease symptoms often appear are often thirsty, urinate (polyuria), weak, infection. Way to detect a person has this disease could use level of sugar in urine or blood after fast. If level of blood sugar after fast less than or equal to 126 mg/dL, then a person can not have diabetes mellitus disease. But, if it has level of blood sugar after fast greater than 126 mg/dL, then this person has symptom of diabetes mellitus disease. There are two types of diabetes mellitus, that are type 1 and type 2. Type 2 is diabetes disease which a body still produce insulin hormone, but little amount. Factors cause this diseases are genetics or overweight factors. This disease can be controlled by manage level of insuline in blood. The way by injecting insulin hormone at scheduled time (Rotella et al., 2015).

A stroke occurs if the flow of oxygen-rich blood to a portion of the brain is blocked. Without oxygen, brain cells start to die after a few minutes. Sudden bleeding in the brain also can cause a stroke if it damages brain cells. If brain cells die or are damaged because of a stroke, symptoms occur in the parts of the body that these brain cells control. Examples of stroke symptoms include sudden weakness; paralysis or numbness of the face, arms, or legs (paralysis is an inability to move); trouble speaking or understanding speech; and trouble seeing. A stroke is a serious medical condition that requires emergency care. A stroke can cause lasting brain damage, long-term disability, or even death. If you think you or someone else is having a stroke, call 9–1–1 right away. Do not drive to the hospital or let someone else drive you. Call an ambulance so that medical personnel can begin life-saving treatment on the way to the emergency room. During a stroke, every minute counts. A stroke is a medical condition in which poor blood flow to the brain results in cell death. There are two main type of stroke, those are ischemic, due to lack of blood flow, and hemorrhagic, due to bleeding. Both result in parts of the brain not functioning properly. Sign and symptom of a stroke may include an inability to move or feel on side of the body, problems understanding or speaking, dizziness, or loss of vision to one side. Signs and symptoms often appear soon after the stroke has occurred. If symptoms last less than one or two hours it is known as a transient ischemic attack (TIA) or ministroke. A hemorrhagic stroke may also be associated with a severe headache. The symptoms of a stroke can be permanent.[11]

Regression is formula that be constructed between independent variable and independent variable. Formally, regression can be defined as relation between response variable versus predictor variables. Response variable can have categorical and interval or ratio type. Then predictor variable can be measured by nominal or ordinal scale, and the other have interval or ratio scale. Regression has response variable is categorical, it can be differentiated into two groups, that is response variable with two categories and response variable with more than two categories. Response variable which has two categories value namely by response variable of binary, while response variable which has more than two categories named by response variable of multinomial.[1,2,5,6,8,] In Lee et al. (2013) said that binary response variable can be made binary regression. Binary regression is regression has response variable yield binary outcomes. Binary outcomes be able to be produced from predictor variable with data measurement scale is arbitrary.

Link function is function of multiple linear regression that is linear toward both predictor variable and parameter. Link functions in this writing are logit link function, normit link function and cloglog link function. Logit link function has domain $[0,1]$ and range $(-\infty,\infty)$, while normit function is quantile function related with normal standard distribution, that is normal distribution with has mean 0 and standard deviation 1. Mathematically, normit function is reciprocal of cumulative distribution function of normal standard distribution.[2,7,9,12,13,19] By O'connor et al. (2016) explained that if the random variable X has both the probability density function $f(x)$ and cumulative distribution function $F(x)$, then It has reliability function at x. The reliability of random variable X is $R(x)$ where It be defined by a complement of probability of cumulative distribution function.

Some of journals related with this paper that are in Rozzaghi (2013) discuss logistic regression model be compared with probit regression model applied at data mining. Logistic regression model initially comes from logit transformation, while probit regression model initially comes from probit transformation. The result that logit transformation can be collected to class of canonic link function, but probit transformation can not be collected to class of canonic link function. Then at Cakmayapan, et al., (2013) explained that comparison logit regression model and probit regression model by Monte Carlo simulation with different sample size, different cutoff point, and different correlation between dependent variable and independent variable. This material discusses models, residuals, deviations, and Pseudo-R square. Comparison link function with binary response variable about symmetric assumption that probit link function better than logit link function. Because value of AIC of probit link function smaller than value of AIC of logit link function. While for loglog link function better than cloglog link function. Because value of AIC of loglog link function smaller than value of AIC of cloglog link function (Damisa, et al. 2017). At Shariff, et al. (2009) that logit regression model more robust than probit regression model for data are not normal distribution.

This paper aims to find significant factors which cause diabetes mellitus disease, then determine ordinal regression model. Also, probability and reliability function of a patient has stroke disease toward albuminuria level. The method used to three link functions, that are logit link function, normit link function, and cloglog link function. The best models are resulted by three link function, respectively. Test of homogeneity prediction result of link functions uses linear hypothesis test. These factors cause to diabetes mellitus and stroke could be used to prevent diseases, in order to all persons are healthy and happy.

2. Material and method

2.1. Material

These data are secondary data about patient of diabetes mellitus and stroke diseases. Source of data from section of medical record a hospital at Semarang middle java. We take a sample random of 95 patients of the incidence and risk factor diabetes mellitus and stroke. The research variables consist of age, sex, systolic blood pressure (sbp), diastolic blood pressure (dbp), body mass index (bmi), smoke, alcoholic (ach), high density lipoprotein cholesterol (hdl), low density lipoprotein cholesterol (ldl), albuminuria (albu), and diabetes mellitus (dm). Variable dm = 1 if yes and = 0 if no, sex = 1 if male and = 0 if female, smoke = 1 if current smoking, = 0 otherwise, ach = 1 current drinking and = 0 if not, albu =1 if macroalbuminuria, = 2 if normal, = 3 if microalbuminuria. First, the response variable is diabetes mellitus, and the factors are age, sex, systolic blood pressure, diastolic blood pressure, body mass index, smoke, alcoholic, high density lipoprotein cholesterol, low density lipoprotein cholesterol, albuminuria. Second, the response variable is albu, and the covariates are bmi and hdl.

2.2. Method

We make relation between link function toward ordinal regression model and reliability function. Let Y_i is the outcome of the *i*-th patient. Assume that Y_i is classified into the *k*-th level, that is $k = 1, 2, \ldots, m$. Suppose that for each n patients are measured covariate variables, $\mathbf{x} = (x_{i1}, x_{i2}, \ldots, x_{ip})'$. These variables have either qualitative or quantitative type. Ordinal regression model and reliability function can be made by link function. The link functions chosen are logit link function, normit link function, and complementary log-log link function. Formally, ordinal regression model and reliability function can be served as follow.[8,9,10,14,19]

Thre are three types of link function, that are:

a. Logit link function

The logit link function is

$$\log\left(\frac{P(Y_i \le k \mid \mathbf{x}_i)}{1-P(Y_i \le k \mid \mathbf{x}_i)}\right) = a_k + \sum_{j=1}^{p} b_j x_{ij},\ k = 1,2,\ldots,m-1.$$

If we put on $u_{ki} = a_k + \sum_{j=1}^{p} b_j x_{ij}$, that function is equivalent with

$$P(Y_i \le k \mid \mathbf{x}_i) = \frac{\exp\left(a_k + \sum_{j=1}^{p} b_j x_{ij}\right)}{1+\exp\left(a_k + \sum_{j=1}^{p} b_j x_{ij}\right)} = \frac{\exp(u_{ki})}{1+\exp(u_{ki})},\ k = 1,2,\ldots,m-1.$$

Therefore,

$$P(Y_i = k \mid \mathbf{x}_i) = P(Y_i \le k \mid \mathbf{x}_i) - P(Y_i \le k-1 \mid \mathbf{x}_i)$$

$$= \begin{cases} \dfrac{\exp(u_{1i})}{1+\exp(u_{1i})} & k = 1 \\ \dfrac{\exp(u_{ki})}{1+\exp(u_{ki})} - \dfrac{\exp(u_{k-1i})}{1+\exp(u_{k-1i})} & k = 2,\ldots,m-1 \\ 1 - \dfrac{\exp(u_{m-1i})}{1+\exp(u_{m-1i})} & k = m \end{cases}$$

and

$$R(Y_i = k \mid \mathbf{x}_i) = 1 - P(Y_i \le k \mid \mathbf{x}_i) = 1 - \frac{\exp(u_{ki})}{1+\exp(u_{ki})},\ k = 1,2,\ldots,m-1.$$

b. Normit link function

The normit link function is

$$\Phi^{-1}(P(Y_i \le k \mid \mathbf{x}_i)) = a_k + \sum_{j=1}^{p} b_j x_{ij},\ k = 1,2,\ldots,m-1.$$

If we put on $u_{ki} = a_k + \sum_{j=1}^{p} b_j x_{ij}$, that function is equivalent with

$$P(Y_i \le k \mid \mathbf{x}_i) = \Phi(u_{ki}),\ k = 1,2,\ldots,m-1.$$

So

$$P(Y_i = k \mid \mathbf{x}_i) = P(Y_i \le k \mid \mathbf{x}_i) - P(Y_i \le k-1 \mid \mathbf{x}_i)$$

$$= \begin{cases} \Phi(u_{1i}) & k = 1 \\ \Phi(u_{ki}) - \Phi(u_{k-1i}) & k = 2,\ldots,m-1 \\ 1 - \Phi(u_{m-1i}) & k = m \end{cases}$$

where $\Phi(.)$ is the cumulative standard normal distribution function,

and

$$R(Y_i = k \mid \mathbf{x}_i) = 1 - P(Y_i \le k \mid \mathbf{x}_i) = 1 - \Phi(u_{ki}),\ k = 1,2,\ldots,m-1.$$

c. Complementary log-log link function

The complementary log-log (cloglog) link function is

$$\log(-\log(1-P(Y_i \le k \mid \mathbf{x}_i))) = a_k + \sum_{j=1}^{p} b_j x_{ij},\ k = 1,2,\ldots,m-1,$$

If we put on $u_{ki} = a_k + \sum_{j=1}^{p} b_j x_{ij}$, that function is equivalent with

$$P(Y_i \le k \mid \mathbf{x}_i) = 1 - \exp(-\exp(u_{ki})), \quad k = 1, 2, \ldots, m-1.$$

Then

$$P(Y_i = k \mid \mathbf{x}_i) = P(Y_i \le k \mid \mathbf{x}_i) - P(Y_i \le k-1 \mid \mathbf{x}_i)$$

$$= \begin{cases} 1 - \exp(-\exp(u_{1i})) & k = 1 \\ \exp(-\exp(u_{k-1i})) - \exp(-\exp(u_{ki})) & k = 2, \ldots, m-1 \\ \exp(-\exp(u_{m-1i}) & k = m \end{cases}$$

and

$$R(Y_i = k \mid \mathbf{x}_i) = 1 - P(Y_i \le k \mid \mathbf{x}_i) = \exp(-\exp(u_{ki})), \quad k = 1, 2, \ldots, m-1.$$

3. Result and discussion

First, processing data analysis by logit link function. We take on ten covariates, that are age, sex, sbp, dbp, bmi, smoke, ach, hdl, ldl, and albu, while as a response variable is dm. This research want to know about some covariate cause dm significantly. After processing data, the result that covariates age, sex, sbp, smoke, ach, idl aren't significant toward dm, but covariates bmi, hdl, and albu are significant toward dm. The process table result is presented as follow.

Table 1. Result of all factors cause diabetes mellitus

Covariate	Estimate	z-value	Pr(>| z |)	Explanation
Intercept	0.4543	0.130	0.896	-
age	0.0165	0.481	0.631	Not significant
sex	0.0814	0.135	0.893	Not significant
sbp	-0.0209	-0.887	0.375	Not significant
dbp	0.0035	0.101	0.920	Not significant
bmi	0.0687	1.659	0.097	Significant
smoke	-0.4875	-0.835	0.404	Not significant
ach	-0.2857	-0.456	0.649	Not significant
hdl	-0.0678	-2.520	0.012	Significant
ldl	-0.0030	-0.419	0.675	Not significant
albu	1.6705	3.286	0.001	Significant

We take on level of significance is $\alpha = 10\%$. This result can be informed that a person has a diabetes mellitus disease is caused by variables bmi, hdl, and albu. Body mass index is ratio of high and weight. A person should have a body ideal, thai is he has bmi approximate 3. High density lipoprotein cholesterol is bad, so It should be prevented in order healthy, and, albuminuria that means is macroalbuminuria. A patient has macroalbuminuria if he has high albuminuria in his urine. Then we look for a the best model of this problem, that is as response variable is dm and significance covariate variables are bmi, hdl, and albu. The link function still by logit link function. Its result is in table 2 below.

Table 2. Result of significant factors cause diabetes mellitus

Covariate	Estimate	z-value	Pr(>| z |)	Explanation
Intercept	-1.3857	-0.851	0.3946	-
bmi	0.0681	1.819	0.0389	Significant
hdl	-0.0661	-2.741	0.0061	Significant
albu	1.4581	3.510	0.0005	Significant

In this time we take on level of significance is $\alpha = 5\%$. The best model can be constructed that is

$$\text{dm} = -1.3857 + 0.0681\ \text{bmi} - 0.0661\ \text{hdl} + 1.4581\ \text{albu}$$

In this research we assume that if patients have diabetes mellitus disease, they have stroke disease, too. So, if we study stroke disease, the covariate albu be chosen as response variable at ordinal regression model. We conclude that variable albu as a response variable of ordinal regression, covariates bmi and hdl are predictor variables. Variable albu is divided by three levels, that are albu =1 if macroalbuminuria, albu = 2 if normal, and albu = 3 if microalbuminuria. This discussion used three link functions. The link functions are the logit link function, the normit link function, and the complementary log-log link function. Next, the research wants to count probability of patient with albu = 1, albu = 2, and albu = 3, by link functions, respectively. We process these data and the result be presented about coefficients and intercepts at table 3.

Table 3. Coefficients and intercepts of ordinal regression by link function

Coefficients	Link function		
	logit	normit	cloglog
bmi	0.0471	0.0260	0.0170
hdl	0.0064	0.0022	-0.00081
Intercepts			
albu=1|albu=2	2.569	1.3954	0.6265
albu=2|albu=3	3.6264	2.0003	1.1649

We use that result to discuss ordinal regression models and reliability function by logit link function, normit link function, and cloglog link function. This problem there are three categories, so $k = 1,2,3$ or $m = 3$. It processing use the three link functions by $k = 1,2,3$, respectively. Explaination models be presented as follow.

1. By the logit link function

   For k = 1

   The ordinal regression model is

$$P(Y_i = 1 \mid \mathbf{x}_i) = \frac{\exp(A)}{1 + \exp(A)}$$

   Where $A = 2.569 + 0.0471bmi + 0.0064hdl$

   The reliability functionn is

$$R(Y_i = 1 \mid \mathbf{x}_i) = 1 - \frac{\exp(A)}{1 + \exp(A)}.$$

   For k = 2

   The ordinal regression model is

$$P(Y_i = 2 \mid \mathbf{x}_i) = \frac{\exp(B)}{1 + \exp(B)} - \frac{\exp(A)}{1 + \exp(A)}$$

   Where $B = 3.6264 + 0.0471bmi + 0.0064hdl$

   The reliability functionn is

$$R(Y_i = 2 \mid \mathbf{x}_i) = 1 - \frac{\exp(B)}{1 + \exp(B)}$$

   For k = 3

   The ordinal regression model is

$$P(Y_i = 3 \mid \mathbf{x}_i) = 1 - \frac{\exp(A)}{1 + \exp(A)}$$

The reliability function is

$$R(Y_i = 3 | \mathbf{x}_i) = 0.$$

2. By the nomit link function
For k = 1
The ordinal regression model is

$$P(Y_i = 1 | \mathbf{x}_i) = \Phi(C)$$

Where $C = 1.3954 + 0.026bmi + 0.002175hdl$
The reliability function is

$$R(Y_i = 1 | \mathbf{x}_i) = 1 - \Phi(C)$$

For k = 2
The ordinal regression model is

$$P(Y_i = 2 | \mathbf{x}_i) = \Phi(D) - \Phi(C)$$

Where $D = 2.0003 + 0.026bmi + 0.002175hdl$
The reliability function is

$$R(Y_i = 2 | \mathbf{x}_i) = 1 - \Phi(D)$$

For k = 3
The ordinal regression model is

$$P(Y_i = 3 | \mathbf{x}_i) = 1 - \Phi(C)$$

The reliability function is

$$R(Y_i = 3 | \mathbf{x}_i) = 0$$

3. By the complementary log-log link function
For k = 1
The ordinal regression model is

$$P(Y_i = 1 | \mathbf{x}_i) = 1 - \exp(-\exp(E))$$

Where $E = 0.6265 + 0.0170bmi - 0.00081hdl$
The reliability function is

$$R(Y_i = 1 | \mathbf{x}_i) = \exp(-\exp(E))$$

For k = 2
The ordinal regression model is

$$P(Y_i = 2 | \mathbf{x}_i) = \exp(-\exp(E)) - \exp(-\exp(F))$$

Where $F = 1.1649 + 0.0170bmi - 0.00081hdl$
The reliability function is

$$R(Y_i = 2 | \mathbf{x}_i) = \exp(-\exp(F))$$

For k = 3
The ordinal regression model is

$$P(Y_i = 3 | \mathbf{x}_i) = \exp(-\exp(E))$$

The reliability function is

$$R(Y_i = 3 | \mathbf{x}_i) = 0$$

3.1. Probability and Reliability of Outcomes by Link Functions

Application of ordinal regression models be tried to put on some arbitrary values about covariates of bmi, hdl. If we take on values both bmi = 3.2 and hdl = 52 obtained table 4.

Table 4. Probability values of categorical responses

Probability	Link Function		
	Logit	Normit	Cloglog
$P(Y_i = 1 \mid \mathbf{x}_i)$	0.95487740	0.944274	0.8495566
$P(Y_i = 2 \mid \mathbf{x}_i)$	0.02897333	0.041702	0.1114830
$P(Y_i = 3 \mid \mathbf{x}_i)$	0.04512259	0.055726	0.1504434

Based on table 4, we could explain that:
a. By three link functions, probability of a patient with macroalbuminuria has stroke greater than microalbuminuria, and a patient with microalbuminuria has stroke greater than normal.
b. Probability of patient with macroalbuminuria by logit, normit, and clogloc link functions is decrease, respectively.
c. Probability of patient with normal by logit, normit, and clogloc link functions is increase, respectively.
d. Probability of patient with microalbuminuria by logit, normit, and cloglog link functions is increase, respectively.

We use reliability function to count probability value about a patient has macroalbuminuria, normal, and microalbuminuria, respectively. The chosen covariates are bmi, hdl. If we take on arbitrary values such as, both bmi = 3.2 and hdl = 52 obtained table 5 below.

Table 5. Reliability values of categorical responses

Reliability	Link Function		
	Logit	Normit	Cloglog
$R(Y_i = 1 \mid \mathbf{x}_i)$	0.0451226	0.055726	0.1504434
$R(Y_i = 2 \mid \mathbf{x}_i)$	0.0161493	0.014024	0.0389604
$R(Y_i = 3 \mid \mathbf{x}_i)$	0	0	0

Based on table 5, we could say that:
a. By three link functions, reliability of a patient with macroalbuminuria, normal, and microalbuminuria have stroke, respectively, is decrease.
b. Reliability of patient with macroalbuminuria by logit, normit, and clogloc link functions, respectively, is increase.
c. Reliability of patient with microalbuminuria by logit, normit, and clogloc link functions, respectively, is zero.

3.2. Comparative of result probability values

This research aim to compare difference of probability values of three link functions by linear hypothesis. The hypothesis is written as follow.

Linear Hypothesis:

H_0: all of link function result estimation probability values are same.

H_1: at least one of link function result estimation probability value is different.

The output result is table 6.

Table 6. Comparison between methods

Method	Estimate	Std. Error	t-value	$\Pr(>\|t\|)$
B – A	0.00433	0.40020	0.011	1.000
C – A	0.0277	0.40020	0.069	0.997
C – B	0.0233	0.40020	0.058	0.998

If we take significant level α = 5%, then all of $\Pr(>|t|)$ values greater than 5%, so H_0 is accepted. It means that all of link function methods yield estimation probability value is same.
If above linear hypothesis is presented by visual, the figure is as follow.

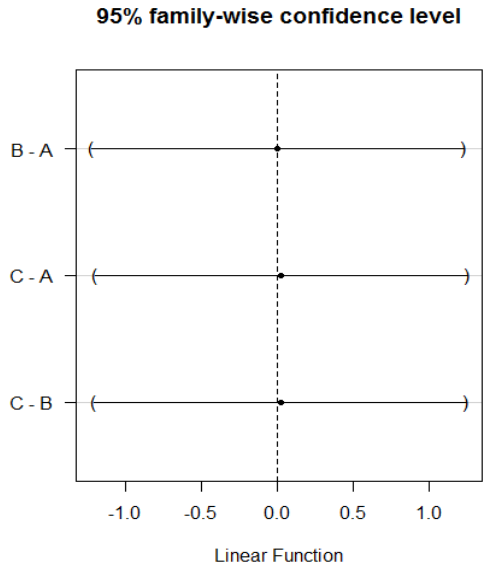


Figure 1. Differences between two link functions for predict probability.

3.3. AIC value
Akaike information criterion (AIC) (Akaike, 1969), is defined as

$$AIC = l(\hat{\mathbf{b}}) - 2p$$

where $l(\hat{\mathbf{b}})$: loglikelihood of β prediction, p: the number of parameters in distribution.

Criterion of AIC be used to measure exactly of choosing the best regression model. The indication of AIC is smaller AIC value is better for elected regression model to use [3]. AIC value of logit link function, normit link function, and cloglog link function are, respectively, 167.6826, 168.3965, 169.6107. These results are same by the result of linear hypothesis analysis that AIC values are not different meanwhile their AIC values are not equal. Therefore, logit model, normit model and cloglog model could be used to predict probalility with result almost same.

4. Conclusion

If we take covariates age, sex, systolic blood pressure, diastolic blood pressure, body mass index, smoke, alcoholic, high density lipoprotein cholesterol, low density lipoprotein cholesterol, and albuminuria to patients have diabetes mellitus, the factors cause that disease are body mass index, high density lipoprotein cholesterol, and albuminuria. The best model result is

$$\text{dm} = -1.3857 + 0.0681\,\text{bmi} - 0.0661\,\text{hdl} + 1.4581\,\text{albu}$$

Because diabetes mellitus be caused by body mass index, high density lipoprotein cholesterol and albuminuria. Then people in order to be avoided diabetes mellitus, they are suggested by controlling

obesity, level of cholesterol, and albumin in blood. This method by regulate diet, enough rest and regular exercise. The result of probability of patient stroke that by three link functions, probability of a patient with macroalbuminuria has stroke greater than microalbuminuria, and a patient with microalbuminuria has stroke greater than normal. Probability of patient with macroalbuminuria by logit, normit, and clogloc link functions is decrease, respectively. Probability of patient with normal by logit, normit, and clogloc link functions is increase, respectively. Probability of patient with microalbuminuria by logit, normit, and cloglog link functions is increase, respectively. While, the result of reliability of patient stroke that by three link functions, reliability of a patient with macroalbuminuria, normal, and microalbuminuria have stroke, respectively, is decrease. Reliability of patient with macroalbuminuria by logit, normit, and clogloc link functions, respectively, is increase. Reliability of patient with microalbuminuria by logit, normit, and clogloc link functions, respectively, is zero. By linear hypothesis obtained conclusion that all of link function methods yield estimation probability value is same. AIC value of logit link function, normit link function, and cloglog link function are, respectively, 167.6826, 168.3965, and 169.6107. These results are same by the linear hypothesis; meanwhile their AIC values are not equal. Therefore, logit model, normit model and cloglog model could be used to predict probability and reliability with result almost the same.

References

[1] Agresti A 2007 *An Introduction to Categorical Data Analysis* (New Jersey: John Wiley & Sons, Inc)
[2] Agresti A 2015 *Foundations of Linear and Generalized Linear Models* (New York: John Wiley & Sons)
[3] Akaike H 1969 *Annals of the Institute of Statistical Mathematics* 21 243-247
[4] Cakmayapan S and Goktas A 2013 *Journal of Social and Economic Statistics* 2(1)
[5] Conover W J 1980 *Practical Nonparametric Statistics* (Canada: John Wiley & Sons, Inc.)
[6] Cox D R 1970 *Analisys of Binary Data* (London: Methuen)
[7] Damisa S A, Bello S Y and Musa F N 2017 *Biomedical Statistics and Informatics* 2(4) 145–149
[8] Dobson A J 2001 *An Introduction to Generalized Linear Models* (London: Capman & Hall)
[9] Hardin J W and Hilbe J M 2007 *Generalized Linear Model and Extensions* (United State of America: Stata Corp LP)
[10] Kalbfleisch J D and Prentice R L 2012 *The Statistical Analysis of Failure Time Data* (John Wiley & Sons., Inc)
[11] Kozial J A and Feng A C 2016 *Journal Information Srtoke* (American Heart Association, Inc)
[12] Lawless JF 2003 *Statistical Models and Methods for Lifetime Data* (Canada: John Wiley & Sons, Inc.)
[13] Lee E T and Wang J W 2013 *Statistical Methods for Survival Data Analysis* (New Jersey: John Wiley & Sons, Inc.)
[14] Lu W 2007 *Life Data Analysis* 13(1) 75-90
[15] O'connor A N, Madurres M and Mosleh A 2016 *Probability Distributions Used in Reliability Engineering* (Maryland: University of Maryland College Park)
[16] Rotella C M, Pola L and Mannuci E 2015 *International of Journal of Endocrinology Metabolism.* 11(3) 137-144
[17] Rozzaghi M 2013 *Journal of Modern Applied Statistical Methods* 12(1) 1-164-169
[18] Shariff A A, Zaharim A and Sopian K 2009 *European Journal of Scientific Research* 27(4) 548-553
[19] Tsay RS 2005 *Analysis of Financial Time Series* (New Jersey: John Wiley & Sons, Inc.)
[20] World Health Organization Department of Noncommunicable Disease Surveilance 2010 *Definition, Diagnosis and Classification of Diabetis Mellitus and Its Complication* (Switzerland)