Statistika, Vol. 10 No. 2, 121 – 127
Nopember 2010

Some Results on Statistical Analysis from Unit of Record, Hospital Universiti Sains Malaysia (HUSM)

WAN MUHAMAD AMIR W AHMAD[1], NOR AZLIDA ALENG[2], ZALILA ALI[3]

[1,2]Jabatan Matematik, Fakulti Sains Dan Teknologi, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu.
[3]Pusat Pengajian Sains Matematik, Universiti Sains Malaysia, (USM), 11800 Pulau Pinang, Malaysia.
E-mail: wmamir@umt.edu.my

ABSTRACT

Most of the patients which is visiting HUSM for the treatment at the same time suffer with more than one diseases. Because of that, many of the researchers are trying to find the association of the factor that contribute to such a case. In this case study we are trying to find the association for a certain health factor. It's contribution will have a major impact in the area of medical statistics.
Keywords : Risk Estimation, logistic regression, MANCOVA

1. INTRODUCTION

Health sciences research always involved the study of relation or association between the qualitative variables such a relationship of hypertension status to diabetes, heart attack with cholesterol level and so on. These analysis that can be done through the categorical data analysis. The finding of the study especially in medical or health sciences is very important in order to get the information of the joining variables for the inference purposes. In medical research, this method is quite famous especially when the researcher is trying to get the information about association of the risk from one disease to another. The finding of the relation can be used to educate the public consciousness to be more awareness and the early action can be taken to prevent the spread of the disease starting from the early stage. In this study, we are using the data from the record department of Hospital University Science Malaysia.

2. MATERIAL AND METHODS

To accomplish the above said objective, the data's were collected from unit of record, Hospital University Science Malaysia (HUSM) and the statistical analyses includes pearson chi-square, risk estimation, logistic regression, clustered bar chart and MANCOVA. The identified variables are summarized in Table 1.

Table 1 Description of the Data

Variable Name	Variable Description and Coding
1. CHOLES	*Total cholesterol (mg/dl)*
2. ICCHD	*Incident of CHD during 6 years of follow-up (no/yes)*
3. BMI	*Body mass index (weigh(kg)/[height(m)]^2)*
4. HEARTATCK	*History of heart attack (no/yes)*
5. DIABETES	*Diabetes status (normal/diabetic)*
6. DIABP	*Diastolic blood pressure (mm Hg)*
7. SYSBP	*Systolic blood pressure (mm Hg)*
8. TANMED	*Taking anti-hypertensive medications (no/yes)*
9. HYPER	*Status of hypertension : Normal/hypertension*
10. WEIGHT	*Weight (kg)*
11. SMOKING	*Smoking status (no/yes)*
12. INSULIN	*Serum insulin level (IU/ml)*
13. GLUCOSE	*Serum fasting glucose level (mg/dl)*
14 HIDL	*High Density Lipo*

In this study, we have collected one thousand set of data based on the thirteen variables. All the variables with the details of explanation are given by the Table 2.

Results

Table 2: Characteristics of Variables Association

Association of Variables	***Pearson Chi-Square Value***	***P***
History of heart attack with		
a) Hypertension Status	*0.234*	*0.628*
b) Diabetes Status	*0.004*	*0.953*
c) Smoking Status	*0.067*	*0.795*
d) Incident of Coronary Heart Disease	*4.681*	***0.030****
e) Body Mass Index	*0.433*	*0.510*
Diabetes Status with		
a) Hypertension Status	*21.262*	***0.000****
b) Smoking Status	*2.193*	*0.139*
c) Incident of Coronary Heart Disease	*9.706*	*0.002*
d) Body Mass Index	*30.873*	***0.000****
Hypertension Status with		
a) Smoking Status	*0.718*	*0.397*
b) Incident of Coronary Heart Disease	*20.063*	***0.000****
c) Body Mass Index	*11.710*	***0.001****
Incident of Coronary Heart Disease with		
a) Body Mass Index	*0.131*	*0.717*

****Significant at*** $p < 0.05$

In order to get the relationship between the two variables we carry out the test of association between the variables. All the possible way of variables calculation is done carefully and all the results are being listed in the Table 3. We found that five variables have the significance association that are history of heart attack with the incident of coronary heart disease $(p = 0.030)$, diabetes status with hypertension status $(p = 0.000)$, diabetes with body mass index status $(p = 0.000)$, hypertension status with body mass index $(p = 0.001)$ and hypertension status with the incident of coronary heart disease $(p = 0.000)$. Table 4 summarizing the results of risk estimation of the significant selected variables. This analysis is conducted in order to get the clear view of how the variables related. The details of the explanation is given in Table 4.

Table 3 Results of Risk Estimation of Selected Variables

		Value	Explanation
1	*Odds ratio for family history of heart attack (no / yes)*	*1.446*	***Case control study**: The odds of developing Coronary Heart Disease among the patient that having heart attack is 1.5 times higher compare to among the non-heart attack.*
	For cohort Incident CHD = yes	*1.283*	***Cohort study:** Heart attack patient had 1.3 times higher risk of developing Coronary Heart Disease compared to non-heart attack.*
2	*Odds ratio for hypertension status (normotensive / hypertensive)*	*1.898*	***Case control study**: The odds of developing diabetes among the patient that having hypertension is 2 times higher compare to among the non-hypertension.*
	For cohort diabetes status =diabetic	*1.555*	***Cohort study:** Hypertension patient had 2 times higher risk of developing diabetic compared to normotensive person.*
3	*Odds ratio for BMI status (Normal/ High)*	*2.612*	***Case control study**: The odds of developing diabetes among the patient that having higher BMI is 2 times compare to a normal person.*
	For cohort diabetes status =diabetic	*1.796*	***Cohort study:** Higher BMI patient had 2 times higher risk of developing diabetic compared to normal BMI person.*
4	*Odds ratio for Incident CHD (Yes / No)*	*2.142*	***Case control study**: The odds of developing hypertension among the patient that having Incident of CHD is 2 times compare to a person that having no Incident CHD.*
	For cohort hypertension status =hypertension	*1.362*	***Cohort study:** Patient with incident of CHD had 1 times higher risk of developing hypertension compared to patient that never have the incident of CHD.*
5.	*Odds ratio for hypertension status (normotensive / hypertensive)*	*1.849*	***Case control study**: The odds of developing higher BMI among the patient that having hypertension is 2 times higher compare to among the non-hypertension.*
	For cohort BMI status=High	*1.683*	***Cohort study:** Patient with higher BMI had 2 times higher risk of developing hypertension compared to normotensive patient.*

Table 4. Final Model of the Associated Factor for Hypertension by Multiple Logistics Regression

	β	Df	S.E	Wald	p	Exp(β)
Total cholesterol (mg/dl)	0.0056	1	0.038	2.1401	0.044	1.0056
Incident of CHD during 2 years of follow-up (no/yes)	0.8648	1	0.3613	5.7292	0.017	2.3746
Diabetes status (normal/diabetic)	0.6336	1	0.3522	3.2369	0.072	1.8844
Diastolic blood pressure (mm Hg)	0.0321	1	0.155	4.3126	0.038	1.0326
Systolic blood pressure (mm Hg)	0.2206	1	0.0179	152.681	0.000	1.2469
Taking anti-hypertensive medications (no/yes)	5.7904	1	0.4701	151.704	0.000	327.14
Smoking status (no/yes)	0.2736	1	0.2709	0.0198	0.013	1.3147
Glucose level (mg/dl)	-0.0071	1	0.0057	1.5564	0.012	0.9929
Constant	-34.1943	1	2.8115	147.923	0.000	

β : Constant beta coefficient Wald: Wald statistics p: p value
S.E: Standard Error Df: Degree of Freedom Exp(β): exponent beta

In this study we also are modeling the hypertension status into a mathematical model by using logistics regression model. From the modeling method we come out a significant and appropriate model that is given by Table 5. The full model can be written as shown in (1)

$$\begin{aligned}\text{logit}(\hat{\pi}) = &-34.1943 - 0.0071\text{GLUCOSE} + 0.2736\text{SMOKING} + 5.7904\text{TANMED} \\ &+ 0.2206\text{SYSBP} + 0.0321\text{DIABP} + 0.6336\text{DIABETES} \\ &+ 0.86448\text{ICCHD} + 0.0056\text{CHOLES}\end{aligned} \quad (1)$$

From the mathematical model above, we can also define factors that contribute to the hypertension. All the independents variable in the model can be assume as factor since they have a strong correlate to the hypertension status.

Table 5. Final Model of the Associated Factor for Diabetes by Multiple Logistics Regression

	β	Df	S.E	Wald	p value	Exp(β)
Glucose level (mg/dl)	0.5924	1	0.0551	115.564	0.0000	1.8084
Serum insulin level (IU/ml)	0.0497	1	0.0213	5.4497	0.0196	1.0509
Body mass index (weigh(kg)/[height(m)]^2)	-0.0759	1	0.0502	2.2865	0.0305	0.9269
Constant	-63.5166	1	5.9677	113.281	0.0000	0.9269

β : Constant beta coefficient Wald: Wald statistics p: p value
S.E: Standard Error Df: Degree of Freedom Exp(β): exponent beta

For the diabetes status we can write the model as

$$\text{logit}(\hat{\pi}) = -63.5166 - 0.0759\text{BMI} + 0.0497\text{INSULIN} + 0.5924\text{GLUCOSE} \quad (2)$$

All the significant p values in the Table 4 and Table 5 show that the variables have the strong association in giving the significant results. We also plot the clustered bar chart for every significant variable and the results are given in Figure 1 to Figure 5.

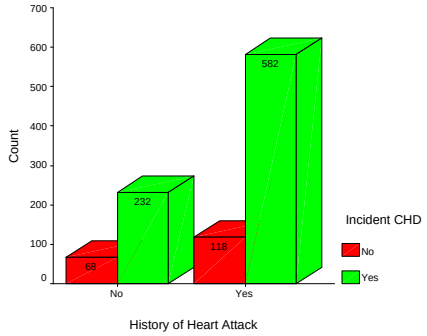

Figure 1 Clustered Bar Charts of Incidence CHD with History of Heart Attack

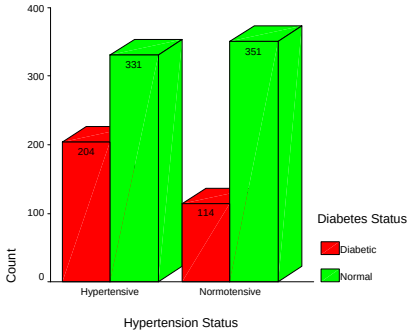

Figure 2 Clustered Bar Charts of Diabetes Status with Hypertension Status

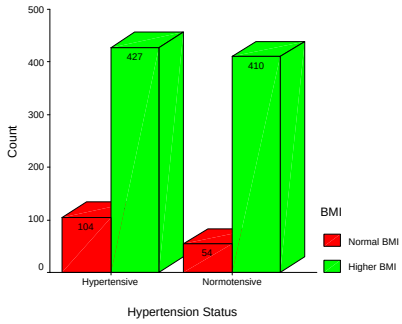

Figure 3 Clustered Bar Charts of BMI Stage with Hypertension Status

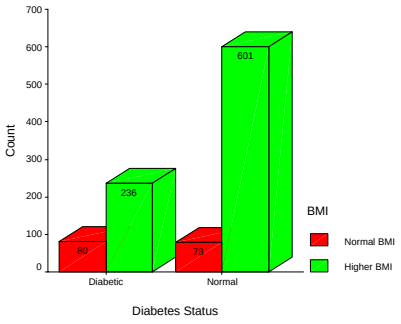

Figure 4 Clustered Bar Charts of Diabetes Status with BMI Stage

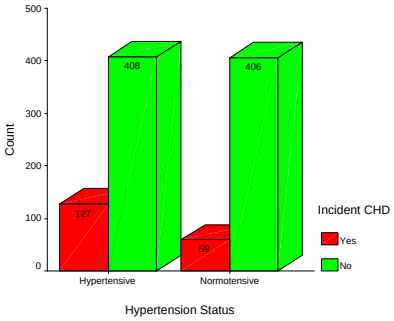

Figure 5 Clustered Bar Charts of Incidence CHD with Hypertension Status

Figure 1 shows the clustered bar charts of patients which are having history of heart attack with incident of coronary heart disease. 582 of patients which are have history of heart attack also having the incident of coronary heart disease. This indicates that there are strong connections or association between histories of heart attack with incident of CHD. The clustered bar charts in Figure 2 show the diabetes status of patient with hypertension. We can see clearly that hypertension status occurs most frequent in patients who are suffering from diabetes and clustered bar charts Figure 3 shows that most of the patient who are suffering from hypertension having higher BMI. This indicates that people who are having higher BMI are risk to get hypertension. Lastly Figure 4 shows that hypertension status is related most to the incident of CHD compared to normotensive patients.

We run the MANCOVA analyses and the results are given by the Table 6.

Table 6. MANCOVA Analysis For Selected Variables.

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Body mass index	1158.979[a]	9	128.775	10.692	.000
	HIDL cholesterol	5995.655[b]	9	666.184	4.646	.000
	Serum fasting glucose	598603.390[c]	9	66511.488	157.191	.000
Intercept	Body mass index	193342.793	1	193342.793	16052.798	.000
	HIDL cholesterol	606774.487	1	606774.487	4231.620	.000
	Serum fasting glucose	4051565.100	1	4051565.100	9575.302	.000
INSULIN	Body mass index	134.209	1	134.209	11.143	.001
	HIDL cholesterol	1586.659	1	1586.659	11.065	.001
	Serum fasting glucose	18694.735	1	18694.735	44.182	.000
SMOKING	Body mass index	98.145	2	49.072	4.074	.017
	HIDL cholesterol	645.430	2	322.715	2.251	.106
	Serum fasting glucose	1020.936	2	510.468	1.206	.300
DIABETES	Body mass index	424.526	2	212.263	17.624	.000
	HIDL cholesterol	2280.742	2	1140.371	7.953	.000
	Serum fasting glucose	248408.005	2	124204.003	293.539	.000
SMOKING * DIABETES	Body mass index	39.969	4	9.992	.830	.506
	HIDL cholesterol	994.800	4	248.700	1.734	.140
	Serum fasting glucose	1291.787	4	322.947	.763	.549
Error	Body mass index	11923.738	990	12.044		
	HIDL cholesterol	141956.681	990	143.391		
	Serum fasting glucose	418895.361	990	423.127		
Total	Body mass index	711151.147	1000			
	HIDL cholesterol	2414474.000	1000			
	Serum fasting glucose	13437041.000	1000			
Corrected Total	Body mass index	13082.717	999			
	HIDL cholesterol	147952.336	999			
	Serum fasting glucose	1017498.751	999			

a. R Squared = .089 (Adjusted R Squared = .080)

b. R Squared = .041 (Adjusted R Squared = .032)

c. R Squared = .588 (Adjusted R Squared = .585)

The output indicates that there is a significant different:

- between the sample mean of insulin and the sample mean of body mass index.
- between the sample mean of insulin and the sample mean of HIDL cholesterol.
- between the sample mean of insulin and the sample mean of serum fasting glucose.
- between the sample mean of body mass index and the level of smoking.
  - The mean of BMI index for patient with never smoke is 26.46.
  - The mean of BMI index for patient with former smoke is 26.63.
  - The mean of BMI index for patient with current smoke is 25.21
- between the sample mean of body mass index and the level of diabetes.
  - The mean of BMI index for non-diabetes patient is 25.7961.
  - The mean of BMI index for borderline diabetic patient is 27.53.
  - The mean of BMI index for diabetic patient is 27.91.
- between the sample mean of HIDL cholesterol and the level of diabetes.
  - The mean of HIDL cholesterol for non-diabetes patient is 48.84.
  - The mean of HIDL cholesterol for borderline diabetic patient is 45.39.
  - The mean of HIDL cholesterol for diabetic patient is 44.71.
- between the sample mean of serum fasting glucose and the level of diabetes.
  - The mean of serum fasting glucose for non-diabetes patient is 98.04.
  - The mean of serum fasting glucose for borderline diabetic patient is 115.18.
  - The mean of serum fasting glucose for Diabetic patient is 165.60.

3. CONCLUSION

This study emphasis the effect of common issue of health problem that are diabetes and hypertension. A number of patients have been proposed for computational of getting the significant results. From the study we have show clearly how the significant results were related. All the significant results are being summarized in Table 1 to Table 7 and Figure 1 to Figure 5. All the results from our studies can be used to educating the public about the important of health.

References

Belle, V. G., Fisher, L. D., Heagerty, P. J., Lumley, T. (2004). Biostatistics: A Methodology for the Health Sciences. 2nd ed. Wiley, New Jersey.

Pagano, M., Gauvreau, K. (2000). Principles of Biostatistics. 2nd ed. United States of America .Duxbury.

Kirkpatrick, L. A., Feeney, B. C. (2001). A Simple Guide to Spss for Windows. Revised Edition. Chicago: Spss Inc.

Kutner, M. H., Neter, J. & Wasserman, W. (1990). Applied Linear Statistical Models, 3rd ed. United States of America: R. R. Donnelley & Sons Company.

Montgomery, D. C. (2001). Design and Analysis of Experiments, 5th ed. New York: John Wiley & Sons.

Zar, J. H. (1999).Biostatistical Analysis. 4th ed. New Jersey. Prentice Hall.