Imperial Journal of Interdisciplinary Research (IJIR)
Vol-3, Issue-6, 2017
ISSN: 2454-1362, http://www.onlinejournal.in

Statistical Analysis to Investigate Incidence of Hypertension among Sudanese Patients: A Case Study of Patients Attending to Al Gadarif Hospital, Al Gadarif State, Sudan (2016)

Dr: Adel Ali Ahmed [(1)], Dr: Elmogiera F. Elsyed Elawad [(2)] & Ust: Salma ALfadil Hasab Sido[(3)]

[1] Associate Professor, University of Gezira, Faculty of Economics and Rural Development, Department of Applied Statistics and Demography.

[2] Associate Professor, University of Gezira, Faculty of Economics and Rural Development, Department of Applied Statistics and Demography

[3] Assistance Professor, Gadarif University, Faculty of Economics, Department of Applied Statistics and Population Studies.

Abstract: *This study depends mainly on primary data through a sample of (100) patients (cases) and (100) control from AlGadarif hospital in AlGadarif state. This study aims to apply logistic regression to discover the relationship between predictor's variables and dependent variable hypertension incidence. Study methodology included chi-square test to know characteristics of study population. From the result family history, sleepless, use of salt, age, sex, body mass index, residence, smoking, job, marital status and use of medicine are significant. The researcher applied logistic regression to match the results. The study recommended that guidelines to detect and control of hypertension through identify the risk factors of hypertension among individuals who have hypertension life style changed including stopping smoking, limiting salt, maintaining a healthy weight and awareness about the dangerous of disregard the disease.*

1. Introduction:

Regression models have become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables.

The aims of application of any methods focus on data and statistical reasoning rather than on either the presentation of as many methods as possible or the mathematical reasoning should be the most important objective of any researcher. The statistical analysis of lifetime of response time data has become a topic of considerable interest of statistical and workers in different areas.

Logistic regression is part of category of statistical models called generalized linear models. This broad class of models includes ordinary regression and analysis of variance as well as covariance and log linear regression.

Logistic regression allows one to predict outcome from a set of variable that may be continuous, discrete and discontinuous or a mix of any of these. Generally, the dependent or response variable is discontinuous such as presence/absence or success/failure. Discriminate analysis can only be used wit continuous independent variable. Thus, in instances where the independent variables are categorical, or mix of continuous and categorical, logistic regression is preferred (David. W. Hosmar, 2006).

The use of logistic regression modeling in biomedical research, business and finance, criminology, ecology, engineering, health, policy, linguistics and wild life biology.

2. Study problem:

High blood pressure, also known as hypertension affects millions even children and teens. Blood pressure reading are usually given as two numbers 120 over 80 (written as 120/80 mmHg) one or both of these numbers can be too high (David zieve, 2011).

Logistic regression is one of the most popular and effective ways to analyze data with binary outcome. This study used logistic regression model to assess the probability of Hypertension.

Logistic regression model techniques become useful in many types of life because it is means of using data to gain insight into real problems.

3. Importance of the study:

The importance of the study comes from the importance of the hypertension, because it worried the human and threats them. Our emphasis here is on the role of statisticians to express the dangerous of hypertension through statistical functions.

4. Study objectives:

1. To reveal the concept of hypertension through logistic regression.
2. To determine the factors that affects the hypertension.
3. To apply the logistic regression so as to discover the relation between variable denotes hypertension.

5. Study methodology:

Source of date: A questionnaire was launched for data collection, patients and controls. In particular this questionnaire designed to collect on risk factors variables. Questionnaire covers all the types of variables. The areas of control were selected randomly.

Analysis of data: The purpose of analysis of data is to investigate the relationship (dependency) between hypertension and risk factors (independent). Results are reported using chi-square for dependency. To estimate the coefficient for the models and the other statistical test so as to distinguish between the variables that related to hypertension we used (SPSS) Statistical Package and service solutions.

Statistical model: Logistic regression will used to dependent variable (hypertension) and independent variables: (Sex – Age – Tribe – Residence – Level of education Marital status – Family history – Alcohol – Smoking – Not being physically active – Use of salt – Sleepless - Chronic - Heigh and Weight).

Sampling design and sample size:

The study based on randomly sample that will be collected from population under hypertension and cases control in Gadrif hospital. Then the sample size for this study is 100 patients.

6. DATA ANALYSIS AND APPLICATION:

6.1 Statistical analysis:

The comparison between two groups (case and control) of data were made for all the independent variables, the categorical independent variables were managed in arranged in table and chi-square tests were applied to compare the difference between case and control. The preliminary analysis of the data set was carried out.

Family History: The difference between patients and control group statistically was significant, (P-value = 0.0001) 64% for patients and 32% for control are yes while 36% of patients and 63% of control are no. The different between the two groups is significant (P-value = 0.0001).

Use of Salt: The difference between patients and control group statistically was significant (P-value = 0.0001) 32% for patient and 83% of control are normal usage of salt while 66% of patients and 6% of control usage little of salt while 2% of patients and 11% of control usage too much salt. The difference between the two groups is significant (P-value = 0.0001).

Sleepless: The difference between patients and control group statistically was significant (P-value = 0.0001) 48% for patients and 6% of control are yes and 52% of patients and 94% control are no. The difference between the two groups is significant (P-value = 0.0001).

Sex: The difference between patients and control group statistically was significant (P-value = 0.006) 23% for patients and 42% for control are male while 77% of patients and 58% are female. The difference between the two groups is significant.

Residence: The difference between patients and control group statistically was significant (P-value = 0.022) 65% of patients and 48% of control are live in city while 35% for patients and 52% of control are live in village. The difference between the two groups is significant (P-value = 0.022).

Smoking: The difference between patients and control group statistically was significant (P-value = 0.025) 17% for patients and 6% of control are yes while 23% of patients and 94% of control are no. The difference between the two groups is significant (P-value = 0.025).

Body Mass Index (BMI): The difference between patients and control group statistically was significant (P-value = 0.039) 29% for patients and 30% of control have normal (BMI) while 6% of patients and 18% of control have abnormal (BMI) while 34% of patients and 34% of control are have fat (BMI) while 17% of patients and 12% of control are have medium (BMI) and 14% of patients and 6% of control have overweight (BMI). The difference between the two groups is significant (P-value = 0.039).

Job: The distribution of patients and control according to job represent most of the patients and control are unemployed 77% for patients and 63% of control while 23% for patients and 37% of control are employed. The difference between the two groups is significant (P-value = 0.044).

Tribe: The difference between patients and control group statistically was not significant (P-value = 0.517) 21% for patients and 19% of control are north while 28% of patients and 21% of control are west, while 14% of patients and 40% of control are middle. The difference between the two groups is not significant (P-value = 0.517).

Level of Education: The difference between patients and control group was not significant (P-value = 0.793) 42% of patients and 35% of control illiterate while 12% of patients and 11% of control are khalowa, while 28% of patients and 34% of control are primary, while 12% of patients and 15% of control are secondary, and 6% of patients and 5% of control are university. The difference between the two groups is not significant (P-value = 0.793).

Alcohol: The difference between patients and control group was not significant (P-value = 0.497) 0% of patients and 0% of control are yes while 100% of patients and 100% of control are no. The difference between the two groups is not significant (P-value = 0.497).

Marital Status: The distribution of patients and control according to marital status was representing most of the patients and controls are married. 72% of patients and 79% while 6% of patients and 9% of control are single. While 4% of patients and 4% of control are divorced and 18% of patients and 8% of control are widow. The difference between the two groups is not significant (P-value = 0.189).

Activity: The distribution of patients and control according to activity was representing most of the patients and controls are no. 86% of patients and 90% of control, while 14% of patients and 10% of control are yes. The difference between the two groups is not significant (P-value = 0.515).

Use of Medicine: The difference between patients and control group statistically was not significant (P-value = 0.256) 59% of patients and 50% of control are yes while 41% of patients and 50% of control are no. The difference between the two groups is not significant (P-value = 0.256).

Age Group: The difference between patients and control group statistically was significant (P-value = 0.004) 1% of patients and 4% of control age group, 2% of patients and 6% of control age group2, 3% of patients and 17% of control age group3, 20% of patients and 22% of control age group4, 23% of patients and 21% of control age group5, 16% of patients and 12% of control age group6, 16% of patients and 13% of control age group7, 11% of patients and 3% of control age group8, 8% of patients and 2% of control age group9, . The difference between the two groups is not significant (P-value = 0.004).

Table (1):Chi-square test for dependency between hypertension and independent variable

Variable		Case		Control		P-value
Family history	Yes	64	64%	37	37%	0.00**
	No	36	36%	63	63%	
	Total	100	100%	100	100%	
Use of salt	Normal	32	32%	83	83%	0.00**
	Little	66	66%	6	6%	
	Too much	2	2%	11	11%	
	Total	100	100%	100	100%	
Sleepless	Yes	48	48%	6	6%	0.00**
	No	52	52%	94	94%	
	Total	100	100%	100	100%	
Sex	Male	23	23%	42	42%	0.006**
	Female	77	77%	58	58%	
	Total	100	100%	100	100%	
Residence	City	65	65%	48	48%	0.022*
	Village	35	35%	52	52%	
	Total	100	100%	100	100%	
Smoking	Yes	17	17%	6	6%	0.025*
	No	83	83%	94	94%	
	Total	100	100%	100	100%	
Body Mass Index	Abnormal	6	6%	18	18%	0.039*
	Normal	29	29%	30	30%	
	Fat	34	34%	34	34%	
	Medium	17	17%	12	12%	
	Over weight	14	14%	6	6%	

	Total	**100**	**100%**	**100**	**100%**	
Job	**Employed**	**23**	**23%**	**37**	**37%**	**0.044***
	Unemployed	**77**	**77%**	**63**	**63%**	
	Total	**100**	**100%**	**100**	**100%**	
Tribe	**North**	**21**	**21%**	**19**	**19%**	**0.517**
	West	**28**	**28%**	**21**	**21%**	
	East	**14**	**14%**	**20**	**20%**	
	Middle	**37**	**37%**	**40**	**40%**	
	Total	**100**	**100%**	**100**	**100%**	
Level of education	**Illiterate**	**42**	**42%**	**35**	**35%**	**0.793**
	Khalwa	**12**	**12%**	**11**	**11%**	
	Primary	**28**	**28%**	**34**	**34%**	
	Secondary	**12**	**12%**	**15**	**15%**	
	University	**6**	**6%**	**5**	**5%**	
	Total	**100**	**100%**	**100**	**100%**	
Alcohol	**Yes**	**0**	**0%**	**2**	**2%**	**0.497**
	No	**100**	**100%**	**98**	**98%**	
	Total	**100**	**100%**	**100**	**100%**	
Marital status	**Married**	**72**	**72%**	**79**	**79%**	**0.189**
	Single	**6**	**6%**	**9**	**9%**	
	Divorced	**4**	**4%**	**4**	**4%**	
	Widow	**18**	**18%**	**8**	**8%**	
	Total	**100**	**100%**	**100**	**100%**	
Activity	**Yes**	**14**	**14%**	**10**	**10%**	**0.515**
	No	**86**	**86%**	**90**	**90%**	
	Total	**100**	**100%**	**100**	**100%**	
Use of medicine	**Yes**	**59**	**59%**	**50**	**50%**	**0.256**
	No	**41**	**41%**	**50**	**50%**	
	Total	**100**	**100%**	**100**	**100%**	
16-23	**group (1)**	**1**	**1%**	**4**	**4%**	**0.004****
24-31	**group (2)**	**2**	**2%**	**6**	**6%**	
32-39	**group (3)**	**3**	**3%**	**17**	**17%**	
40-47	**group (4)**	**20**	**20%**	**22**	**22%**	
48-55	**group (5)**	**23**	**23%**	**21**	**21%**	
56-63	**group (6)**	**16**	**16%**	**12**	**12%**	
64-71	**group (7)**	**16**	**16%**	**13**	**13%**	
72-79	**group (8)**	**11**	**11%**	**3**	**3%**	
80-87	**group (9)**	**8**	**8%**	**2**	**2%**	
	Total	**100**	**100%**	**100**	**100%**	

Source: Own calculation

was binary. Its value (1) represented the case and (0) the control. All the variables measured on hypertension were tested by using chi-square test. Some of the significant and strong associations were mentions.

6.2 Logistic regression analysis:

All data were coded and entered into computer. Data were scrutinized for advanced analysis. For this research on hypertension the outcome variable

Table (2) Family history model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Family history	-1.108	-294	14.214	0.00	3.030	0.186 – 0.588
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Family history	262.495		14.764			0.00

Source: Own calculation.

According to the table (2) above, Wald statistic = (14.214) with P-value (0.00) which statistically is significance it means the variable (family history) influence the hypertension. The odd ratio (0.030) and 95% CI (0.186 – 0.588) which mean the people have family history are (0.03) times more than those who have no family history, this can be support by the test of the overall significance of the model chi-square = (14.764) with P-value = (0.00) which means that the model is statistically significant.

Table (3) Sleepless model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Sleepless	-2.67	0.466	32.833	0.00	14.5	0.028 – 0.172
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			Sig
Sleepless	227.82		49.441			0.00

Source: Own calculation.

According to the table (3) above, Wald statistic = (32.833) with P-value (0.00) which indicate the statistically significance for the coefficient. Independent variable (sleepless) influence the hypertension. The odd ratio (14.5) which means respondents from hypertension (14.5) times more than other.

This can be support by the test of the overall significance of the model chi-square = (49.441) with P-value = (0.00) so the model is statistically significant.

Table (4) Use of salt of model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Use of salt	1.236	0.269	21.072	0.00	3.44	2.030 – 5.831
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Use of salt	252.722		24.537			0.00

Source: Own calculation.

According to the table (4) above, Wald statistic = (21.072) with P-value (0.00) which indicate statistically significance for the coefficient. The independent variable (use of salt) influence hypertension incidence positively. The odd ratio (3.44) which means people who use of salt too much have positive hypertension (3.44) times more than who use of salt normal and little, this can be support by the test of the overall significance of the model chi-square = (24.537) with P-value = (0.00) which mean that the model is statistically significant.

Table (5) Age model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Age	0.350	0.086	16.641	0.00	1.199	1.679
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Age	258.465		18.794			0.00

Source: Own calculation.

The estimated model of group is shown in table (5) above, and also the test of the model. From the table reports Wald statistic = (16.641) with P-value (0.00). So the age statistically is significance. From the result group plays an important role to effect the hypertension. The odd ratio (1.419) which means group have positive effect to hypertension. The overall model (chi-square) = (18.794) with P-value (0.00), which mean that the model statistically is significant.

Table (6) Sex model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Sex	0.886	0.312	8.042	0.005	2.424	1.315 – 4.471
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Sex	268.94		8.319			0.00

Source: Own calculation.

From the estimated coefficient of the model in table (6) Wald statistic = (8.042) with P-value (0.005) which mean sex statistically is significance. This variable (sex) plays an important role to affect hypertension. Odd ratio (2.424) this mean female (2.424) times more than male. The overall model chi-square = (8.319) with P-value = (0.004) indicate that the model statistically is significant.

Table (7) Residence model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Residence	-0.699	0.290	5.817	0.016	2.01	0.282 – 8.77
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Residence	271		5.911			0.015

Source: Own calculation.

From the estimated coefficient of the model in table (7) Wald statistic = (5.817) with P-value (0.016) the residence statistically is significance at (1%) level of significance. From this result residence plays an important role to hypertension. Odd ratio = (2.01) which means that those who live in the city has affect (2.01) times more than those who live in village. The overall model chi-square = (5.91) with P-value = (0.015) which mean that the model statistically is significant.

Table (8) Smoking model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Smoking	-1.186	0.498	5.477	0.019	3.209	0.117 – 0.877
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			Sig
Smoking	271.092		6.166			0.013

Source: Own calculation.

According to the table (3.8) Wald statistic = (5.477) with P-value (0.019) that the coefficient statistically significance. The independent variable (smoking) influence hypertension. The odd ratio (3.209) and 95% CI (0.117 – 0.877) which means those who smoking has positive hypertension (3.209) times more than those who no smoking. This can be support by the test of the overall significance of the model chi-square = (6.166) with P-value = (0.013) so the model is statistically significant.

Table (9) Job model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Job	0.676	0.315	4.600	0.032	1.966	1.060 – 3.647
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Job	272.560		4.699			0.030

Source: Own calculation.

According to the table (9), Wald statistic = (4.600) with P-value (0.032) which indicate the statistically significance of the coefficient. The odd ratio (1.699) which means people unemployed has positive (1.699) times more than employed. This can be support by the test of the overall significance of the model chi-square = (4.699) with P-value = (0.030) which that the model is statistically significant.

Table (10) Body Mass Index model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Body Mass Index	0.371	0.131	7.960	0.005	1.449	1.120 – 1.875
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Body Mass Index	266.805		0.454			0.033

Source: Own calculation.

The estimated model of Body Mass Index is shown in table (10) above, and also the test model. From the table reports Wald statistic = (7.960) with P-value (0.005) so body mass index statistically is significant. This result of body mass index plays an important role to affect hypertension. The odd ratio (1.449) which indicates that those who have high body mass index (1.449) times more than other types. This over all model chi-square = (10.454) with P-value = (0.033) which mean that the model is statistically significant.

Table (11) Marital status model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Marital status	0.249	0.139	3.214	0.073	1.283	0.972 – 1.685
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			Sig
Marital status	273.931		3.327			0.068

Source: Own calculation.

From the estimated coefficient of the model in table (11) Wald statistic = (3.214) with P-value (0.073) which is significant and the variable play an important role to the development of hypertension. To test the overall model significance chi-square = (3.327) with P-value = (0.0680) mean that the model statistically is significant.

Table (12) Use of medicine:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Use of medicine	0.364	0.285	1.629	0.202	1.439	0.397 – 1.45
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Use of medicine	275.623		1.636			0.201

Source: Own calculation.

From the estimated coefficient of model in table (12) Wald statistic = (1.629) with P-value (0.202) which it use of medicine statistically is significance and the variable play an important role to develop hypertension. To test the overall model significance chi-square = (1.636) with P-value = (0.201) which mean that the model statistically significant.

Table (13) Tribe model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Tribe	-0.066	0.088	0.562	0.454	1.086	0.787 – 1.113
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Tribe	276.696		0.562			0.00

Source: Own calculation.

From the estimated coefficient of the model in table (13) above, Wald statistic = (0.562) with P-value (0.454) which it mean tribe statistically is not significance and the variable cannot play an important role to develop of hypertension. To test the overall model significance chi-square = (0.562) with P-value = (0.453) it mean that the model statistically it is not significant.

Table (14) Education model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Level of education	-0.100	0.112	0.798	0.372	1.105	0.276 – 1.127
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig

Level of education	276.458	0.801	0.371

Source: Own calculation.

From the education coefficient of the model in table (14) Wald statistic = (0.798) with P-value (0.372) which is means education statistically is not significant and the variable cannot play an important role to improve hypertension. To test the overall model significance chi-square = (0.801) with P-value = (0.371) means that the model is statistically is not significant.

Table (15) Alcohol model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Alcohol	2.122	2.842	0.00	0.999	1.648	0.00
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			sig
Alcohol	274.466		2.793			0.00

Source: Own calculation.

From the estimated coefficient of the model in table (15) Wald statistic = (0.00) with P-value (0.999) which is not statistically significant and variable cannot play an important role to improve of hypertension. To test the overall model significance chi-square = (2.793) with P-value = (0.095) means that the model statistically is not significant. This variable from risk factors of hypertension often not really.

Table (16) Activity model:

The estimated model coefficients						
Effect	B	SE	Wald	Sig	OR	95% CI for OR
Activity	-0.382	0.441	0.751	0.336	1.465	0.228 – 1.619
The log likelihood (Test of the Model)						
Effect	-2 log likelihood		Chi-square			Sig
Activity	276.498		0.761			0.383

Source: Own calculation.

From the estimated coefficient of the model in table (16) Wald statistic = (0.0751) with P-value (0.386) which is activity statistically is not significant and the variable cannot play an important role to develop hypertension. To test the overall model significance chi-square = (0.761) with P-value = (0.383) which mean that the model statistically not significant.

7. CONCLUSION AND RECOMMENDATIONS

7.1 Conclusion:

The technique of logistic regression is most commonly used for analysis of epidemiological studies. Logistic regression models have been applied to this case-control research and the results obtained from univariate model.

From the results of the models, family history, sleepless, use of salt, age, sex, body mass index, residence, smoking, job, marital status and use of medicine is significant which mean these variables play an important role to develop hypertension. The result can be acceptable and it considered as a risk factors of hypertension.

From the result we found that tribe, level of education, alcohol and activity are not significant which mean these variables do not play an important role to affect of the hypertension.

7.2 Recommendations:

The research has the following recommendations:

- Guidelines to detect and control hypertension through identify the risk factors of hypertension.
- Among individuals who have hypertension life style such stopping smoking, limiting salt, maintaining a healthy weight.
- The awareness about the dangers of the disease.
- The case-control data used in this research, together with hypertension registry data can be used to construct such models of absolute risk.
- The interrelation between hypertension incidence and risk factors of hypertension need to be evaluated by the other studies.

References:

1. Christopher B.S.(2006):Introduction to mathematical modeling of crop growth, how the equation are derived and assembled in to a computer model, Ph.D., dissertation, Faculty of Agriculture, University of Putra, Malaysia.
2. Davied. W. Hosmer and Stanley Lemeshow(2006): Applied logistic regression, Second edition, Wiley, Inc, New York.
3. John Neter, Michaeld. Kutner, Christopher. J.N. achtsheim and William Wassermah(1996): Applied linear Statistical Models, Fourth edition.
4. Michey, J. and Greenland,S.(1989): A study of the impact of confounder selection criteria on effect estimation. American Journal of Epimiology.
5. Rothmand . K. J and Greenlend. S.(1998): Modern epidemiolpgy, third edition Lippincott. Raven. Philadelphia.