The Use of The Binary Logistic Regression Model for Determining Factors That Affect on Anemia Patients in Misan City

Sabah Hasan Jasim

College of Basic Education- Misan University

Abstract

Anemia is a decease or disorder that effect on men caused by lower in account of RBCs of deceased the ratio of Hb in blood . This article demonstrates some factors that may be effect on anemia patients in misan city by use application of binary logistic regression applied by SPSS version (22). About one hundred blood samples **50** case infected and **50** of no infected cases with anemia disease taken from laboratories in misan city. Results show that HGB ,sex and RBC was statistically significant while other factors was not significant in table(3) , HGB P-value (0.00) ,df (1) ,Exp B (0.481) also RBC P-value(0.003), df (1) ,Exp(7.732) finally the sex p-value (0.002). df (1), Exp (4.947) therefore HGB , sex and RBCs was more effect than other factors . In ratio of area under the ROC curve was (0.842%) that mean was a good results. The aim of the study is to give aresearchers and students, with guideline for using probability infected with diseases by logistic modeling in searches.

Key- words : Logistic regression, anemia, binary logistic, logistic modeling.

استخدام نموذج الانحدار اللوجستي الثنائي في تحديد العوامل الموثره على مرضى فقر الدم في محافظه ميسان

صباح حسن جاسم / قسم الرياضيات / كلية التربية الأساسية / جامعة ميسان

الخلاصة

يعتبر فقر الدم من الأمراض الشائعة التي تؤثر على صحة الإنسان نتيجة لقلة المعدل الطبيعي لكريات الدم الحمراء أو نتيجة لقلة في نسبه الهيموغلوبين الدم . تضمن هذا البحث دراسة بعض العوامل المؤثرة على مرضى فقر الدم في مركز محافظه ميسان حيث تم جمع العينات من احد المختبرات في المحافظة. وباستخدام نموذج الانحدار اللوجستي الثنائي و البرنامج الإحصائي SPSS (22) حيث كانت عينات الدراسة 100 عينة من الدم ، 50 عينة مصابه بمرض بفقر الدم و 50 عينه غير مصابين بمرض فقر الدم كمجموعه ضابطه للبحث واعتماداً على القيمة الثنائية (1،0) وسجلت النتائج في جدول رقم (3) بأن نسبة الدلالة الإحصائية لهيموغلوبين الدم (HGB) P-value = 0.000 عند أرجحيه = EXP (B) 0.481 ودرجة حرية df =1 أما بالنسبة للدلالة الإحصائية للجنس (SEX) P-value = 0.002 عند نسبة أرجحيه EXP (B) =7.732 عند درجة حرية df=1 وأخيراً فأن نسبة الدلالة الإحصائية لكريات الدم الحمر (RBCs) P-value = 0.003 عند أرجحيه EXP (B) = 4.947 ودرجة حرية df =1 لذلك يعتبر عامل نسبة الهيموغلوبين والجنس وكريات الدم الحمر هي من أكثر العوامل المؤثرة في مرضى فقر الدم أما بالنسبة لعوامل كريات الدم البيضاء (WBCs) والصفائح الدموية (PLC) فلا تعتبر من العوامل الموثره إحصائيا وليس لها تأثير على مرض فقر الدم . إن الغرض من هذه الدراسة هو لمساعدة الباحثين والدارسين والأكاديميين باستخدام الانحدار اللوجستي الثنائي كدليل في دراسة البحوث .

مفتاح الكلمات : الانحدار اللوجستي ، فقر الدم ، اللوجستي الثنائي ، النموذج اللوجستي .

1-Introduction:

Anemia Is a lower in a total amount of red blood cells or hemoglobin(Hb) in the blood .or hemoglobin concentration below establishes cutt-off level. The symptoms of anemia are feeling with tired short of breath, weakness, or a low ability to do exercises [1].It is widespread public health as economic and social development although the prevalence of anemia different and fitting data are often lacking and given in aresource poor areas significant proportions of young and women of childbearing age are anemic [2].

There are threee kinds of anemia are dive to decreasee red blood cell production and increased red blood cell break down causes of anemia were blood loss include trauma and gastro-intestinal bleeding decreased include thalasemia , low iron count in blood , loss of vitamin B12 and other causes [3].It can be classified according to amount of hemoglobin and the size of red blood cells in each blood cell. If the ratio of Hb in the blood cells are small it is causes amicrocystic b anemia or the large ratio it is a Macrocytic anemia. Diagnosis of anemia disease in men is based on a hemoglobin ratio is less than (13 to 14 g/dL), while in women it must be less that (12 to 13g/dL) [4].

The logistic regression was proposed as use in late 1960 and 1970 and it became routinely uses in a statistical packages in the early 1980 [5].Since that time the use of logistic regression in especially in higher education and social sciences [6]. Finally the logistic regression is describing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables [7].

The Aims of this search use the binary logistic regression as a model for detect probability of some factors that association with anemia and use this modal for descriptive analysis for these factors and applied logistic modeling by spss (22) as guild line for researchers.

2- Concept of logistic regression models:

"The word of logistic regression mean also logistic model or logit model describes the relationship between multiple independent variables and dependent variable, and estimates the probability of event by use of logistic curve to fit data there are two types of models at logistic regression binary logistic regression and multinomial logistic the binary logistic regression is use when dependent variable was (dichotomous) and independent variables are categorical or continuous .In multiple logistic regression may be employed.The logistic regression is popular model and it based on provides estimates in range Zero ,One and give a S-shape description of effect of some risk factors on risk event" [8].

2.1. The Odds:

Is the ratio of the probability that an event will ooccur and refer to the probability of an events is p or will not occur it is mean the probability of the event not is $(1-P)$. Then the odds value give as:

$$" \text{ odds of } \{\text{Event}\} = \frac{P}{1-P} \quad "$$

In logistic regression of variable p is also probability outcome and an explanatory variable x is model relating p and x in the equation $P = a + \beta x$. this is a simple logit model and refer is not a good model because extreme values of x will give values of $a + \beta x$ that does not fall between 0 and 1 .The logistic regression solution to this problem is to transform the odds using the natural logarithm [9]. Logistic regression we model the natural log odds as a linear function of the explanatory Variable:

$$\text{"logit(y)} = \ln(\text{odds}) = \ln\left(\frac{P}{1-P}\right) = a + \beta\chi \quad \ldots\ldots\ldots\ldots \quad (1)\text{"}$$

Taking the antilog of equation (1) on both sides, one can derive an equation for the prediction of the probability of the occurrence of interested outcome as

$$P = P\ (Y = \text{Interested Outcome}/X = \chi, \text{aspecific vlaue})$$

$$= \frac{e^{A+\beta\chi}}{1+e^{A+\beta\chi}} = \frac{1}{1+e^{-(A+\beta\chi)}}$$

Extending the logic of the simple logistic regression to multiple predictors, one may construct a complex logistic regression as

$$\text{"logit}(Y) = \ln\left(\frac{P}{1-P}\right) = a + \beta_1 X_1 + \ldots + \beta_K X_K\text{"}$$

$$\text{“}P = P\ (Y = \text{Interested Outcome}/X_1 = X_1, \ldots X_K = X_K)\text{"}$$

$$\text{“} = \frac{e^{A+\beta_1 X_1 + \ldots + \beta_K X_K}}{1+e^{A+\beta_1 X_1 + \ldots + \beta_K X_K}} = \frac{1}{1+e^{-(A+\beta_1 X_1 + \ldots + \beta_K X_K)}}\text{”}$$

2.2. odds ratio (OR) :

Is a comparative measure of two odds relative to different events two events A and B, the corresponding odds of(A ,B)occurring is

$$\text{"Odds ratio } \{A\text{vs.}B\} = \frac{\text{odds}\{A\}}{\text{odds}\{B\}} = \frac{P_{A/(1-P_A)}}{P_{B/(1-P_B)}}\text{"}$$

An (OR) association between an exposure and an out-come. The ratio of if "OR=1 indicates does not affect odds of outcome. and OR>1 indicates exposure with higher odds of outcome. But OR<1indicates exposure associated with lower odds of outcome. For example, the variable infected with anemia is coded as1 (yes = infected) and 0 (no= non infected), Logistic regression is one way to generalize the OR beyond two -binary variables" [10].

3. The logistic regression-curve graph (S-shape):

"Logistic regression is a method for accurate a regression curve ,y =f(x), when y consists of binary coded (0, 1 noninfected , infected) with anemia disease . When the response is a binary variable and x is numerical, and logistic curve to the relationship between x and y. Logistic-curve graph is like S-shaped or sigmoid curve" [11].

A logistic curve begins in three stages , first slowly and then as a linear growth, finally as exponential growth line, then the line slowly again to a stable line [12].

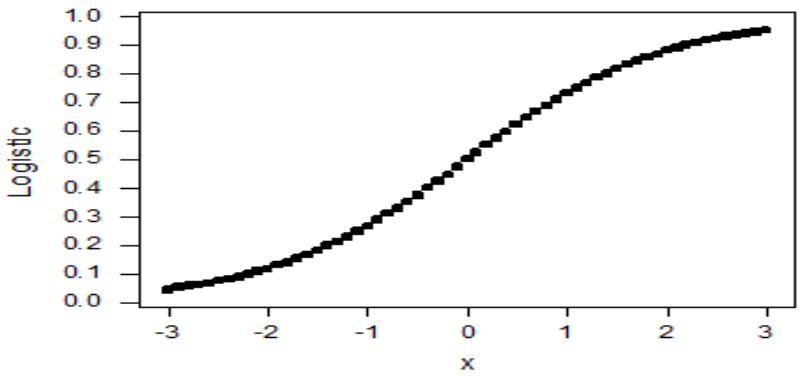

Figure (1): Logistic regression curve graph as S-Shape

5- Discussion and ROC curve data analysis:

In this study use a binary logistic regression test in patients of anemia disease based on some factors. This research hypothesis which predicted that which more factors that probability to causes anemia. In table (1) the ratio of Cox & Snell R^2 and Naelkerke R^2 ranges from 29.35% to 39.1% . In classification table (2) the cut value is p< 0.5 this mean that if the probability of cases classify in to yes category is greater than $p < 0.5$ or no category and the ratio of sensitivity and Specificity for these results was (74%).Table (3) show variable in the equation is use to determining statistical significance for each the independent variable also sex $P = 0.002$, $Exp = 4.947$, RBCs $P = 0.003$,$Exp = 7.732$, and HGB

$P = 0.000\ Exp\ (0.481)$ all in df=1 that mean HGB ,sex and RBCs are more affect factors in infected cases while WBCs $p = 0.610$, Exp =0.946 , and PLT $P = 0.253$,$Exp = 0.996$ was not statistical significance . Finally by use table (3) data the logic equation was as the following:

$$"\ln(\frac{P(x)}{1-P(x)})" = 0.249 + 1.599X_1 - 0.056X_2 + 2.045X_3 - 0.732X_4 - 0.004X_5$$

The odds ratio in gender male more time by 4.947 times more than female that infected by anemia disease.

Roc curve is mean receiver operating characteristics the performance of classification schemes that categorize cases in two groups. The Roc curve is refer as visual index of accuracy of tests in searches [13].The area under the curve (AUC) represents the probability that the assay as a result of randomly chosen negative cases .In AUC a table (4) a significant a ratio of each model is less than (*0.05)* so all are doing better results. In figure (2) and table(4) show the AUC was (*0.824*) and asymptotic significance was (*0.00*) that mean good results.

"Table (1): Model summary"

"Model Summary"			
Step 1	*"-2 Log likelihood"*	*"Cox & Snell R Square"*	*"Naelkerke R Square"*
	103.930a	*293*	*391*
"a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001."			

"Table (2): Classification table for infected and non-infected cases"

" Classification Table a "					
			"Predicted"		
			case		*"Percentage Correct"*
	"Observed"		*non infected*	*infected*	
Step 1	*case*	*non infected*	*37*	*13*	*74*
		infected	*13*	*37*	*74*
	"Overall- Percentage"				*74*
"a. The cut value is .500"					

"Table (3) Model of the Equation"

"Variables in the Equation"							
		B	*S.E.*	*wald*	*Df*	*Sig.*	*Exp-(B)*
Step 1[a]	*sex(1)*	*1.599*	*.518*	*9.512*	*1*	*.002*	*4.947*
	WBEs	*-.056*	*.110*	*.260*	*1*	*.610*	*.946*
	RBCs	*2.045*	*.682*	*8.982*	*1*	*.003*	*7.732*
	HGB	*-.732*	*.201*	*13.199*	*1*	*.000*	*.481*
	PLT	*-.004*	*.004*	*1.304*	*1*	*.253*	*.996*
	Constant	*.249*	*1.772*	*.020*	*1*	*.888*	*1.282*
"a. Variable(s) entered on step" 1: sex, WBCs, RBCs, HGB, PLT.							

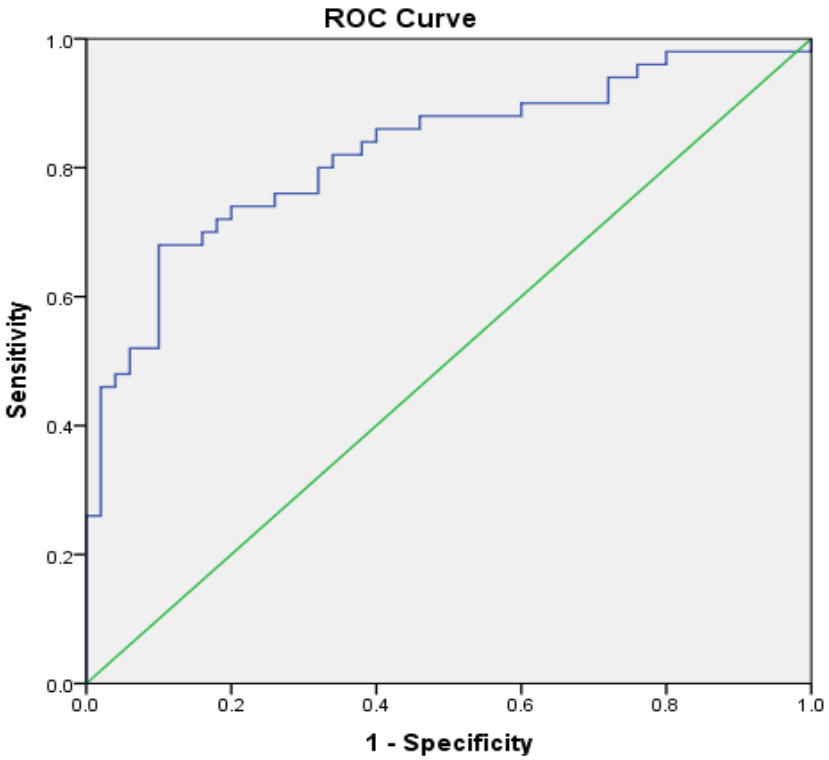

Figure (2): Specificity and sensitivity in Roc curve graph in anemia patients

"Table (4): The ratio of Predicted probability and Area under curve (AUC)"

"Area Under the Curve"				
"test Result Variable(s): Predicted probability"				
"Area"	*"Std. Error"*	*"Asymptotic Sig.[b]"*	*"Asymptotic 95% Confidence Interval"*	
			"Lower Bound"	*"Upper Bound"*
.824	*.042*	*.000*	*.742*	*.907*
"a. Under the nonparametric assumption"				
"b. Null hypothesis: true area = 0.5"				

Index Patients Samples and some factors that affect on Anemia

	CAS	SEX	WBC	RBC	HGB	PLT	PRE*	PGR*
	0	1	8.70	4.70	14.00	192	.15537	0
	0	1	7.50	4.56	12.30	189	.34182	0
	1	2	6.00	4.03	10.20	166	.82905	1
	1	2	5.80	3.91	10.30	180	.77058	1
	0	1	10.20	4.42	13.20	227	.12854	0
	1	2	5.60	10.70	36.00	176	.02455	0
	1	1	9.20	3.64	9.60	163	.36708	0
	1	2	7.30	4.36	10.70	181	.85208	1
	1	2	5.50	4.18	9.20	182	.92938	1
	1	1	10.00	3.39	5.40	483	.64675	1
	0	1	10.20	4.61	11.70	245	.37650	0
	0	1	9.30	4.72	14.00	224	.13912	0
	0	1	6.20	4.11	12.60	158	.16941	0
	1	2	7.10	4.23	9.20	215	.92047	1
	1	2	5.00	3.65	10.30	184	.66985	1
	1	2	4.90	4.45	10.00	177	.93079	1
	0	2	7.71	3.74	12.40	295	.21895	0
	0	1	11.30	4.54	13.00	131	.23628	0
	0	1	8.60	4.81	13.80	208	.20028	0
	0	1	8.53	4.27	12.40	158	.22328	0
	1	2	7.26	4.15	9.70	216	.87056	1
	0	1	9.26	5.75	14.80	240	.40911	0
	1	2	4.80	3.91	9.50	204	.85202	1
	0	2	7.89	4.48	11.90	252	.68600	1
	0	1	6.76	5.21	15.70	151	.16655	0
	0	2	6.38	4.30	12.60	198	.55390	1

	1	2	9.20	3.65	9.40	240	.70922	1
	1	2	11.60	5.57	10.70	232	.97740	1
	0	1	9.80	4.35	11.60	330	.21348	0
	0	2	9.99	4.06	11.70	248	.49209	0
	1	2	4.70	3.89	9.40	214	.85140	1
	0	1	8.85	5.10	15.00	255	.13185	0
	0	2	9.12	4.55	12.30	297	.59161	1
	0	2	10.89	4.14	11.40	436	.37689	0
	0	1	6.94	4.54	13.00	163	.25618	0
	0	2	9.84	4.11	11.40	219	.60411	1
	1	2	9.98	3.49	10.70	236	.39802	0
	0	2	7.64	4.99	13.60	173	.71752	1
	0	2	11.18	4.14	11.50	434	.35810	0
	0	2	13.28	5.06	14.00	271	.51193	1
	1	2	8.76	3.23	8.30	277	.66909	1
	1	2	7.93	3.68	10.70	135	.62752	1
	1	1	6.48	4.61	8.90	255	.84690	1
	1	1	10.58	4.78	10.40	283	.64817	1
	1	2	10.52	3.52	9.10	367	.55683	1
	1	2	16.28	4.23	10.30	447	.53434	1
	0	2	7.74	4.40	12.90	221	.50673	1
	0	1	7.15	4.61	14.00	192	.14302	0
	0	1	7.81	4.92	14.10	307	.14701	0
	0	1	5.69	4.40	13.80	178	.12654	0
	1	2	12.97	3.31	10.20	289	.30790	0
	0	1	6.98	4.45	11.80	202	.36803	0
	0	1	15.24	4.16	11.20	221	.22471	0
	0	2	8.71	3.74	12.30	269	.24169	0
	0	1	16.61	4.74	12.90	202	.21554	0
	1	2	7.15	4.10	9.50	213	.87751	1
	1	2	8.18	4.68	10.50	268	.89384	1
	1	2	9.00	10.40	31.90	210	.16372	0
	1	2	7.97	4.30	10.70	271	.76957	1
	1	1	9.07	4.68	8.30	457	.78313	1
	0	1	12.44	5.11	13.90	273	.20799	0
	1	1	11.82	3.88	8.70	274	.49587	0
	0	2	6.35	4.19	12.10	276	.50639	1
	1	2	8.20	3.98	10.90	286	.58127	1
	1	2	10.70	4.10	10.20	179	.80271	1
	1	2	8.10	3.79	9.80	312	.65450	1

	0	1	7.20	4.41	12.20	254	.24050	0
	0	1	7.61	4.58	12.80	162	.29506	0
	0	1	8.80	5.25	15.70	278	.10104	0
	1	2	11.43	4.23	8.40	271	.92773	1
	1	1	8.27	3.45	10.00	185	.21958	0
	1	1	9.61	3.85	10.90	246	.19083	0
	1	2	6.96	3.74	10.20	240	.64921	1
	1	2	10.51	3.72	10.60	292	.46525	0
	1	2	4.87	4.87	9.30	148	.98363	1
	1	2	9.12	4.47	10.40	223	.87147	1
	1	2	9.18	4.44	10.40	217	.86703	1
	1	2	5.59	3.73	9.30	392	.66387	1
	1	1	9.73	3.69	10.90	166	.19209	0
	1	1	4.69	3.81	10.40	162	.37150	0
	1	2	11.37	3.24	9.70	242	.42638	0
	1	1	6.72	3.37	10.40	201	.15366	0
	1	2	6.87	4.20	10.80	283	.71879	1
	1	2	7.19	5.23	10.30	194	.97755	1
	1	2	6.46	3.75	6.40	370	.94732	1
	1	2	11.22	4.08	10.20	283	.70860	1
	0	2	8.71	4.38	14.60	222	.21135	0
	0	1	9.17	3.99	11.40	362	.11967	0
	0	1	9.71	4.78	14.10	233	.13770	0
	0	2	12.48	4.43	12.20	324	.47372	0
	0	2	6.66	4.12	11.70	249	.56789	1
	0	1	7.43	4.96	13.70	170	.31506	0
	0	2	9.31	4.86	12.90	246	.68416	1
	0	1	11.24	4.85	14.10	259	.13145	0
	0	2	8.03	4.34	11.20	250	.73267	1
	0	2	6.28	4.60	11.10	240	.85245	1
	0	2	6.74	4.15	11.60	351	.49179	0
	0	1	6.13	4.40	13.30	188	.16337	0
	0	2	8.67	4.30	12.30	169	.60631	1
	1	2	5.62	3.66	9.80	132	.78272	1

PTL* Platelet Total Count WBCs*White Blood Cells RBCs*Red Blood Cells HGB* Hemoglobin Blood Concentration PRP* Predicted probability PGR* Predicted group

References:

1-Stedmans medical dicitionary (28th ed).Lippincott William and wilkin .2006.p.Anemia.ISBN 9780781733908.

2-WHO /unicef /unu.Iron deficiency anemia: assessment, prevention and control. Geneva, world Health organization, 2001.

3- Janz , TG; Johson , RL, Rubenstein ,SD; 2013: Anemia in the emergency department evaluation and treatment . Emergency medicine practice .15 (11):1-14;94i2 15-6.

4- Bhutla ZA, Das, JK,Rizvi , A,Gaffey MF;Walel, RE; 2013.Evidence based interventions for improrevement of material and child nutrition . Nutrition interventions Reviews Growp. 382(9890): 452-77.

5-Cabrea, A.F.(1994). Logistic regression analysis in higher education: An applied perspective .Higher Education Handbook of Theory and Research, Vol. 10. 225 -256.

6-Austin ,J.T; Yaffee ,R.A&Hinkle D,E;1992:Logestic regression for research in higher education .Higher education :Handbook of theory and Research , 8,379-410.

7- Peng , C.Y. Manz, B.D., & Keck ,J.(2001). Modeling categorical variables by logistic regression .American Journal of Health Behavior, 25(3),278-284.

8- Kleinbaum, D. G., & Klein, M., (2010). Logistic regression(statistics for biology and health) (3rd ed.). New York, NY: Springer-Verlag New York Inc.

9- Peng, C. J., Lee, K. L., & Ingersoll, G. M., (2002). An introduction to logistic regression analysis and reporting. The Journal of Educational Research, 96(1), 3-14.

10- Peng, C. J., & So, T. H., (2002). Logistic regression analysis and reporting: A primer. Understanding Statistics, 1(1), 31-70.

11- Eberhardt, L. L., & Breiwick, J. M., (2012). Models for population growth curves. ISRN Ecology, 2012, 1-7. http://dx.doi.org/doi:10.5402/2012/ 815016.

12- Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. Critical Care (London, England), 9(1), 112-118. http://dx.doi.org/ 10.1186/cc3045.

13- Honely ,J.A and B.J Mcneil ., 1982: The meaning and use of the Area under receiver operating characteristics curve .Radiology,143:1,29-36.