JKAU: Sci., **vol. 12,** pp.41-49 (1420 A.H. / 2000 A.D.)

Joint Influence of Several Factors on Blood Sugar Levels

B.M. Assas
Department of Statistics, Faculty of Science,
King Abdulaziz University, Jeddah, Saudi Arabia

Abstract. This paper is an application of linear regression analysis to a data set. The object is to study the joint effect of some factors on blood sugar level of diabetic patients. The factors included are likely to have an effect on blood sugar level. Data was randomly selected from two private clinics in the Holy City of Makkah. It represents 450 Saudi diabetic male and female patients. Factors considered here are age, bodymass, family history, number of children, marital status and food habits. The one factor regression models for blood sugar level on each factor listed above were studied. Then, accordingly the study pass on to considering the general multiple regression analysis for the combined (male and female) and separately for male and female patients. Finally, the stepwise regression technique is used to find out if it is desirable to include all the factors that we have listed or whether some factors can be eliminated as relatively unimportant.

Introduction

It is instructive to study the effect of some factors in isolation on the blood sugar level of diabetic patients. This has been done in many studies[1, 3-6]. For example they found that age, bodymass, family history and marital status etc. were important factors. We now embark on the study of the joint effects of several factors on the blood sugar level. The factors considered here are age, bodymass, family history, marital status and food habits. Data from 450 diabetic patients, of whom 220 were male and 230 female, was collected from two private clinics in the Holy City of Makkah. [500 questionnaires were distributed to patients: 450 replied].

Multiple regression analysis is performed where the blood sugar level is the dependent variable and the factors listed above are the concomitant variables. This is done separately for male and female patients so as to bring out any sex differential. We then try to reduce the set of concomitant variables by excluding those factors which are relatively less influential, using the stepwise regression technique[2]. A combined multiple regression analysis is then provided combining the male and female diabetic patients. Since some of these factors need coding (because they are qualitative) we discuss this first.

The Factors, Label and Coding

The following provides a list of labels and coding together with their explanations wherever needed.

Age	:	X_1 (in years)		
Bodymass	:	X_2	=	(Weight in kilograms) / (height in meters)^2 so that the unit is kg/m^2
Family history	:	X_3	=	1 if either father or mother of patient or both were also diabetic
			=	0, otherwise
Number of children	:	(X_4, X_5)	≡	(0,0) if there are no children
			≡	(1, 0) if there is at least one male child
			≡	(0, 1) if there is no male child but there are one or more female children.

In Saudi society as in various other societies parents suffer some stress if no children are born. Also absence of a male child sometimes causes concern. We can recognize three broad types exposed to different sorts of stress. The above coding (in preference to counting of male and female children) is done to differentiate the 3 classes.

Marital status	:	(X_6, X_7)	≡	(0, 0) if the person stays married at the time of interview (This also includes 7 unmarried persons out of 450 patients).
			≡	(1, 0) if the person is divorced
			≡	(0, 1) if the person is widow/ widower
Food habit	:	(X_8)	=	1 if the food is high in carbohydrate

	=	0 if the food is low in carbohydrate
(X_9)	=	1 if the food is high in protein
	=	0 if the food is low in protein
(X_{10})	=	1 if the food is high in fat
	=	0 if the food is low in fat

The above coding, using dummy variables 0 and 1, will allow one to make relative comparisons between the coefficients for these factors in the full regression model

General Linear Regression Analysis

We first provide the consolidated one factor regression of Y on X for each of X_1, X_2,...,X_{10} together with the coefficient of determination R^2 (in %). The value of R^2 expressed in percentage gives the amount of variation in Y which can be explained by the Factor X. The regression line $Y = \alpha + \beta X$ together with R^2 for each of the factors are as follows:

Age	:	$Y = 450 - 3.34\ X_1$	$R^2 = 8.7\%$
Bodymass	:	$Y = 49.4 + 7.12\ X_2$	$R^2 = 28.2\%$
Family history	:	$Y = 83.3 + 123\ X_3$	$R^2 = 35.0\%$
Number of children	:	$Y = 339 - 133\ X_4$	$R^2 = 41.7\%$
		$Y = 246 + 86.3\ X_5$	$R^2 = 16.2\%$
Marital status	:	$Y = 257 + 111\ X_6$	$R^2 = 16.9\%$
		$Y = 262 + 102\ X_7$	$R^2 = 12.1\%$
Food habits	:	$Y = 271 + 11.5\ X_8$	$R^2 = 0.3\%$
		$Y = 270 + 15.1\ X_9$	$R^2 = 0.5\%$
		$Y = 258 + 32.3\ X_{10}$	$R^2 = 24\%$

It would appear that, considered singly, number of male children (reduced to 0, 1 or absence or presence of male child) and family history account for the variation in blood sugar level more effectively than bodymass in our data set. The classification of high and low amounts of fat explains only about 24% of the blood sugar level. However left at that it may give a misleading picture of the way in which these factors may operate on the blood sugar level when considered simultaneously. We accordingly pass on to considering the general multiple regression analysis combined (male and female) and then separately for male and female patients.

The fitted multiple regression for males and females combined together is as follows:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10}$$

$$= 237 + 0.001\, X_1 + 2.87\, X_2 + 65.4\, X_3 - 89.5\, X_4 - 23.1\, X_5 + 59.2\, X_6 + 50.6\, X_7 + 2.98\, X_8 - 10.2\, X_9 + 19.5\, X_{10} \quad (1)$$

Thus age, bodymass, family history, divorce, being widow/widower, higher level of carbohydrate and fat all affect the blood sugar level positively. Presence of male and female children and high protein diet all affect blood sugar level negatively. Signs of the regression coefficient often create interpretational problems. However, in this case, the signs are exactly as one would expect. Further details are also interesting. The size of the regression coefficients for age and bodymass (X_1 and X_2) are affected by the units in which they are measured. Notice the large regression coefficient relating to male child (X_4) as compared to the regression coefficient relating to female child (X_5). This is due to the special importance attached to the presence of at least one male child. Likewise comparing the coefficient of (X_8) carbohydrate and (X_{10}) fat we notice that the effect of fat is roughly six times more than that of carbohydrate. Whereas the contribution to blood sugar level for a high level of carbohydrate is nearly 3 units, that for fat is 19.5 units. Divorce (X_6) and being a widow/widower (X_7) seems to produce similar sort of stress level for diabetic patients (compare the coefficients 59.2 and 50.6).

The fitted regressions for male and female patients are as follows. For females it is

$$Y = 238.87 + 0.175\, X_1 + 2.72\, X_2 + 75.7\, X_3 - 84.4\, X_4 - 29.8\, X_5 + 45.2\, X_6 + 41.7\, X_7 + 3.4\, X_8 - 2.25\, X_9 + 24.6\, X_{10} \quad (2)$$

and for males

$$Y = 199.29 + 0.136\, X_1 + 4.10\, X_2 + 43.3\, X_3 - 84.1\, X_4 - 8.4\, X_5 + 84.5\, X_6 + 62.6\, X_7 + 4.03\, X_8 - 10.6\, X_9 + 13.0\, X_{10} \quad (3)$$

Comparing the regression coefficients for males and females in (2) and (3) we notice the following:

Age and bodymass act in very similar ways for both male and female patients (again, age turns out to be an insignificant factor in these data sets). Family history turns out to be relatively more important for female patients than for male patients. Male children are important in a similar way for both male and female patients. However the presence of female children seems to have a stronger harmonious effect on female patients than on male patients. Surprisingly male patients are much more affected by divorce than the female patients and likewise patients are more affected by widow/widower state than the female patients. Likewise female patients are more affected by fat and protein than the male patients. A consolidated statement of the regression coefficients together with 't' values for testing the significance of the regression coefficients are given in Table 1.

TABLE 1. Regression coefficients t values for male and female and combined male and female patients.

Concomitant variable	Male		Female		Combined	
	Coeff.	t	Coeff.	t	Coeff.	t
Constant	199.29	6.19	238.87	7.23	237	11.05
X_1	0.136	– 0.27	0.175	0.25	0.001	0.00
X_2	4.10	4.90	2.72	4.29	2.87	5.76
X_3	43.3	– 3.69	75.7	6.43	65.4	7.87
X_4	– 84.1	– 6.23	– 84.4	– 5.77	– 89.5	– 8.87
X_5	– 8.4	– 0.65	– 29.8	– 2.02	– 23.1	– 2.35
X_6	84.5	5.62	45.2	3.42	59.2	5.89
X_7	62.6	4.08	41.7	2.85	50.6	4.69
X_8	4.03	0.43	3.4	0.34	2.98	0.43
X_9	– 10.6	– 1.09	– 2.25	– 0.23	– 10.2	– 1.47
X_{10}	13.0	1.39	24.6	2.57	19.5	2.86
R^2	74.7%		69.5%		69.9%	

Looking at the significant t values, the important common factors for both male and female patients are body mass, family history, presence of female children, being divorced and being widow/widower. Presence of female child is only important for female patients and high fat content in food also seems to affect only female patients. What is surprising is the value of R^2 which is very high in each case.

Around 70% of the variation is explained in female patients and the combined group by the set of 10 factors (for male patients it is 75%).

Stepwise Regression

We now raise the question of whether it is desirable to include all the 10 factors that we have listed or whether we can effectively reduce this set by eliminating some factors as relatively unimportant. A standard procedure of achieving this is to use the stepwise regression technique[2]. Essentially the procedure starts with inclusion of one factor. Then at each stage a new factor is introduced if it has a significant additional effect. Then a check is made to see whether any of the factors already included have become insignificant and therefore need to be dropped. The process stops when there is no real need to introduce a new factor or a real need to drop an old factor. On rare occasions however it so happens that the final set selected by this method is not necessarily the best set of that order. For example if eventually a set of 5 variables are selected by this method it may be possible to select a different set of 5 variables which explains the variation in the dependent variable in a better way. However such cases are rare and the final set selected by this procedure is usually quite good. We now present the results for the stepwise regression separately for male and female patients and then a combined result. The sequence in which the concomitant variables are introduced into the stepwise regression together with the value of R^2 (as a percentage) at each stage are given in Tables 2, 3 and 4:

TABLE 2.

Male

Step	Model including	R^2
1	X_4	43.76
2	X_4, X_2	59.98
3	X_4, X_2, X_3	66.58
4	X_4, X_2, X_3, X_6	71.06
5	X_4, X_2, X_3, X_6, X_7	74.07

TABLE 3.

Male

Step	Model including	R^2
1	X_4	39.73
2	X_4, X_3	55.70
3	X_4, X_3, X_2	65.66
4	X_4, X_3, X_2, X_{10}	61.66
5	$X_4, X_3, X_2, X_{10}, X_7$	67.08
6	$X_4, X_3, X_2, X_{10}, X_7, X_6$	68.56
7	$X_4, X_3, X_2, X_{10}, X_7, X_6, X_5$	69.42

TABLE 4.

Combined

Step	Model including	R^2
1	X_4	41.71
2	X_4, X_3	55.20
3	X_4, X_2, X_3	63.65
4	X_4, X_2, X_3, X_6	66.16
5	X_4, X_2, X_3, X_6, X_7	68.23
6	$X_4, X_2, X_3, X_6, X_7, X_{10}$	69.10
7	$X_4, X_2, X_3, X_6, X_7, X_{10}, X_5$	69.64

The regression coefficients for the final models, together with their t values, are given in Table 5. The first thing that strikes us is that two of the factors, namely (X_5) presence female child and (X_{10}) high level of fat, are included for the female patients but not for the male patients. Perhaps these factors are not very important in influencing the blood sugar level for male patients. Family history seems to affect the female patients much more than the male patients. However (X_6) Divorce and (X_7) widow/widower seems to affect the male patients more than female patients. Understandably these states cause more stress to male patients than females. With only 5 variables in the case of male patients we reach a value of R2 (74.07%) which is very close to the value of R2 based on 10 variables (74.7%). Similar conclusions hold for female patients and the combined group, each of which retain 7 variables. [The reduction in the number of concomitant variables is substantial in the case of male patients].

TABLE 5. Regression coefficients in the models selected by the stepwise regression for male, female and combined male and female patients.

Concomitant variable	Male		Female		Combined	
	Coeff.	t	Coeff.	t	Coeff.	t
X_2	4.17	5.08	2.79	4.44	2.85	5.74
X_3	– 43	– 4.16	– 74	– 7.01	– 64.7	– 8.56
X_4	– 78	– 7.54	– 84	– 5.81	– 88.3	– 8.8
X_5			– 29	– 2.04	– 22.2	– 2.27
X_6	88	5.97	45	3.45	59.4	5.94
X_7	61	4.00	42	2.88	– 22.2	– 2.27
X_{10}			24.5	2.59	18.9	2.78
R^2	74.07		69.42		69.64	

These three models indicate that some of the 10 variables are not worthy of inclusion. However, it would be interesting to explore these data sets further, using Principal Component Analysis.

Acknowledgement

I am grateful to Dr. A. Meccawi, Department of Surgery, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia for providing related references from Medical Journals and Professor M. Samiuddin, Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia for his comments.

References

[1] **Assas, B.M.,** Food Habits of Diabetes, *J. Fac. Commer., Ain Shams University,* **2:** 43-5 (1996).

[2] **Draper, N.R.** and **Smith, H.,** *Applied Regression Analysis.* John Wiley, New York (1981).

[3] **Assas, B.M., Meccawi, A.** and **Samiuddin, M.,** On Some Factors Affecting Blood Sugar Level, *J. Fac. Commer., Ain Shams University, Faculty of Commerce,* **1:** 1-9 (1997).

[4] **Miyakawa, M.,** Risk Factors for Diabetes Mellitus Evaluated by Long-Term Observation, Nippon Eiseigaku Zassbi, *Japanese Journal of Hygiene,* December, **50**(5): 986-97 (1995).

[5] **Assas, B.M., Samiuddin, M., Samra, A.A.** and **Meccawi, A.,** Hereditary Influences on Diabetics, *Pak. J. Statistics,* **12**(3): 259-264 (1996).

[6] **Steptoe, A.** and **Wardle, J.,** What the Experts Think: A European Survey of Expert Opinion About the Influence of Life Style on Health, *European Journal of Epidemiology,* April, **10** (2): 195-203 (1994).

التأثير المشترك لبعض العناصر على مستوى السكر في الدم

بكري معتوق عساس
قسم الإحصاء – كلية العلوم – جامعة الملك عبدالعزيز
جـــــدة – المملكة العربية السعودية

المستخلص . لقد ساهمت تغيرات نمط الحياة ونظام الغذاء والتحديثات الجارية في المملكة العربية السعودية خلال العقود الثلاثة الأخيرة في ارتفاع معدلات انتشار كثير من الأمراض الخطيرة والتي منها مرض ارتفاع مستوى السكر في الدم الذي يعتبر الآن مرضا شائعا في المملكة .

وقد تم جمع بعض البيانات عن مجموعة من المتغيرات الملائمة لدراسة مرض ارتفاع مستوى السكر في الدم من عينة من المصابين بهذا المرض والمسجلين في بعض المستشفيات بالمنطقة الغربية من المملكة ، ولقد أمكن التوصل إلى معلومات مفيدة عن المرض وعلاقته بخصائص المرضى .