A Statistical Study of Socio-economic and Physical Risk Factors of Myocardial Infarction

Alamgir
Department of Statistics, University of Peshawar,
Peshawar, NWFP, Pakistan

Salahuddin
Department of Statistics, University of Peshawar,
Peshawar, NWFP, Pakistan

Abstract

A sample of 506 patients from various hospitals in Peshawar was examined to determine significant socio-economic and physical risk factors of Myocardial Infarction (heart attack). The factors examined were smoking (S), hypertension (H), cholesterol (C), diabetes (D), family history (F), residence (R), own a house (OH), number of dependents (ND), household income (I), obesity and lack of exercise (E). The response variable MI was binary. Therefore, logistic regression was applied (using GLIM and SPSS packages) to analyze the data and to select a parsimonious model. Logistic regression models have been obtained indicating significant risk factors for both sexes, for males and for females separately. The best-selected model for both sexes is of factors S, F, D, H and C. The best-selected model for males is of factors CIFH, S, H, D, C and F, while the best-selected model for females is of factors D, H, C and F.

Key words: Logistic regression, Stepwise Procedures, Brown Method

Introduction

More people are at risk now a day to Myocardial Infarction (MI) than previously thought, particularly in developing world. More than 50 percent of deaths and disability from heart disease and strokes, which together kill more than 12 million people worldwide each year. Major risk factors for MI are high blood pressure, high cholesterol, obesity and smoking, World Health Organization (WHO) says. WHO estimates that 25 percent more healthy life years will be lost to cardiovascular disease globally by 2020? The brunt of this increase will be borne by developing countries.

Verna Rose (1999) showed that women had a higher short-term mortality rate after myocardial infarction than men, but only when the myocardial infarction occurs before the age of 75 years. This was his conclusion based on analysis of data from 384,878 patients enrolled in the National Registry of Myocardial Infarction from June 1994 through January 1998. Both before and after adjustment for medical history, admission data, type of treatment, time to hospital arrival and hospital characteristics, the risk of hospital mortality was higher in women than in men up to the age of 75 years. He showed that the risk of death in women increased 7 percent for every five years of younger age. The investigators believe that younger women who have had a myocardial infarction are in a high-risk group deserving special study.

Our country is also a developing one. To see the effects of these risk factors on MI in N.W. F. P., we have carried out this study based on data obtained from various hospitals in Peshawar.

Methods and Materials

A sample of 506 patients collected from different hospitals in Peshawar was examined for MI to investigate the risk factors of myocardial infarction (a binary response variable) like smoking, hypertension, cholesterol, diabetes, family history, residence, own a house, number of dependents, household income and lack of exercise etc. The same person collected the data on 506 patients in order to avoid bias. The appropriate technique, used for model selection, in the case of binary response variable is logistic regression. Here we consider general linear models in which the outcome variables are measured on a binary scale. The logistic regression model was first suggested by Berkson (1944), who showed how the model could be fitted using iterative weighted least squares. Logistic regression is now widely used in medical research since many studies involve two-category response variables.

The logistic regression model can be written as

$$p = p(x) = \frac{e^{(\beta_0 + \sum \beta_i x_i)}}{1 + e^{(\beta_0 + \sum \beta_i x_i)}} \qquad , \; i = 1,2,...,m$$

for a single explanatory variable X, the above model takes the form

$$p(x) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

Like classical regression models, inferences can be drawn for the logistic regression. The main contribution in this case was that of Wald (1943) who provided general asymptotic results for maximum likelihood estimator. It follows that parameter estimators in the logistic regression models have large-sample normal distributions due to maximum likelihood estimation and its large sample properties. Thus, a large sample 100(1-α)% confidence interval for parameter of the logistic regression model has the form $\hat{\beta} \pm z_{\alpha/2}\sigma(\hat{\beta})$, where $\sigma(\hat{\beta})$ stands for the estimated asymptotic standard error.

For logistic regression model with “s” parameters, let $\psi = (\psi_1, \psi_2, ..., \psi_q)$ denote a subset of normal parameters. Suppose we want to test H_0: $\psi = 0$. Let M_1 denotes the fitted model, and let M_2 denote the simpler model with ψ=0. Large sample tests can use Wilk’s (1938) likelihood ratio approach, with test statistic based on twice the log of the ratio of maximized likelihood for M_1 and M_2. Let L_1 denote the maximized log likelihood for M_1 and let L_2 denote the maximized log likelihood for M_2. Under H_0, the statistic -2 ($L_2 - L_1$) has a large sample chi-squared distribution with degree of freedom equal to q. Alternatively, by the large sample property of maximum likelihood estimators of parameter, the statistic

$$\psi' [\mathrm{Cov}\,(\psi)]^{-1}\,\psi$$

has the same limiting null distribution in large samples (Wald,1943). This is called a Wald statistic. When ψ has a single element, this chi-squared statistic with one degree of freedom is the square of the ratio of the parameter estimate to its estimated standard error, that is $\left(\frac{estimate}{s.e(estimate)}\right)^2$. The parameters of the model are estimated using the method of maximum likelihood. Once $\hat{\beta}$ has been obtained, the estimated value of the linear systematic component, also known as linear predictor, of the model is

$$\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots\dots\dots\dots + \hat{\beta}_m X_{mi}$$

from this, the fitted probabilities $\hat{p}_i$ can be found using $\hat{p}_i = \frac{e^{\hat{\eta}_i}}{1+e^{\hat{\eta}_i}}$.

Stepwise procedures were applied, using SPSS and GLIM packages, to select a final model. Stepwise procedures are sequential in the sense that they assume a current model and look to add to, or delete terms one at a time from that model, see Benedetti and Brown (1978), Goodman (1971).

A variety of methods for arriving at an appropriate initial model are available from which to begin the search for a well-fitting model. One of the suggested methods is that given by Brown (1976). It is most famous and simple technique for obtaining an initial model. Brown suggested two tests to look the importance of each possible term. These are tests of marginal association and a test of partial association. We applied Brown method to select an initial model for the backward elimination procedure. The variable used to determine an initial model are S, I, F, D, H, C, R and ND. Different factors for initial model, which were either partially or marginally significant at different stages (at 5% level of significance), were obtained. The significant factors both at marginal and partial tests are S, F, D, I, H and C. As far as the interactions of explanatory variables are concerned, the only four factors interactions, which were significant partially or marginally are IFHC and IFDC respectively. Partial test provided the model with factors S, F, H, D, C, I and IFHC. Marginal test provided the model with factors S, H, F, D, C, I and IFDC. For backward elimination procedure we got the model with factors S, H, F, D, C, I, IFHC and IFDC.

Backward elimination procedure

The best model is obtained in four steps automatically using SPSS package. The model selected through SPSS is with factor S, H, C, F and D. This model contains the main effects only. Thus the significant risk factors for MI are smoking, hypertension, cholesterol level, family history of patients and diabetes. Other factors turned out to be insignificant.

Table 1: Model Summary

Model	-2Loglikelihood	Deviance (G^2)	df	p- value	R^2
S, F, D, H, C	419.577	72.75	53	0.0372	0.7461

Table 2: Model with main effects

Model Terms	$\hat{\beta}$	Standard error	Wald	df	p-value
Smoking	1.194	.290	16.926	1	.000
Hypertension	-2.143	.305	49.257	1	.000
Cholesterol	4.014	1.027	15.277	1	.000
Family History	-1.315	.321	16.803	1	.000
Diabetes	-2.223	.408	29.672	1	.000
Constant	-3.837	1.048	13.413	1	.000

We see that all the coefficients are highly significant. Our final model is

$$\text{Logit}(\hat{p}) = -3.837 - 2.223\,D + 4.014\,C - 1.315\,F - 2.143H + 1.194\,S.$$

Model fitting for males and females

We have also investigated the effect of these factors on MI for males and females separately. Using SPSS package for backward elimination process, we obtained a model for males with factors CIFH, S, H, D, C and F. Again, the same factors turned out to be significant for MI but one interaction term CIFH appeared in the model. That is, these factors have a joint effect on the occurrence of MI. We fit this model to have parameters estimates along with standard errors given in Table 3.

Table 3: Model with interaction

Model Terms	$\hat{\beta}$	Standard error	Wald	df	p-value
Cholesterol*Income*Family History*Hypertension	1.829	.939	3.792	1	.052
Smoking	-.792	.316	6.282	1	.012
Hypertension	-2.605	.565	21.256	1	.000
Cholesterol	3.495	1.035	11.389	1	.001
Family History	-1.493	.433	11.904	1	.001
Diabetes	1.646	.546	9.090	1	.003
Constant	-4.160	1.148	13.126	1	.000

Thus our selected model is
Logit ($\hat{p}$) = -4.160 + 1.646 D + 3.495 C - 1.493 F - 2.605 H - 0.792 S+ 1.829C*I*F*H.

Models for females

We also consider selecting an appropriate model for females through backward elimination process, using SPSS package. The final model with factors H, C, D and F has been selected in five steps. The important risk factors for MI in this case are hypertension, cholesterol, diabetes and family history of patient. We noted that the smoking factor has been ignored which is obvious. Out of female patients, only one female was smoker. So smoking has no significant effect on MI. Table 4 contains information about variables in the models, estimated coefficients, their standard errors and significance of the coefficients.

Table 4: Model for females

Model Terms	$\hat{\beta}$	Standard error	Wald	df	Sig.
Hypertension	-2.483	.458	29.400	1	.000
Cholesterol	8.517	19.580	.189	1	.664
Family History	-1.508	.590	6.533	1	.011
Diabetes	.-2.985	.636	22.026	1	.000
Constant	-6.542	19.579	.112	1	.738

We now fit out final selected model using SPSS. The log likelihood of the model is 138.169. The most significant variables for female patients are hypertension, family history, cholesterol and diabetes. Thus our final model is

Logit ($\hat{p}$) = -6.542 - 2.985 D + 8.517 C - 1.508 F – 2.483 H.

Conclusion

The purpose of this study was to determine the significant risk factors of Myocardial infarction (MI). A total of 506 patients were examined and their personal and medical data were taken. For each patient, the phenomena of myocardial infarction was studied in relation to different risk factors, namely, smoking, diabetes, hypertension, cholesterol, income, family history of patients for MI, number of dependents, obesity etc. Other variables under consideration were sex, residence and own a house. Out of the total of 506 patients, 323 were male and 183 were female patients. Of 323 male patients, the number of MI patients were 212 and 111 patients had no MI. Of female patients, 88 had MI and 95 patients had no MI. Thus, we can conclude that males are more likely to have MI than females.

The response variable in this study was myocardial infarction, a binary variable taking the value 1 for the patient who had MI and 0 otherwise. As the response variable was binary, therefore, logistic regression was applied to the data for all the patients. Brown procedure was used to have an initial model. Backward elimination method resulted into a logistic regression model containing variables S, H, F, C, and D. This indicated that smoking, hypertension, family history of the patient, cholesterol and diabetes were important factors for MI. Residence (R) and number of dependents (ND) and income variables turned out to be insignificant for MI.

Separate logistic regression models were also fitted for each sex using the same backward elimination method. For males, we got the final model with variables S, F, D, H, C and ICFH. Thus, smoking, hypertension, family history of MI, diabetes, cholesterol and income were important risk factors for MI. The model also indicates that there is a joint effect of four factors, namely, income, cholesterol, family history of the patient and hypertension. For females, the model selected is with variables F, C, H and D. The model showed that family history of patient, cholesterol, hypertension and diabetes were important risk factors for MI.

References

1. Benedetti, J. K. and Brown M. B. 978. Strategies for the selection of log-linear models. Biometrics, 34: 680-686.
2. Berkson, J. 1944. Application of the logistic function to bioassay. Journal of American Statistical Association, 39: 357-65.
3. Brown, M. B. 1976. Screening effects in multidimensional contingency tables. Applied Statistics, 25: 37-46.
4. Goodman, I. A. 1971. The analysis of multinomial contingency tables: Stepwise Procedures and direct estimation methods for building models for multiple Classifications. Techno metrics, 13: 33- 61.
5. Verna R. 1999. Women under age 75 have a higher mortality rate after MI than Men. 48th Annual Scientific Session: Journal of the American College of Cardiology.
6. Wald, A. (1943). Test of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of American Mathematical Society, 54: 426-482.
7. Wilks, S.S. (1938). The large sample distribution of the likelihood ratio for testing Composite hypotheses. Annals of Mathematical Statistics, 9: 60-62.
8. WHO, 2002. Cardiovascular Death and Disability can be reduced more than 50 Percent. Press Release Series No.83.Geneva.