Analysis of factors Associated with Cholelithiasis in Multan District

Kinza Israr, Anam Javaid, Shahbaz Nawaz, Sumbal Javaid, Muhammad Kashan

In-Text Citation: (Israr, Javaid, Nawaz, Javaid, & Kashan, 2018)
To Cite this Article: Israr, K., Javaid, A., Nawaz, S., Javaid, S., & Kashan, M. (2018). Analysis of factors Associated with Cholelithiasis in Multan District. *International Journal of Academic Research in Business and Social Sciences*, *8*(9), 514–523.

Analysis of factors Associated with Cholelithiasis in Multan District

Kinza Israr[1], Anam Javaid[2], Shahbaz Nawaz[3], Sumbal Javaid[4], Muhammad Kashan[5]

[1]Department of Statistics (BS.student), The Women University Multan, Pakistan.
[2]Department of Statistics (Lecturer), The Women University Multan, Pakistan.
[3]Department of Statistics, (visiting Lecturer) Bahauddin Zakariya University Multan Pakistan.
[4]Medical Student in Nishter Medical University Multan, Pakistan.
[5]BS. Student in Department of Biotechnology, Bahauddin Zakariya Univrsity Multan Pakistan.
Email:kinzarajpoot15@gmail.com,anamjavaid0786@yahoo.com, Shahbazgurmani91@gmail.com, sumbal.javaid@yahoo.com, Kashanjaved01@gmail.com

Abstract

Gallstones basically considered as crystal like collections that are formed by merging of normal and abnormal gallbladder content. There can be two types of gallstones exist *i.e.* cholesterol stones & pigment stones. The current paper focuses on symptoms of the disease, major cause for the disease and treatments preferred by the majority patients and factor analysis is performed for this purpose by taking sample of size 170 data from different hospitals in Multan by using convenience sampling. Correlation matrix is used to see the correlation between different factors. Eigen values are constructed for each factor and scree plot is used for their graphical representation. Component loadings and rotated component analysis is performed after that a component plot and factor loading plot is obtained for a quick analysis.

Keywords: Biostatistics, Factor analysis, Gallstone, Multivariate analysis, Treatment, Patients.

Introduction

Gallstone is a common problem now a days and many persons especially females are effecting from this disease. many factors are related for the cause of this disease like the age of forty or elder, being overweight, pregnancy, eating high fat diet, taking high cholesterol diet etc. so the main problem of this study is to find out the factors causing this disease so that to some extent the people do not get effect from this disease. Many studies has been done in literature for the identification of factors related to GSD as Sharif et al., (2010) studied the primary liver cancer risk that were associated with gallstone and with Cholecystectomy by taking 4,487,662 subjects and by applying multivariate meta-regression analysis they found that gallstone history and cholecystectomy can increased liver cancer

515

risk. Wang et al., (2010) investigated the association of binding protein 4 (RBP4) with a variety of obesity- related diseases by taking 100 Cholesterol gallstone patients and 147 healthy controls. They applied multivariate analysis, odds ratio, confidence intervals and found significantly as lower serum RBP4 level as significantly lower in patients of gallstone as comparison with controls. Tsai et al., (2012)studied relationship between gallbladder status and recurrent biliary complication for the patients by taking 1625 patients of CBD stones by making its different groups by applying chi square test, ANOVA, Kaplan –Meier Analysis and log- rank test. Fracanzani et al.,(2012) studied the presence gallstone disease as associated with diabetes and predicted more severe liver disease by taking 373 males between January 2003 to June 2010and found 20% patients as having GSD by applying multivariate analysis they found the Patients with NAFLD had high prevalence of GD that characterized subject with glucose altered regulation and with more advanced liver disease. Dai et al.,(2013) studied the association between gallbladder stones and between the chronic hepatitis C by taking 1701 residents in Southern Taiwan and by applying multivariate analysis technique they identified different independent factors associated with GB stones in males and females. Huang et al.(2013) studied the status of health and explored the risk factors by taking Chinese rural population during 2007 by taking interviews. They found different morbidity rate for different diseases among which the gallstone were 0.54% and found the unhealthy lifestyle like usage of tobacco and overweight as the partially contribution of highly prevalence of Chronic disease. Jun et al. (2015) studied gallstone incidence after risk factors for formation of gallstone, gallstone after gastrectomy and cholecystectomy surgical outcomes after gastrectomy by taking total 2480 gastric cancer patients and found the 5.2% patients had the occurrence of gallstone formation who had undergone gastrectomy for cancer of gastric. Kim et al.,(2017) studied the associated factors with severity of acute pancreatitis (AP) according to alcohol and gallstones by taking 905 patients including 660 for alcohol and 245 for gallstone. They applied multivariate analysis and found diffrerengt factors like Current smoking, pancreatic necrosis, and bacteremia can heighten the clinical sequence of AP. Zhang et al., (2017) identified the risk factors that were associated with relevant to persistent upper abdominal pain(UAP) after cholecystectomy by taking 1714 symptomatic patients and applied multivariate analysis they found Females with longer ancient and more frequent preoperative Upper abdominal pain were more likely to grow postcholecystectomy UAP. Fagagnini et al., (2017) investigated risk factors for gallstones and for kidney stones by applying univariate and multivariate logistic regression analysis of taking 2323 patients and found the different significant factors for inflammatory bowel diseases (IBD).

From the above literature, previous work that has been done can be seen by some researchers that many researchers worked out for risk factor identification related to GSD. One of the contributions in literature can be add from this study as for the risk factor identification, a case study has been conducted related to GSD and from the analysis, and the main factors that are causing this disease can be seen.

Methodology

In literature, many researchers used multivariate analysis for study of GSD because there are many factors involved for study the factors of any disease. In this study also multivariate analysis has been used to identification of the risk factors by having such a questions in survey that how and why people

get effected from this disease and what kind of treatment they are taking also to which area they are belonging and from how many time they are effecting from this disease so for the answers of the above types of questions, data was taken from different private and govt. hospital from Multan by using convenience sampling. There were 170 observations of Patients in sample study. For the analysis, Multivariate analysis is carried out and to see the main risk factors of the disease, factor analysis is carried out and correlation matrix is obtained to see the relationship between any two factors. The first important identified factor will be the main risk factor this disease.

Results and Discussion

Factor Analysis

Factor analysis is considered as a data reduction method. For factor analysis, it is compulsory to see the correlation between different factors because the purpose of factor analysis is to find out the independent latent variables as it is a method for finding the variability between correlated variables in terms of low number of factors because now the identified factors will be able to provide almost the same information and the main factor will be the one that explain more variation. The purpose of factor analysis is to summarize the information that is in original variables from a smaller set of new factors having minimum loss of information so to identify in this case study; it is performed on SPSS by taking all factors included in the questionnaire. The following table show the correlation between different factors.

Table 1.   Frequency Distribution of Correlation Matrix of Main Factors of GSD

Correlations	Age in Years	Gender	Pregnancy	Complexion	Height in kg	weight after treatment	Causes of GSD	Genes Problem	Any Other Disease	Smoking
Age in Years	1.000	-.043	.027	-.143	.044	.138	-.139	.058	-.250	-.177
Gender	-.043	1.000	-.068	.086	-.111	-.251	.019	-.076	.258	.601
Pregnancy	.027	-.068	1.000	-.014	-.065	-.184	-.128	-.045	-.051	-.055
Complexion	-.143	.086	-.014	1.000	-.266	-.125	-.083	-.031	-.175	.141
Height in kg	.044	-.111	-.065	-.266	1.000	.243	.045	.121	.083	-.073
weight after treatment	.138	-.251	-.184	-.125	.243	1.000	.123	-.060	.032	-.081
Causes of GSD	-.139	.019	-.128	-.083	.045	.123	1.000	.208	.103	.093
Genes Problem	.058	-.076	-.045	-.031	.121	-.060	.208	1.000	-.062	-.125
Any Other Disease	-.250	.258	-.051	-.175	.083	.032	.103	-.062	1.000	.270
Smoking	-.177	.601	-.055	.141	-.073	-.081	.093	-.125	.270	1.000

The above matrix shows the correlation among the original variables. If the correlation between any two variables is too high (> 0.9) then one of that variable can be removed. The sign shows the positive

and negative direction of the relationship. It can be seen from the above result that no any two factors are highly correlated so there is no need to remove any variable.

B. Communalities

The communalities table are given below:

Table 2.   Communalities

	Initial	Extraction
Age in Years	1.000	.830
Gender	1.000	.830
Pregnancy	1.000	.642
Complexion	1.000	.670
Height in kg	1.000	.490
weight after treatment	1.000	.693
Causes of GSD	1.000	.597
Genes Problem	1.000	.788
Any Other Disease	1.000	.657
Smoking	1.000	.730

Initial communalities are estimates of the variance in each variable accounted for by all components or factors. For principal components extraction, this is always equal to 1.0 for correlation analysis. Extractions communalities are estimates of the variance in each variable accounted for by the components. Variables with high values are well represented in common factor space while variables with lower values are not. They are the reproduced variance for the number of components that we have saved. The communalities in this table are all high except height, which indicates that the extracted components represent the variables well.

Total Variance Explained

Total explained variation from different factors can be seen from the following table.

Table 3.   Total variance explained

Component	Initial Eigen values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.006	20.056	20.056	2.006	20.056	20.056	1.777	17.771	17.771
2	1.596	15.960	36.016	1.596	15.960	36.016	1.437	14.367	32.137
3	1.188	11.878	47.895	1.188	11.878	47.895	1.258	12.576	44.713
4	1.092	10.923	58.818	1.092	10.923	58.818	1.249	12.491	57.204
5	1.046	10.456	70.273	1.046	10.456	70.273	1.207	12.070	70.273
6	.818	8.177	77.450						
7	.779	7.791	85.241						
8	.645	6.448	91.689						
9	.490	4.903	96.592						
10	.341	3.408	100.000						

The Total column gives the Eigen value, or amount of variance in the original variables accounted for by each component. The % of Variance column gives the ratio, expressed as a percentage, of the variance accounted for by each component to the total variance in all of the variables. The Cumulative % column gives the percentage of variance accounted for by the first n components. Here in this case, it is 70.273% almost 71% means that if we used just these five extracted components instead of original ten components then 71% explained variation can be obtained from these factors by leaving out the other five factors. For the initial solution, there are as many components as variables, and in a correlations analysis, the sum of the Eigen values equals the number of components. Eigen values considered as variances of factors. Eigen values greater than 1 has been extracted, because the factor with Eigen value greater than or equal to one is able to explain more variation than a single observed variable so the Eigen value of 2.006 mean that the factor will explain variance of 2.006 as of ten variables that explain more variation among ten variables so the first five principal components form the extracted solution so instead of using the original ten variables, now these five factors can be used for the study as there will be minimum loss of information by using the data reduction technique. Rotation sum of square loadings shows the extracted components. They explain nearly 71 % of the variability in the original ten variables, so it can considerably reduce the complexity of the data set by using these components, with almost of 29% loss of information. The rotation maintains the cumulative percentage of variation explained by the extracted components, but that variation is now spread more evenly over the components. The large changes in the individual totals suggest that the rotated component matrix will be easier to interpret than the unrotated matrix but there is a little bit change in individual component of both types so both type of matrix will give almost same explanation.

Scree Plot
For the graphical representation, the scree plot is obtained by using Eigen values of each of the component.

Figure 1.   Scree Plot

The Eigen value of each component in the initial solution is plotted on Scree Plot that helps us to determine the optimal number of components. Here it can be shown that on the steep slope, 5 components can be extracted while other 5 on the shallow slope will contribute little to the solution and the important five components that are extracted can be viewed as from the scree plot.

Component Matrix

Table 4.   Component Matrix

Variables	Component				
	1	**2**	**3**	**4**	**5**
Age in Years	-.416	-.128	-.306	-.112	.732
Gender	.784	.081	-.132	.019	.438
Pregnancy	-.042	-.401	-.228	.637	-.146
Complexion	.292	-.495	.385	-.426	-.101
Height in kg	-.316	.565	-.207	.152	.068
weight after treatment	-.386	.486	-.181	-.514	-.109
Causes of GSD	.067	.516	.569	.040	-.024
Genes Problem	-.217	.187	.645	.366	.396
Any Other Disease	.450	.526	-.229	.237	-.261
Smoking	.793	.184	-.096	-.115	.213

This table contains component loadings, which are the correlations between the variable and the component (that has been extracted). Because these are correlations, possible values range from -1 to +1. It can be seen from the above result for each variable.

Rotated Component Matrix

The rotated component matrix for the main factors is given below:

Table 5.   Rotated component matrix

Factors of GSD	Component				
	1	2	3	4	5
Age in Years	-.005	.185	.890	.058	-.037
Gender	.903	-.066	.049	-.084	.009
Pregnancy	-.157	.107	-.025	-.761	-.162
Complexion	.048	-.809	-.065	.092	-.029
Height in kg	-.094	.657	.029	.194	.103
Weight after treatment	-.239	.309	.029	.717	-.162
Causes of GSD	.056	.079	-.335	.285	.628
Genes Problem	-.088	.066	.161	-.112	.859
Any Other Disease	.412	.432	-.542	-.004	-.080
Smoking	.829	-.088	-.159	.082	-.050

Here the rotated component matrix helps us to determine what the components represent. The first component is most highly correlated with gender and smoking and they are less correlated with the other two components. The second component is most highly correlated with height. The third component is most highly correlated with age. Fourth is with weight and fifth is with genes problem. So by using the above analysis, we can focus on gender, smoking, height, age, weight and genes problem and also with causes of GSD because the components are representative of all ten original variables, and the components are not linearly correlated with each other.

Component Plot in Rotated Space
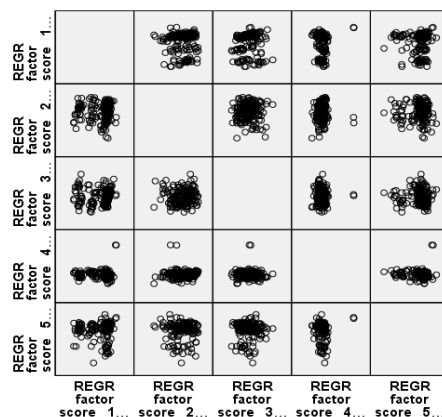The component plot in rotated space is the main factors are given below:
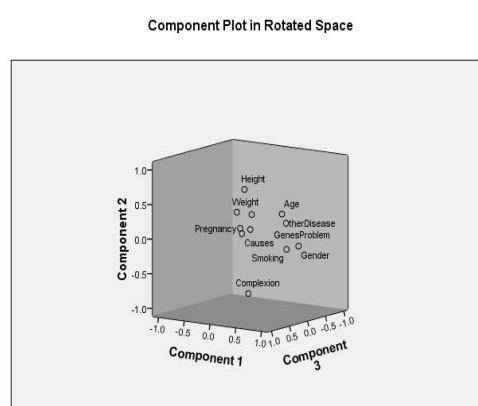
Figure 2. Component plot in rotated space

Values close to 1 or -1 shows that the variables have strong effect while values less than 0.5 show there variable do not have strong effect.

H.   Factor Loading Plot
The factor loading plot is the main factors are given below:

Figure 3.   Factor loading plot

We can check the component scores for outliers and non linear association between components. The above plot shows this about five main factors in our resulting analysis

Conclusion

From the above analysis, correlation matrix indicates the correlation between different factors in this study related to Gallstone disease, total explained variation from these factors can also be seen. Scree plot can also help us in extraction of important components. Factor loading plot help us in component scores for outliers and non-linear association between components.

References

Sharif, A. W., Williams, H. R. T., Lampejo, T., Khan, S. A., Bansi, D. S., Westaby, D., Thillainayagam, Andrew V.., Thomas, Howard C., Cox, I. Jane a &Taylor-Robinson, S. D. (2010). Metabolic profiling of bile in cholangiocarcinoma using in vitro magnetic resonance spectroscopy. *Hpb*, *12*(6), 396–402. https://doi.org/10.1111/j.1477-2574.2010.00185.x

Wang, S. N., Yeh, Y. T., Wang, S. T., Chen, Y. L., Chuang, S. C., Ker, C. G., & Lee, K. T. (2010). Decreased Retinol Binding Protein 4 Concentrations are Associated With Cholesterol Gallstone Disease. *Journal of the Formosan Medical Association*, *109*(6), 422–429. https://doi.org/10.1016/S0929-6646(10)60073-3

Fracanzani, A. L., Valenti, L., Russello, M., Miele, L., Bertelli, C., Bellia, A., Masetti, Chiara., Cefalo, Consuelo., Grieco, Antonio., Marchesini, Giulio & Fargion, S. (2012). Gallstone disease is associated with more severe liver damage in patients with non-alcoholic fatty liver disease. *PLoS ONE*, *7*(7). https://doi.org/10.1371/journal.pone.0041183

Tsai, T. J., Lai, K. H., Lin, C. K., Chan, H. H., Wang, E. M., Tsai, W. L., Cheng, Jin Shiung., Yu, Hsien Chung., Chen, Wen Chi and Hsu, P. I. (2012). The relationship between gallbladder status and recurrent biliary complications in patients with choledocholithiasis following endoscopic treatment. *Journal of the Chinese Medical Association*, *75*(11), 560–566. https://doi.org/10.1016/j.jcma.2012.08.003.

Huang, X., Chen, M., Tan, H., Xiao, S., & Deng, J. (2013). The morbidity rate of chronic disease among chinese rural residents: Results from Liuyang cohort. *Medical Principles and Practice*, *22*(4), 362–367. https://doi.org/10.1159/000346437

Dai, C. Y., Lin, C. I., Yeh, M. L., Hsieh, M. H., Huang, C. F., Hou, N. J., Hsieh, Ming Yen., Huang, Jee Fu., Lin, Zu Yau., Chen, Shinn Cherng., Wang, Liang Yen., Chang, Wen Yu., Chen, Jong Shyong., Yu, Ming Lung & Chuang, W. L. (2013). Association between gallbladder stones and chronic hepatitis C: Ultrasonographic survey in a hepatitis C and B hyperendemic township in Taiwan. *Kaohsiung Journal of Medical Sciences*, *29*(8), 430–435. https://doi.org/10.1016/j.kjms.2012.12.004

Jun, K.-H., Kim, J.-H., Kim, J.-J., Chin, H.-M., & Park, S.-M. (2015). Retrospective Analysis on the Gallstone Disease after Gastrectomy for Gastric Cancer. *Gastroenterology Research and Practice*, *2015*, 1–7. https://doi.org/10.1155/2015/827864

Kim, D. B., Chung, W. C., Lee, J. M., Lee, K.-M., Oh, J. H., & Jeon, E. J. (2017). Analysis of Factors Associated with the Severity of Acute Pancreatitis according to Etiology. *Gastroenterology Research and Practice*, *2017*, 1–5. https://doi.org/10.1155/2017/1219464

Fagagnini, S., Heinrich, H., Rossel, J.-B., Biedermann, L., Frei, P., Zeitz, J., Spalinger, Marianne., Battegay, Edouard., Zimmerli, Lukas., Vavricka, Stephan R., Rogler, Gerhard., Scharl, Michael &

Misselwitz, B. (2017). Risk factors for gallstones and kidney stones in a cohort of patients with inflammatory bowel diseases. *Plos One*, *12*(10), e0185193. https://doi.org/10.1371/journal.pone.0185193

Zhang, J., Lu, Q., Ren, Y. F., Dong, J., Mu, Y. P., Lv, Y., & Zhang, X. F. (2017). Factors relevant to persistent upper abdominal pain after cholecystectomy. *Hpb*, *19*(7), 629–637. https://doi.org/10.1016/j.hpb.2017.04.003