Pak. j. life soc. Sci. (2024), 22(2): 11199-11218 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.2.00848

**RESEARCH ARTICLE**

# Using Binary Logistic Regression to Detect Health Insurance Fraud

Baraah Samara *

Ph.D Stude , Ph.D Program of Strategic Management, Faculty of Graduate Studies, Arab American University (AAUP), Ramallah, Palestine

***ARTICLE INFO***

Received: Aug 24, 2024

Accepted: Nov 1, 2024

***Keywords***

Fraud Detection
Logistic Regression
Insurance
Prediction
Decision Support System

***Corresponding Author:**

Baraasamara26@gmail.com

**ABSTRACT**

This study aims to develop a predictive model that detects fraud in the health insurance of individuals treated by internists in their private clinics or hospitals to reduce the flow of health insurance losses with insurance companies. The methodology used is logistic regression, where 26 independent variables are identified to predict fraudulent members. All available datasets from the insurance companies database from Jan to Nov 2022 were used to test our proposed model, including 161 inter$t providers, 11,385 beneficiaries, and 122,928 transactions. Results: The variables can be used effectively to identify fraudulent members. 99.60% of predictions are correct. Despite doing an awful job of predicting more accurately than a random prediction, the model has high accuracy in predicting legitimate claims (99.9%) and fraudulent claims (79.3%). It is important for the insurance industry to develop and promote decision support systems, especially when detecting fraud and assessing risks. To provide better medical services to actual patients, save money, and improve the healthcare experience for patients with real needs, future work requires the development of other fraud models related to other variables and service providers, which can reap many benefits.

## INTRODUCTION

Over the years, healthcare has evolved into one of the most profitable industries [1].It is one of the most dynamic industries with the most active marketplace characterized by the need to facilitate the delivery and sharing of information regarding the healthcare resources, transactions, and other components needed by the players in this sector to operate efficiently as posited by [1].One of the most significant aspects that has changed with time in healthcare is the cost of healthcare. According to statistics by the Centers for Medicare and Medicaid Services (CMS) and the National Center for Health Statistics in the U.S., between 1965 and 2008, there was a significant increase in the amount of money spent on healthcare services. Approximately $42 billion was spent in 1965 on healthcare services by U.S. consumers [1].This increased by over 1,657 percent to about $738 billion by 1991 and over one trillion dollars by 1994. Ten years later, in 2004, it rose to $1.6 trillion. It exceeded $2 trillion in 2008, with every consumer spending approximately $6,280 per year, which was more than a sixth of the entire U.S GDP [1]Many factors have occasioned this increase, but one crucial factor is healthcare service providers' fraud and abuse of insurance payments.

An effective fraud detection mechanism is not only crucial for insurance companies but is also helpful for patients and the economy at large. While there is no sign of this stopping soon, decision support system techniques offer a glimmer of hope to patients, governments, and insurance service providers. These tools reduce the amount of the premium that consumers must pay. With the increase in healthcare spending in many countries, efficient handling of fraud is necessary to detect anomalies and abuse in this sector. The right strategies will satisfy people, and only genuine claims will be given.

Healthcare fraud schemes come in different varieties. However, authors have been put into three categories based on the person involved to ease understanding [2]. These categories include provider fraud, consumer fraud, and payer fraud. Provider fraud is the most common type in the healthcare industry [2]. A study by the authors estimated that more than 10 percent of expenditures in healthcare systems in the US is wasted on fraud annually and found in the healthcare system, fraud and abuse cost between $59.9 billion and $83.9 billion, making it a priority area that must be addressed with urgency. [3]. Surprisingly, approximately $25 million is lost to fraud yearly in this industry [4]. For this reason, the model sets out to detect fraudulent claims in Palestine, thereby saving insurance companies and the insurers themselves. The World Health Organization's goal of ensuring that insurance is available to everyone at all times, wherever they need it, is achieved by reducing the flow of insurance losses and by relying on the data of internists, who make up the largest proportion of insurance claims so that the benefits are spread out more widely.

## 1. LITERATURE REVIEW

The use of regression to detect anomalous patterns among healthcare practitioners has been of paramount importance in classifying fraudulent claims using various dimensions, including the probability of anomalies, number of claims, highest claims amount and general amounts of claims, and the number of health centers [5]. This is critically important as it allows for estimating the probability of anomalies in each record. When the probability exceeds 50%, it would categorize the record as anomalous. In a study conducted by the authors in Turkey, more than 6595 recorded claims had at least 50% probabilities [5]. It was also found that the high probabilities of claims were attributed to excessive charges on patients in public and private hospitals. Therefore, an important predictor of insurance claims in Turkey is the excessive charges the healthcare facilities impose on patients [5].
Liu & Vasarhelyi (2013) conducted an experiment using various datasets, including the insurance providers' institutional information, subscribers' information, physicians' information, payer information, reports of diagnosis, and claim information. The regression analysis was based on the payment amount stated in the claim and the distance between the client and the health service provider. About 74 claims were estimated from instances when the abnormal claims were distinct from the normal claims [6].

Considering the nature of the high abnormality of the insurance claims, Musal (2010) suggests that the fraud associated with insurance claims is now at a high level, forming a major problem in healthcare financing. This literature review seeks to analyze data from several insurance beneficiaries that use infusion therapy medications. This exercise is sought to be completed using regression methods that can detect fraudulent claims in medical insurance [7].

Private insurance companies and governmental health departments are the primary raw data sources for detecting insurance/claims fraud. In insurance claims, the service provider and the subscriber are active, but the healthcare provider is the sole determinant of the service cost. Doctors and physicians are particularly directly involved in fixing the claims costs. Raw data, including diagnoses, cost of service, Lab ID, and other relevant information, contain a vast amount of attributes that can describe the behavior of the service provider and subsequently point out any fraudulent claims the provider and the subscriber commit in the health care service [6].

## 2. Binary Logistic Regression Model

In the absence of a quantitative response model for binary variables, logistic regression is an appropriate technique for assessing the effects of categorical variables on the dependent variable. The method is well understood, easily applied, and provides a solid foundation for newer assessment approaches [8]. As a nonlinear approach, logistic regression can only be used to model dichotomous variables since the

classifying variable has to be either 0 or 1 [9] and out of 32 classification algorithms, logistic regression has been ranked as the second most accurate in terms of accuracy [10].
In insurance fraud investigations, binary choice models are used to predict an insurance claim's likelihood to be fraudulent. Several indicators are used by investigators to identify individuals who are more likely to submit fraudulent claims based on estimated probabilities [11]. As a matter of fact, the dependent variable used in this study is referred to as a "binary variable," and logistic regression is one of the most commonly used methods to analyze such problems. Studies of fraud have been conducted using binary choice models [8]. A relationship of this type can be described as having the shape of an "S." Several reasons have contributed to the popularity of the logistic regression model, including its logistic function on which it is based, which provides estimates between 0 and 1, as well as its S-shaped representation for the combined effect of several risk factors on an event's risk [12].

## 2.1. Statistical Equation

By using the natural logarithm to calculate the odds, logistic regression, a form of statistical analysis, gives a linear correlation between an event's natural log odds and its explanatory variable [13]. As a result of logistic regression, the natural log odds are modeled as linear functions of the explanatory variables, and the probability of an interesting outcome is predicted as follows [13]:

P=P (Y=interested outcome/X= $\chi$, value of independent variable)

$P = \chi 1/1 + e-(a+\beta\chi_1+ \beta\chi_2+\ldots\ldots)$

## 2.2. Description of the Data

Data-driven approaches based on logistic regression are being applied to insurance claim data to detect fraud. This specific method is intended to identify frequently occurring fraud patterns and detect fraud in claims data, particularly for internists. Health insurance claims are included for patients treated within the insurance medical network, including pathological treatments, laboratories, x-rays, medicines, and other conditions. The methodology will apply to all available datasets from Jan-Nov 2022, including 161 internist providers and 11,385 beneficiaries with 122,928 transactions. There are no missing values in this report's sample data, and all anonymous transactions and identification numbers (such as subscriber name, national identifier, policy number, and health care provider name) have been omitted and replaced with names and aliases. Our initial data set contained 26 attributes. There are binary, numerical, and categorical attributes. In a medical insurance database, attributes include demographics (age and sex), details of services (treatments), and details of policies and claims (benefits and amounts).

Logistic regression was used to calculate the likelihood of fraud using data from 11,385 subscribers and 122,928 claims grouped by aggregation procedures. To determine where participants consume health insurance, various random variables were correlated. A logistic regression approach helps obtain this distinction and produce a homogeneous distribution of healthcare services used and billed. Each subscriber is identified as potentially excluded based on their usage or billing. Here are the steps required to illustrate the first section of the model: 1. Aggregate the cost per member using the various variables in the raw data set and combine them into one item. 2. The dependent variables' regression analysis is conducted based on the continuous independent variable from the previous step.

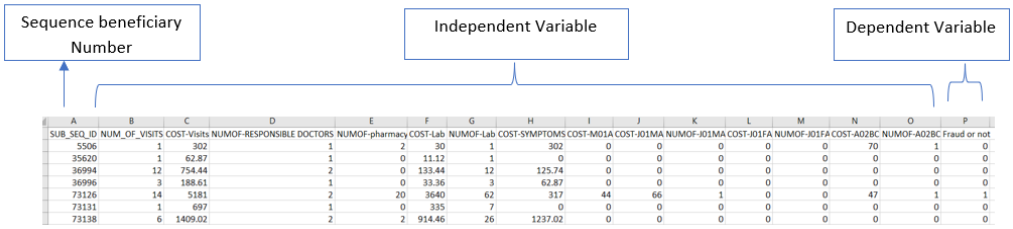
## 2.3. Potentially helpful variables

The following independent variables were selected after the aggregation procedure was used to get new features from the initial data set in light of the positive outlier results relating to cost and other variables presented in the previous exploratory report.
As a result, the new attribute includes the total number of claims, the number of visits, the cost of visits, the number of follow-up visits, the number of visits to a treated internist, the cost per doctor's office; the number of claims per doctor's office; the cost of pharmacy; the number of dispensed medicines; the cost of lab procedures; the cost and number of symptoms per chapter; and the cost and number of therapeutic class, Anti-Inflammatory & Anti-Rheumatic Products, non-Steroids, Fluoroquinolones, Macrolides, Proton Pump Inhibitors.

These variables were collected as independent variables in the first 11 months of 2022. The dependent variable was the patron's state, which was a 1 if the member was fraudulent and a 0 if not. Primary data were obtained to extract the fraud members, from which the participants were classified as either outliers or not based on the cost and then the classification of fraud.

## 2.4. An overview of the potentially helpful variables

In Figure 1, the dependent variable is binary, while the independent variables are continuous in nature.




| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SUB_SEQ_ID | NUM_OF_VISITS | COST-Visits | NUMOF-RESPONSIBLE DOCTORS | NUMOF-pharmacy | COST-Lab | NUMOF-Lab | COST-SYMPTOMS | COST-M01A | COST-J01MA | NUMOF-J01MA | COST-J01FA | NUMOF-J01FA | COST-A02BC | NUMOF-A02BC | Fraud or not |
| 5506 | 1 | 302 | 1 | 2 | 30 | 1 | 302 | 0 | 0 | 0 | 0 | 0 | 70 | 1 | 0 |
| 35620 | 1 | 62.87 | 1 | 0 | 11.12 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 36994 | 12 | 754.44 | 2 | 0 | 133.44 | 12 | 125.74 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 36996 | 3 | 188.61 | 1 | 0 | 33.36 | 3 | 62.87 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 73126 | 14 | 5181 | 2 | 20 | 3640 | 62 | 317 | 44 | 66 | 1 | 0 | 0 | 47 | 1 | 1 |
| 73131 | 1 | 697 | 1 | 0 | 335 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 73138 | 6 | 1409.02 | 2 | 2 | 914.46 | 26 | 1237.02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 1 : an overview of the potentially helpful variables**

## 2.5. Incorporating binary logistic regression into fraud detection

Fraud in health care is a serious issue. Predictive models, such as regression models, are being used in practice in an attempt to better detect fraud. Fraud prevention aims at reducing costs by maximizing the percentage of error-free identification of fraudulent service providers as soon as unusual patterns in the data suggest they are likely to be involved in fraud. This report is based on real-life data collected from insurance companies.

A fraud classification model is used to classify new transactions as either fraudulent or legal using samples of fraudulent and legal transactions. The fraud detection model identifies outliers as possible instances of fraudulent transactions to predict the likelihood that a particular member will be fraudulent.
A model was developed to predict fraudsters based on subscribers and medical providers within the client portfolio. By utilizing the model, the company can decide whether to approve treatments and medical expenses based on predictions of fraudsters from subscribers and medical authorities. In this case, the objective is to determine whether a subscriber has misused insurance or committed fraud in its various manifestations.

## 3. Hypothesis

The dependent variable is fraudulent members, who are counted as 1 if they are and 0 if they are not. One of the independent variables considered is the total cost of medical provider visits (COST visits). The assumption is that the higher the total cost of visits for a subscriber during the eleven months, the higher the odds. Therefore, we would expect that a positive relationship would exist if the insured were a fraudster. That was supported by Table 1 with results of Z-score in Appendix 2.

An insured's medical examination costs are another independent variable (COST-OFFICE-PROC). Insurers will be monitored for fraud using cost records from doctors' clinics since the more visits to doctors provided by the insured, the more likely the insured will use them for opportu$tic reasons, which increases the likelihood of fraud. Therefore, we expect a direct correlation between doctors' office costs and insured fraud. As in Figure 3 in Appendix 2, the cost of the office procedure ranked number 3 relative to HCP Cost.
A third independent variable is the cost of drugs dispensed by the insured (COST-pharmacy); as in Figure 3 in Appendix 2, it ranks second in internal doctor expenses. It is expected that the greater the number of medicines dispensed by the insured, the greater the chance of fraud by subscribers, so the cost of dispensing medicines from pharmacies will be used as an indicator of fraud. since drugs prescribed for non-chronic treatments are being replaced by cosmetics or resold. Pharmacy costs and insured fraud must be directly linked.

The fourth independent variable is the cost of the lab (COST-Lab). Among internal medicine expenses, as shown in Figure 3 in Appendix 2, laboratories account for 40%. It is important to consider the cost associated with the laboratory tests doctors request for patients, including 19 procedures at a total cost of 606,111 shekels. The model must consider outliers based on the Z-Score analysis in Table 2 in Appendix 3. The likelihood of a fraudster is higher if more laboratory tests are performed and the cost increases, so a positive relationship is expected.

The fifth independent variable is the cost of each chapter symptom (COST-SYMPTOMS). A diagnosis entered by an internist is classified by the International Classification System into 19 chapters based on where the disease is located. Compared to the other chapters, the SYMPTOMS, SIGNS, AND ILL-DEFINED CONDITIONS chapter consumed the most s(1,956,593). Thus, this chapter describes cases involving general pain, abdominal pain, chest pain, and other symptoms caused by vitamin deficiency. Insurance costs increase, and fraud increases. This hypothesis is supported by the Z-score result in Table 3 in Appendix 3.

According to z-score analysis, the $6^{th}$, $7^{th,}$ and $8^{th}$ independent variables are for the following pharmacotherapeutic categories, showing extreme disbursement among internists. The misuse and excessive use of antibiotics (COST-J01MA, COST-J01FA) is common, especially since some antibiotics are extremely expensive, making fraud easier. Proton pump inhibitors (COST-A02BC) are also prescribed with other medications because other medications are heavy on the stomach, so fraud is easy to commit with them. Furthermore, they are prescribed in conjunction with other medications because they have a tendency to cause excesses. The likelihood of a fraudster is higher if more of these therapeutic classes are consumed and the cost increases, so a positive relationship is expected. This hypothesis is supported by the Z-score result in Table 4 in Appendix 3.

## 4. Development of the Model

In the preliminary analysis and first step, it was found that the coefficient values in Table 1 were not equal to zero, which means that the zero hypotheses should be rejected and the alternative hypothesis accepted table 1: coefficient values.

### 4.1. An analysis of multicollinearity in regression

Regression models with multicollinearity are characterized by highly correlated independent variables. Model results can be difficult to interpret when independent variables are highly correlated because if one is changed, the other will be affected, resulting in significant variations. If you were required to select the list of significant variables for the model each time, the results would likely differ significantly with slight differences in the data or the model. Coefficient estimates would not be stable, and the model would be difficult to interpret, making it prone to overfitting. Multicollinearity was avoided by considering whether the independent variables were correlated and whether the correlation was greater than 0.8. figure 4 in Appendix 2 shows Pearson correlations for twenty- two potential independent variables and one dependent variable.

The correlation matrix shows that there are seven pairs of independent variables with a correlation equal to or greater than 0.8, which are the number of visits and the number of claims), (cost of pharmacy and number of pharmacies), (cost of lab and number of lab), (cost of symptom and number of symptom), (Cost of -J01MA and number of -J01MA), (Cost of -J01FA and several -J01FA) and (Cost of -A02BC and number of -A02BC).In the matrix, the variables that were most closely related, least important, and not sufficiently correlated with fraud were eliminated. The cost per visit, the pharmacy cost, the cost of laboratory, the symptom cost, the office procedure cost, and the J01MA, J01FA, and A02BC costs were retained.

### 4.2. A Correlations between fraud and each independent variable

(P value = 0.000).

**Table 2: A Correlations between fraud and each independent variable**

| CORREL | COST-Visits | COST-OFFICE-PROC | NUMOF-OFFICE-PROC | COST-Lab | COST-SYMPTOMS | COST-J01FA | COST-A02BC |
|---|---|---|---|---|---|---|---|
| Fraud | 0.64 | 0.53 | 0.35 | 0.24 | 0.35 | 0.06 | 0.21 |

A logistic regression model incorporating all eight independent variables is shown in Table 2. These results indicate that all eight variables are significantly related to fraud.

### 4.3. log-likelihood

Using this measurement, one can assess whether or not the model has improved compared to what it would have been without the addition of the independent variable. As a result, the probability value should be considered when adding an independent variable to an equation.

Model Summary for Binary Logistic Regression

**Table 3: Model Summary for binary logistic regression (Source: own elaboration using real statistics add-in Excel)**

| | LL | 2 log-likelihood |
|---|---|---|
| LL0 | (1,035.9) | 2071.74 |
| Step 1 | (4,905.0) | 9810.06 |
| Step 2 | (176.5) | 352.96 |
| Step 3 | (177.8) | 355.58 |
| Step 4 | (177.8) | 355.58 |

The likelihood value of a model can be calculated by comparing the fits of two models with and without independent variables to determine if they are strongly correlated; a decrease in this probability value indicates that adding this feature improved the model.

Table 3 shows the results of using Pearson product-moment correlation to examine the association between fraud and each independent variable. As shown in Appendix 1, eight related variables were selected for the detection fraud model. A regression model is analyzed backward and stepwise. Using the full model, including all p predictors, each predictor is removed one by one. The variable with the largest p-value is removed, as is the variable with the least statistical significance. The first full model is available in Table 8 in Appendix 3.

The following observations were drawn from Tables 3 and 4:

The initial value of the -2-log obtained for the regression model containing only constants was 2071.74. After analyzing the multicollinearity and ignoring the highly correlated variable, adding all the independent variables to the model gives the -2-log value of 9810.06, much higher than the initial value.

In the second step, one variable, the visit cost, which is the least important variable, was removed from the model, along with its relationship with the dependent variable, where p is greater than 0.05. Thus, the log probability of -2 decreased by 9457 and decreased by one degree of freedom.

The third step consisted of removing the independent variable COST-J01MA, which had a value of p >0.05. Thus, the log probability of -2 increased by 2.62 and decreased by one degree of freedom.

In the end, removing this variable is the best step to get the best model since all its independent variables have an effect on the dependent variables. The best model is derived based on this information, which is the final step. In logistic regression, a model with six independent variables exhibits an improved fit to the data compared to a model with no independent variables (the null model).

## 5. RESULTS AND PERFORMANCE METRICS

### 5.1. Variables equation in the Best Model

The output includes the coefficients, standard errors, Wald z-statistics, and p-values for the six independent variables. Wald's test determines which independent variables are statistically significant in predicting fraud among members. All six independent variables are statistically significant in predicting fraud among members. This test rejects the null hypothesis that the predictor has no effect on the dependent variable if the p-value is less than 0.05.

**Table 4: coefficients, standard errors, Wald z-statistics, and p-values for the independent variables**

| | **coeff b** | **s.e.** | **Wald** | **p-value** | **exp(b)** | **lower** | **upper** |
|---|---|---|---|---|---|---|---|
| Intercept | -11.7288 | 0.708117 | 274.3453 | 0.00 | 8.06E-06 | | |
| COST-OFFICE-PROC | 0.004879 | 0.000329 | 220.0764 | 0.00 | 1.004891 | 1.004243 | 1.005539 |
| COST-pharmacy | 0.005188 | 0.000417 | 154.5553 | 0.00 | 1.005201 | 1.00438 | 1.006024 |
| COST-Lab | 0.005131 | 0.000475 | 116.7175 | 0.00 | 1.005145 | 1.004209 | 1.006081 |
| COST-SYMPTOMS | 0.00061 | 0.000211 | 8.347953 | 0.00 | 1.00061 | 1.000196 | 1.001025 |
| COST-J01FA | 0.002845 | 0.000196 | 210.29 | 0.00 | 1.002849 | 1.002464 | 1.003235 |
| COST-A02BC | 0.004744 | 0.002076 | 5.221339 | 0.02 | 1.004755 | 1.000675 | 1.0088520 |

The following equation of the regression logistic model is constructed:

Logit (Fraud)= -11.7288+0.005188 (COST-OFFICE-PROC) + 0.005188 (COST-pharmacy) + 0.005131 (COST-Lab) +0.00061 (COST-SYMPTOMS) +0.002845 (COST-J01FA) + 0.004744 (COST-A02BC)

According to logistic regression, the log odds of the outcome are changed by a unit increase in the predictor variable. Fraud increases by 0.004879 for every one-unit increase in office procedure costs. The intercept test result was significant, suggesting that the intercept should be incorporated into the model. If the value of the office procedure increases by one unit, the likelihood that an event will result in the member becoming fraud increases by 1.005 times.

The likelihood of fraud increases by 0.005188 for every unit increase in pharmacy cost. Drug dispensing indicators are related to the likelihood of fraud. This factor also demonstrated significant predictive ability (p 0.05). The test of the intercept also revealed significant results (p 0.05), suggesting that the intercept would require inclusion in the model. When the cost of the pharmacy is increased by one unit, the odds of this occurring increase by 1.005 times.
The likelihood of being a fraudulent member increases by 0.002845 for J01FA and 0.004744 for A02BC by one unit. The explanation of variances in fraud detection is statistically significant when exp(B) is 1, which indicates that these types of medicines are 1 time more likely to contribute to fraud.

The chance of becoming a fraud member increases with a one-unit increase in the cost of the symptom. It was determined that the slope coefficient of 0.00061 reflected an increase of one unit in fraud and also indicated that there was significant evidence that the intercept should be included in the model. The intercept test was significant and indicated that the intercept should also be included in the model. When the value of the cost of a chapter symptom is increased by 1 unit, it increases the odds of an event by one time.

### 5.2. Chi-Square

Pearson Chi-Square was used to determine whether the independent variables were related to the dependent variable (Fraud). In this report, the chi-squared for the overall model fit statistic was computed, yielding a p-value of 0, as in Table 5, which was less than the conventional 0.05. This indicates that at least one independent variable contributes to the outcome prediction, and it is, therefore, necessary to reject H0.

### 5.3. A correlation coefficient of R-squared

Based on R squared, the independent variables had an 83% effect on the dependent fraud variable.

**Table 5: Chi -Sq and correlation coefficient of R-squared**

| LL0 | LL1 | Chi-Sq | df | p-value | alpha | sig | R-Sq (L) |
|---|---|---|---|---|---|---|---|
| -1035.87 | -177.79 | 1716.155 | 6 | 0 | 0.05 | yes | 0.828366 |

### 5.4. Classification Table

The following relevant output is the Classification Table "Table 6" used to classify claims as fraudulent or not fraudulent, which compares the model predictions with the actual observations. Overall, 99.60% of predictions are correct, which is not 100% correct. However, this model is doing a terrible job of predicting better than a random prediction would yield a lower percentage; therefore, with these 6 independent variables, the model provides a better explanation of the results. The model performs better when predicting legitimate claims (99.9 %) versus fraudulent claims (79.3%).

**Table 6: Classification Table**

| **Classification Table** | | | |
|---|---|---|---|
| | Fraud | Not Fraud | |
| Fraud | 161 | 7 | 168 |
| Not Fraud | 42 | 12175 | 12217 |
| | 203 | 12182 | 12385 |
| Accuracy | 79.3% | 99.9% | 99.60% |
| Cutoff | 0.5 | | |

A cutoff value of 0.5 distinguishes predictions from those not predicting events. It is possible to use sensitivity = 95.8% in the fraud member detection model, where the percentage of observed positives was predicted to be positives, while specificity = 99.6%. Observed negatives that were predicted to be negatives. A higher level of sensitivity and specificity indicates a better model fit.

ROCs (Receiver Operating Characteristics) are used as a method for evaluating the constructed model's quality. This curve classifies positive and negative outcomes for all possible cutoffs. A model's predictability is measured by AUC, a mathematical measure of the model's predictability. AUC ranges from 0.5 (no predictive ability) to 1.0 (perfect predictive ability). Excellent classification results above the diametrical dividing line of the ROC space and poor classification results below are the two types of AUC. One can detect fraud with high accuracy using an AUC of 0.987, which is extremely high quality, as shown in Figure 2. AUC is said to have a significant effect on predictability.

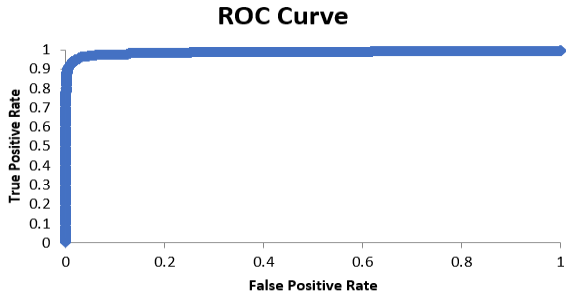



**Figure 2: ROC Curve**

## 6. Test of the Equation

An estimate of fraud probability can be obtained by adding the independent variables to the equation and estimating the estimated equation. Consider a contributor whose total costs include 50 $ for a doctor's clinic, 200 $ for medicines, 300 $ for laboratories, 150 $ for symptomatic diseases, and zero for COST-J01FA and COST-A02BC.

Logit (Fraud)= -11.7288+0.005188 (COST-OFFICE-PROC) + 0.005188 (COST-pharmacy) + 0.005131 (COST-Lab) +0.00061 (COST-SYMPTOMS) +0.002845 (COST-J01FA) + 0.004744 (COST-A02BC)

Logit (Fraud)=-11.7288+0.005188(50) + 0.005188 (200) + 0.005131 (300) + 0.00061 (150) +0.002845(0) +0.004744(0)

Logit (Fraud)= e^ -8.801 = 0.000150582

**logistic function of e^x / (1 + e^x) =** 0.000150582/ (1+0.000150582) =0%

The probability of this member becoming a fraud is equal to 0%. As a result, the model has worked well and successfully.

## 7. CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK.

Decision support systems play a significant role in developing and promoting the insurance industry, particularly by increasing customer acquisition and assessing risks accurately. Stakeholder premiums are affected by fraud; thus, detecting and preventing it is essential. Due to the large amount of data in the healthcare system, it is almost impossible to manually audit all of the transactions to determine whether they are fraudulent. The binary logistic regression fraud detection model developed in this report has been demonstrated to effectively detect fraudulent healthcare transactions requiring precise identification.

Detecting fraud is difficult due to its diversity and complexity. To ensure that a well-functioning healthcare system is capable of detecting fraud and preventing it from developing, new fraud detection models, including those that have not yet been developed, are needed.

Moreover, fraud detection can provide faster approaches and lower computational costs when applied to large datasets. A fraud detection form can recover fraudulent claim costs, reduce medical fees, and prevent unnecessary treatment. A reduction in fraud investigation time results from this.

To provide better medical services to real patients, save money, and improve the healthcare experience for patients with real needs, future work requires the development of other fraud models related to other variables and service providers, through which many benefits can be reaped.

## REFERENCES

Abou-Moghli, A. A., & Shatem, M. (2024). Examining the impact of e-governance on organizational strategy execution in the Jordanian ICT industry. Problems and Perspectives in Management, 22(3), 185.

Salih, A., Alsalhi, L., & Abou-Moghli, A. (2024). Entrepreneurial orientation and digital transformation as drivers of high organizational performance: Evidence from Iraqi private bank. *Uncertain Supply Chain Management*, *12*(1), 9-18.

Ahmad, A. Y. B., Kumari, D. K., Shukla, A., Deepak, A., Chandnani, M., Pundir, S., & Shrivastava, A. (2024). Framework for Cloud Based Document Management System with Institutional Schema of Database. International Journal of Intelligent Systems and Applications in Engineering, 12(3s), 672-678.

Ahmad, A. Y. B., Tiwari, A., Nayeem, M. A., Biswal, B. K., Satapathy, D. P., Kulshreshtha, K., & Bordoloi, D. (2024). Artificial Intelligence Perspective Framework of the Smart Finance and Accounting Management Model. International Journal of Intelligent Systems and Applications in Engineering, 12(4s), 586-594.

Ahmad, A., Abusaimeh, H., Rababah, A., Alqsass, M., Al-Olima, N., & Hamdan, M. (2024). Assessment of effects in advances of accounting technologies on quality financial reports in Jordanian public sector. Uncertain Supply Chain Management, 12(1), 133-142.

Ahmad Y. A. Bani Ahmad, "Firm Determinants that Influences Implementation of Accounting Technologies in Business Organizations," WSEAS Transactions on Business and Economics, vol. 21, pp. 1-11, 2024

Alhawamdeh, H., Al-Saad, S. A., Almasarweh, M. S., Al-Hamad, A. A.-S. A., Bani Ahmad, A. Y. A. B., & Ayasrah, F. T. M. (2023). The Role of Energy Management Practices in Sustainable Tourism Development: A Case Study of Jerash, Jordan. International Journal of Energy Economics and Policy, 13(6), 321–333. https://doi.org/10.32479/ijeep.14724

Allahham, M., & Ahmad, A. (2024). AI-induced anxiety in the assessment of factors influencing the adoption of mobile payment services in supply chain firms: A mental accounting perspective. International Journal of Data and Network Science, 8(1), 505-514.

Cheng, Congbin, Sayed Fayaz Ahmad, Muhammad Irshad, Ghadeer Alsanie, Yasser Khan, Ahmad Y. A. Bani Ahmad (Ayassrah), and Abdu Rahman Aleemi. 2023. "Impact of Green Process Innovation and Productivity on Sustainability: The Moderating Role of Environmental Awareness" Sustainability 15, no. 17: 12945. https://doi.org/10.3390/su151712945

Daoud, M., Taha, S., Al-Qeed, M., Alsafadi, Y., Ahmad, A., & Allahham, M. (2024). EcoConnect: Guiding environmental awareness via digital marketing approaches. International Journal of Data and Network Science, 8(1), 235-242.

Fraihat, B. A. M., Ahmad, A. Y. B., Alaa, A. A., Alhawamdeh, A. M., Soumadi, M. M., Aln'emi, E. A. S., & Alkhawaldeh, B. Y. S. (2023). Evaluating Technology Improvement in Sustainable Development Goals by Analysing Financial Development and Energy Consumption in Jordan. International Journal of Energy Economics and Policy, 13(4), 348

Lin, C., Ahmad, S. F., Ayassrah, A. Y. B. A., Irshad, M., Telba, A. A., Awwad, E. M., & Majid, M. I. (2023). Green production and green technology for sustainability: The mediating role of waste reduction and energy use. Heliyon, e22496.

K. Daoud, D. . Alqudah, M. . Al-Qeed, B. A. . Al Qaied, and A. Y. A. B. . Ahmad, "The Relationship Between Mobile Marketing and Customer Perceptions in Jordanian Commercial Banks: The Electronic Quality as A Mediator Variable", ijmst, vol. 10, no. 2, 2718–2729. Retrieved from https://kurdishstudies.net/menu-script/index.php/KS/article/view/831

Peiran Liang, Yulu Guo, Sohaib Tahir Chauhdary, Manoj Kumar Agrawal, Sayed Fayaz Ahmad, Ahmad Yahiya ,Ahmad Bani Ahmad, Ahmad A. Ifseisi, Tiancheng Ji,2024"Sustainable development and multi-aspect analysis of a novel polygeneration system using biogas upgrading and LNG regasification processes, ,producing power, heating, fresh water and liquid CO2",Process Safety and Environmental Protection

Peiran Liang, Yulu Guo, Tirumala Uday Kumar Nutakki, Manoj Kumar Agrawal, Taseer Muhammad, Sayed ,Fayaz Ahmad, Ahmad Yahiya Ahmad Bani Ahmad, Muxing Qin 2024. "Comprehensive assessment and sustainability improvement of a natural gas power plant utilizing an environmentally friendly combined cooling heating and power-desalination arrangement"Journal of Cleaner Production,Volume , ,436,,140387

Busch, "Electronic health records: An audit and internal control guide," John Wiley & Sons, 2008

Rawte, V. and Anuradha, G., "Fraud detection in health insurance using data mining techniques," in In 2015 International Conference on Communication, Information & Computing Technology (ICCICT), 2015.

Shrank, W.H., Rogstad, T.L. and Parekh, N., "Waste in the US health care system: estimated costs and potential for savings," Jama, vol. 322, no. 15, pp. 1501-1509, 2019.

Kose, I., Gokturk, M. and Kilic, K., "An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance," Applied Soft Computing, vol. 36, pp. 283-299, 2015.

Kirlidog, M. and Asuk, C., "A fraud detection approach with data mining in health insurance," Procedia-Social and Behavioral Sciences, pp. 989-994, 2012.

Liu, Q. and Vasarhelyi, M., "Healthcare fraud detection: A survey and a clustering model incorporating geo-location information," in In 29th world continuous auditing and reporting symposium (29WCARS), Brisbane, Australia., 2013.

Shatem, M., & Abou-Moghli, A. (2024). The moderating role of perceived environmental uncertainty in the impact of corporate governance on strategy implementation: An agency theory perspective. Uncertain Supply Chain Management, 12(3), 1577-1588.

Musal, "Two models to investigate Medicare fraud within unsupervised databases.," Expert Systems with Applications, vol. 37, no. 12, pp. 8628-8633, 2010

Magder, L.S. and Hughes, J.P., "Logistic regression when the outcome is measured with uncertainty," American journal of epidemiology, vol. 146, no. 2, pp. 195-203, 1997.

Marei, A., Ashal, N., Abou-Moghli, A., Daoud, L., & Lutfi, A. (2024). The effect of strategic orientation on operational performance: the mediating role of operational sustainability. *Business Strategy Review*, *5*(1), 346-355.

Liou, "Fraudulent financial reporting detection and business failure prediction models: a comparison," Managerial Auditing Journal., 2008.

Lim, T.S., Loh, W.Y. and Shih, Y.S., "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms," Machine learning, vol. 40, no. 3, pp. 203-228, 2000.

Hosmer, Lemeshow S. Applied logistic regression, 2000.

Kleinbaum, D.G. and Klein, M., "Introduction to logistic regression In Logistic regression," Springer, pp. 1-39, 2010.

Peng, C.Y.J., Lee, K.L. and Ingersoll, G.M., "An introduction to logistic regression analysis and reporting," The journal of educational research, vol. 96, no. 1, pp. 3-14, 2002.

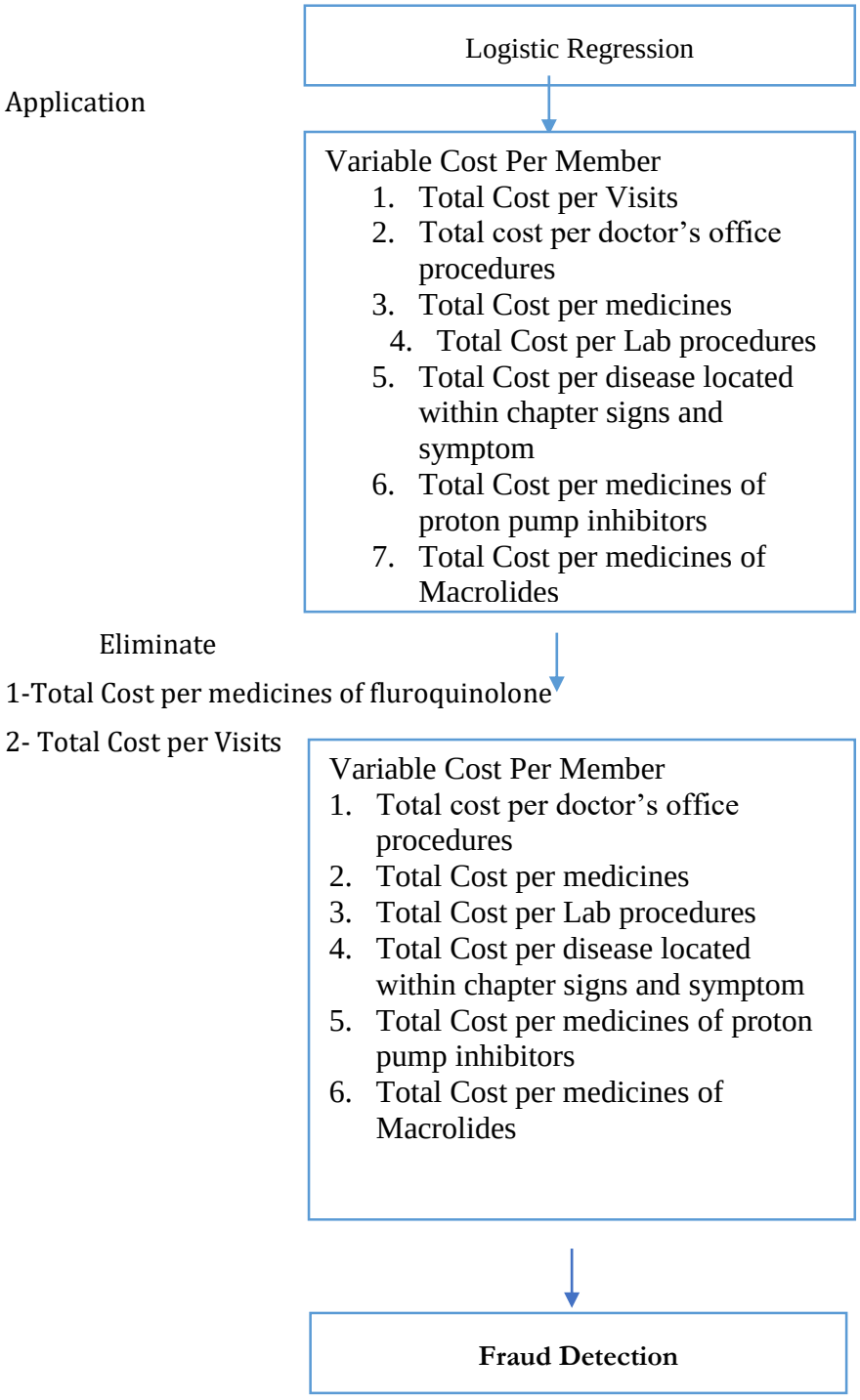
William, P., Ahmad, A. Y. B., Deepak, A., Gupta, R., Bajaj, K. K., & Deshmukh, R. (2024). Sustainable Implementation of Artificial Intelligence Based Decision Support System for Irrigation Projects in the Development of Rural Settlements. International Journal of Intelligent Systems and Applications in Engineering, 12(3s), 48-56.

Yahiya Ahmad Bani Ahmad (Ayassrah), Ahmad; Ahmad Mahmoud Bani Atta, Anas; Ali Alawawdeh, Hanan; Abdallah Aljundi, Nawaf; Morshed, Amer; and Amin Dahbour, Saleh (2023) "The Effect of System Quality and User Quality of Information Technology on Internal Audit Effectiveness in Jordan, And the Moderating Effect of Management Support," Applied Mathematics & Information Sciences: Vol. 17: Iss. 5, Article 12. DOI: https://dx.doi.org/10.18576/amis/170512

Yahiya, A., & Ahmad, B. (2024). Automated debt recovery systems: Harnessing AI for enhanced performance. Journal of Infrastructure, Policy and Development, 8(7), 4893.

Zhan, Y., Ahmad, S. F., Irshad, M., Al-Razgan, M., Awwad, E. M., Ali, Y. A., & Ayassrah, A. Y. B. A. (2024). Investigating the role of Cybersecurity's perceived threats in the adoption of health information systems. Heliyon, 10(1).

**Appendix 1: Logistic Regression Model**



## Appendix 2: Figures

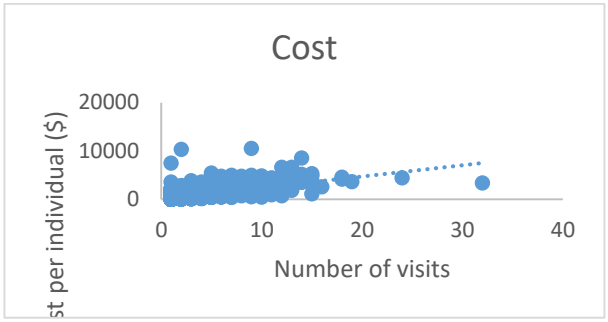



**Figure 3: correlation between number of visits and visit cost per member**

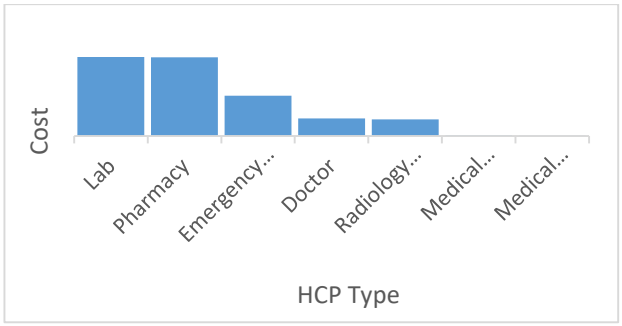



**Figure 4: Distribution of the cost per healthcare provider type.**

| CORREL | Fraud or not | NUM_OF_CLAIMS | NUM_OF_VISITS | COST-Visits | #of Followup Visits | NUMOF-RESPONSIBLE DOCTORS | COST-OFFICE-PROC | NUMOF-OFFICE-PROC | COST-pharmacy | NUMOF-pharmacy | COST-Lab | NUMOF-Lab | COST-SYMPTOMS | NUMOF-SYMPTOMS | COST-M01A | NUMOF-M01A | COST-J01MA | NUMOF-J01MA | COST-J01FA | NUMOF-J01FA | COST-A02BC | NUMOF-A02BC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fraud or not | 1.00 | 0.34 | 0.37 | 0.64 | 0.24 | 0.18 | 0.54 | 0.35 | 0.35 | 0.26 | 0.24 | 0.18 | 0.35 | 0.18 | 0.22 | 0.17 | 0.19 | 0.18 | 0.06 | 0.03 | 0.21 | 0.18 |
| NUM_OF_CLAIMS | 0.34 | 1.00 | 0.80 | 0.70 | 0.27 | 0.40 | 0.23 | 0.73 | 0.60 | 0.71 | 0.63 | 0.71 | 0.44 | 0.55 | 0.35 | 0.38 | 0.38 | 0.41 | 0.33 | 0.33 | 0.41 | 0.46 |
| NUM_OF_VISITS | 0.37 | 0.80 | 1.00 | 0.68 | 0.46 | 0.46 | 0.38 | 0.79 | 0.61 | 0.71 | 0.35 | 0.37 | 0.29 | 0.30 | 0.36 | 0.41 | 0.33 | 0.36 | 0.35 | 0.36 | 0.40 | 0.44 |
| COST-Visits | 0.64 | 0.70 | 0.68 | 1.00 | 0.31 | 0.37 | 0.72 | 0.63 | 0.58 | 0.49 | 0.52 | 0.45 | 0.51 | 0.39 | 0.28 | 0.25 | 0.33 | 0.31 | 0.19 | 0.16 | 0.37 | 0.35 |
| #of Followup Visits | 0.24 | 0.27 | 0.46 | 0.31 | 1.00 | 0.13 | 0.24 | 0.32 | 0.14 | 0.13 | 0.21 | 0.20 | 0.14 | 0.11 | 0.05 | 0.05 | 0.08 | 0.09 | 0.02 | 0.02 | 0.09 | 0.10 |
| IUMOF-RESPONSIBLE DOCTOR | 0.18 | 0.40 | 0.46 | 0.37 | 0.13 | 1.00 | 0.22 | 0.41 | 0.22 | 0.25 | 0.25 | 0.25 | 0.26 | 0.26 | 0.07 | 0.08 | 0.15 | 0.16 | 0.09 | 0.09 | 0.21 | 0.23 |
| COST-OFFICE-PROC | 0.54 | 0.23 | 0.38 | 0.72 | 0.24 | 0.22 | 1.00 | 0.46 | 0.13 | 0.15 | 0.10 | 0.06 | 0.29 | 0.12 | 0.08 | 0.08 | 0.07 | 0.07 | 0.06 | 0.06 | 0.12 | 0.12 |
| NUMOF-OFFICE-PROC | 0.35 | 0.73 | 0.79 | 0.63 | 0.32 | 0.41 | 0.46 | 1.00 | 0.46 | 0.57 | 0.27 | 0.29 | 0.36 | 0.39 | 0.31 | 0.34 | 0.28 | 0.31 | 0.29 | 0.29 | 0.34 | 0.37 |
| COST-pharmacy | 0.35 | 0.60 | 0.61 | 0.58 | 0.14 | 0.22 | 0.13 | 0.46 | 1.00 | 0.82 | 0.06 | 0.07 | 0.17 | 0.16 | 0.52 | 0.49 | 0.55 | 0.51 | 0.36 | 0.33 | 0.54 | 0.52 |
| NUMOF-pharmacy | 0.26 | 0.71 | 0.71 | 0.49 | 0.13 | 0.25 | 0.15 | 0.57 | 0.82 | 1.00 | 0.04 | 0.06 | 0.15 | 0.18 | 0.53 | 0.61 | 0.49 | 0.52 | 0.52 | 0.53 | 0.52 | 0.57 |
| COST-Lab | 0.24 | 0.63 | 0.35 | 0.52 | 0.21 | 0.25 | 0.10 | 0.27 | 0.06 | 0.04 | 1.00 | 0.90 | 0.48 | 0.52 | -0.03 | -0.06 | 0.05 | 0.06 | -0.04 | -0.05 | 0.07 | 0.07 |
| NUMOF-Lab | 0.18 | 0.71 | 0.37 | 0.45 | 0.20 | 0.25 | 0.06 | 0.29 | 0.07 | 0.06 | 0.90 | 1.00 | 0.43 | 0.58 | -0.01 | -0.04 | 0.07 | 0.09 | -0.02 | -0.03 | 0.08 | 0.09 |
| COST-SYMPTOMS | 0.35 | 0.44 | 0.29 | 0.51 | 0.14 | 0.26 | 0.29 | 0.36 | 0.17 | 0.15 | 0.48 | 0.43 | 1.00 | 0.81 | 0.07 | 0.05 | 0.14 | 0.14 | 0.01 | 0.00 | 0.25 | 0.25 |
| NUMOF-SYMPTOMS | 0.18 | 0.55 | 0.30 | 0.39 | 0.11 | 0.26 | 0.12 | 0.39 | 0.16 | 0.18 | 0.52 | 0.58 | 0.81 | 1.00 | 0.07 | 0.07 | 0.14 | 0.16 | 0.03 | 0.02 | 0.23 | 0.25 |
| COST-M01A | 0.22 | 0.35 | 0.36 | 0.28 | 0.05 | 0.07 | 0.08 | 0.31 | 0.52 | 0.53 | -0.03 | -0.01 | 0.07 | 0.07 | 1.00 | 0.83 | 0.35 | 0.31 | 0.25 | 0.24 | 0.24 | 0.25 |
| NUMOF-M01A | 0.17 | 0.38 | 0.41 | 0.25 | 0.05 | 0.08 | 0.08 | 0.34 | 0.49 | 0.61 | -0.06 | -0.04 | 0.05 | 0.07 | 0.83 | 1.00 | 0.26 | 0.25 | 0.32 | 0.33 | 0.22 | 0.25 |
| COST-J01MA | 0.19 | 0.38 | 0.33 | 0.33 | 0.08 | 0.15 | 0.07 | 0.28 | 0.55 | 0.49 | 0.05 | 0.07 | 0.14 | 0.14 | 0.35 | 0.26 | 1.00 | 0.90 | 0.09 | 0.05 | 0.34 | 0.35 |
| NUMOF-J01MA | 0.18 | 0.41 | 0.36 | 0.31 | 0.09 | 0.16 | 0.07 | 0.31 | 0.51 | 0.52 | 0.06 | 0.09 | 0.14 | 0.16 | 0.31 | 0.25 | 0.90 | 1.00 | 0.09 | 0.06 | 0.33 | 0.35 |
| COST-J01FA | 0.06 | 0.33 | 0.35 | 0.19 | 0.02 | 0.09 | 0.06 | 0.29 | 0.36 | 0.52 | -0.04 | -0.02 | 0.01 | 0.03 | 0.25 | 0.32 | 0.09 | 0.09 | 1.00 | 0.91 | 0.19 | 0.24 |
| NUMOF-J01FA | 0.03 | 0.33 | 0.36 | 0.16 | 0.02 | 0.09 | 0.06 | 0.29 | 0.33 | 0.53 | -0.05 | -0.03 | 0.00 | 0.02 | 0.24 | 0.33 | 0.05 | 0.06 | 0.91 | 1.00 | 0.16 | 0.21 |
| COST-A02BC | 0.21 | 0.41 | 0.40 | 0.37 | 0.09 | 0.21 | 0.12 | 0.34 | 0.54 | 0.52 | 0.07 | 0.08 | 0.25 | 0.23 | 0.24 | 0.22 | 0.34 | 0.33 | 0.19 | 0.16 | 1.00 | 0.90 |
| NUMOF-A02BC | 0.18 | 0.46 | 0.44 | 0.35 | 0.10 | 0.23 | 0.12 | 0.37 | 0.52 | 0.57 | 0.07 | 0.09 | 0.25 | 0.25 | 0.25 | 0.25 | 0.35 | 0.35 | 0.24 | 0.21 | 0.90 | 1.00 |

**Figure 5: Pearson correlations for potential independent variables and dependent variable.**

## Appendix 3: Tables

| Sequence ID | Cost | Z SCORE |
|---|---|---|
| 1 | 2,813 | 4.332161 |
| 2 | 2,372 | 3.505702 |
| 3 | 2,210 | 3.202907 |
| 4 | 3,371 | 5.377504 |
| 5 | 4,472 | 7.439718 |
| 6 | 3,202 | 5.06096 |
| 7 | 2,877 | 4.452223 |
| 8 | 2,132 | 3.05681 |
| 9 | 2,183 | 3.152335 |
| 10 | 2,765 | 4.242443 |
| 11 | 2,826 | 4.356698 |
| 12 | 2,152 | 3.094271 |
| 13 | 2,139 | 3.069921 |
| 14 | 4,401 | 7.306732 |
| 15 | 2,842 | 4.386667 |
| 16 | 3,048 | 4.772512 |
| 17 | 3,854 | 6.282181 |
| 18 | 2,182 | 3.150462 |
| 19 | 2,580 | 3.895931 |
| 20 | 2,162 | 3.113001 |
| 21 | 8,552 | 15.08171 |
| 22 | 3,087 | 4.845561 |
| 23 | 2,168 | 3.124239 |
| 24 | 2,981 | 4.647356 |
| 25 | 4,473 | 7.441591 |
| 26 | 2,539 | 3.818256 |
| 27 | 5,409 | 9.194754 |
| 28 | 2,190 | 3.165446 |
| 29 | 2,360 | 3.483863 |
| 30 | 2,692 | 4.104962 |
| 31 | 2,799 | 4.306126 |
| 32 | 4,292 | 7.102946 |
| 33 | 3,029 | 4.736925 |
| 34 | 2,578 | 3.892185 |
| 35 | 2,189 | 3.163573 |
| 36 | 2,648 | 4.023297 |
| 37 | 3,712 | 6.015273 |
| 38 | 4,784 | 8.024106 |
| 39 | 3,759 | 6.104242 |
| 40 | 3,355 | 5.347535 |
| 41 | 2,121 | 3.036207 |

| | | |
|---|---|---|
| 42 | 2,167 | 3.122366 |
| 43 | 2,396 | 3.551292 |
| 44 | 2,312 | 3.393957 |
| 45 | 2,261 | 3.298432 |
| 46 | 3,484 | 5.589138 |
| 47 | 2,279 | 3.332896 |
| 48 | 2,236 | 3.251606 |
| 49 | 3,444 | 5.513299 |
| 50 | 3,798 | 6.177291 |
| 51 | 2,994 | 4.671368 |
| 52 | 2,839 | 4.381048 |
| 53 | 2,289 | 3.350877 |
| 54 | 2,325 | 3.418306 |
| 55 | 2,185 | 3.155144 |
| 56 | 2,253 | 3.283448 |
| 57 | 3,557 | 5.726806 |
| 58 | 5,181 | 8.767702 |
| 59 | 2,791 | 4.291142 |
| 60 | 2,242 | 3.262844 |
| 61 | 4,956 | 8.346268 |
| 62 | 10,530 | 18.78658 |
| 63 | 2,576 | 3.888439 |
| 64 | 3,046 | 4.768766 |
| 65 | 2,238 | 3.254415 |
| 66 | 2,251 | 3.279701 |
| 67 | 2,500 | 3.746088 |
| 68 | 2,185 | 3.156081 |
| 69 | 2,194 | 3.172938 |
| 70 | 4,010 | 6.574375 |
| 71 | 2,289 | 3.34994 |
| 72 | 2,183 | 3.151398 |
| 73 | 2,591 | 3.916534 |
| 74 | 2,641 | 4.010186 |
| 75 | 2,119 | 3.03246 |
| 76 | 4,272 | 7.06511 |
| 77 | 2,653 | 4.032663 |
| 78 | 3,355 | 5.347535 |
| 79 | 2,642 | 4.012059 |
| 80 | 2,892 | 4.480319 |
| 81 | 3,326 | 5.293217 |
| 82 | 3,626 | 5.855128 |
| 83 | 2,326 | 3.420179 |
| 84 | 2,994 | 4.671368 |
| 85 | 4,937 | 8.309744 |

| | | |
|---|---|---|
| 86 | 2,790 | 4.288332 |
| 87 | 3,606 | 5.81793 |
| 88 | 3,647 | 5.894462 |
| 89 | 2,405 | 3.568149 |
| 90 | 3,155 | 4.972927 |
| 91 | 2,972 | 4.630162 |
| 92 | 4,477 | 7.449083 |
| 93 | 2,785 | 4.279904 |
| 94 | 2,937 | 4.564605 |
| 95 | 4,573 | 7.628895 |
| 96 | 3,051 | 4.777195 |
| 97 | 2,301 | 3.373353 |
| 98 | 2,178 | 3.14297 |
| 99 | 3,577 | 5.763349 |
| 100 | 3,506 | 5.630364 |
| 101 | 3,286 | 5.218295 |
| 102 | 4,582 | 7.645752 |
| 103 | 2,305 | 3.380845 |
| 104 | 2,409 | 3.575641 |
| 105 | 2,732 | 4.181382 |
| 106 | 2,582 | 3.899677 |
| 107 | 2,375 | 3.511958 |
| 108 | 2,122 | 3.03808 |
| 109 | 2,185 | 3.156081 |
| 110 | 3,441 | 5.50768 |
| 111 | 6,603 | 11.43116 |
| 112 | 2,759 | 4.230455 |
| 113 | 2,971 | 4.628288 |
| 114 | 3,517 | 5.650967 |
| 115 | 3,012 | 4.705645 |
| 116 | 2,230 | 3.240368 |
| 117 | 2,252 | 3.281574 |
| 118 | 3,151 | 4.965435 |
| 119 | 4,850 | 8.147726 |
| 120 | 2,500 | 3.746088 |
| 121 | 3,587 | 5.781143 |
| 122 | 6,622 | 11.46675 |
| 123 | 2,267 | 3.30967 |
| 124 | 2,203 | 3.189065 |
| 125 | 10,325 | 18.40171 |
| 126 | 2,862 | 4.424127 |
| 127 | 2,516 | 3.776056 |
| 128 | 3,160 | 4.981356 |
| 129 | 4,757 | 7.973534 |

| | | |
|---|---|---|
| 130 | 2,669 | 4.062631 |
| 131 | 2,760 | 4.23216 |
| 132 | 3,928 | 6.420785 |
| 133 | 2,516 | 3.776056 |
| 134 | 2,642 | 4.011123 |
| 135 | 2,124 | 3.041826 |
| 136 | 2,571 | 3.879073 |
| 137 | 4,717 | 7.898612 |
| 138 | 2,309 | 3.388338 |
| 139 | 2,868 | 4.435366 |
| 140 | 2,190 | 3.165446 |
| 141 | 5,274 | 8.940958 |
| 142 | 4,185 | 6.902156 |
| 143 | 3,555 | 5.722142 |
| 144 | 3,493 | 5.606014 |
| 145 | 3,577 | 5.763349 |
| 146 | 2,277 | 3.3284 |
| 147 | 2,253 | 3.282511 |
| 148 | 2,771 | 4.253681 |
| 149 | 3,262 | 5.173342 |
| 150 | 2,608 | 3.948376 |
| 151 | 2,201 | 3.18605 |
| 152 | 2,408 | 3.573768 |
| 153 | 2,286 | 3.345258 |
| 154 | 2,471 | 3.69177 |
| 155 | 2,183 | 3.152335 |
| 156 | 2,704 | 4.128187 |
| 157 | 3,200 | 5.057214 |
| 158 | 4,810 | 8.072805 |
| 159 | 2,229 | 3.238495 |
| 160 | 2,275 | 3.325198 |
| 161 | 3,215 | 5.08531 |
| 162 | 2,302 | 3.375226 |
| 163 | 2,859 | 4.418508 |
| 164 | 3,152 | 4.967308 |
| 165 | 2,489 | 3.725484 |
| 166 | 3,469 | 5.561061 |
| 167 | 2,643 | 4.014232 |
| 168 | 4,991 | 8.412199 |
| 169 | 2,400 | 3.558784 |
| 170 | 3,665 | 5.928177 |
| 171 | 2,106 | 3.007174 |
| 172 | 2,631 | 3.991456 |
| 173 | 2,122 | 3.03808 |

| | | |
|---|---|---|
| 174 | 2,860 | 4.420381 |
| 175 | 2,415 | 3.58688 |
| 176 | 2,820 | 4.34546 |
| 177 | 2,224 | 3.229129 |
| 178 | 2,600 | 3.933392 |
| 179 | 2,154 | 3.098335 |
| 180 | 7,500 | 13.11128 |
| 181 | 2,218 | 3.217891 |
| 182 | 2,200 | 3.184177 |
| 183 | 3,905 | 6.377069 |
| 184 | 4,161 | 6.857203 |
| 185 | 2,121 | 3.036207 |
| 186 | 2,352 | 3.468878 |
| 187 | 2,958 | 4.603939 |
| 188 | 2,698 | 4.116949 |
| 189 | 3,839 | 6.253149 |
| 190 | 2,723 | 4.162839 |

**Table 7: Z-score for cost per member with larger than 3 value**

| Lab Procedure | Z- score |
|---|---|
| ALANINE AMINO (ALT) (SGPT) | 4.501075 |
| ASSAY OF CREATINE | 3.093474 |
| ASSAY OF CREATININE, CREATININE BLOOD | 3.550674 |
| ASSAY THYROID STIM HORMONE, TSH | 3.714474 |
| ASSAY, GLUCOSE, BLOOD QUANT, FBS | 4.229275 |
| COMPLETE CBC W/AUTO DIFF WBC | 6.580077 |
| COMPLETE CBC, AUTOMATED | 6.432476 |

**Table 8: Z-score for the cost of the lab.**

| Chapter | cost | Z- score |
|---|---|---|
| SYMPTOMS, SIGNS, AND ILL-DEFINED CONDITIONS | 1,956,593 | 3.347925 |

**Table 9 : Z-score for cost of the chapter symptom conditions.**

| Therapeutic Class | cost | z-score |
|---|---|---|
| Anti-Inflamma & Anti-Rheumat Products,Non-Steroids | 96,294 | 3.858747 |
| Fluroquinolones | 195,625 | 8.195608 |
| Macrolides | 103,513 | 4.173929 |
| Proton Pump Inhibitors | 200,397 | 8.403953 |

**Table 10: Z-Score for Cost of therapeutic class of medicines.**

| | *coeff b* | *p-value* |
|---|---|---|
| Intercept | -0.37316 | 2.54E-24 |
| COST-Visits | 0.00062 | 0.011909 |

| | | |
|---|---|---|
| COST-OFFICE-PROC | 0.00055 | 0.027148 |
| COST-pharmacy | 0.00576 | 8.27E-38 |
| COST-Lab | 0.00559 | 1.39E-60 |
| COST-SYMPTOMS | 0.000459 | 0.001917 |
| COST-M01A | 0.021172 | 3.8E-47 |
| COST-J01MA | 0.008788 | 1.6E-23 |
| COST-J01FA | 0.17009 | 8.31E-12 |
| COST-A02BC | 0.004187 | 0.00022 |