A Binary Logistic Regression Analysis of Citations Received by Patents:   
Exploring Alternative Approaches to Measuring and Predicting Technology Transfer Outcomes

Malcolm S. Townes

Saint Louis University

Malcolm,

While your data set might not meant for a binary logistic regression analysis, you have done an excellent job in this assignment as always. I wrote some comments in the text below.

You have a choice of either preparing a paper or a poster. The model in the 3rd assignment will be used in the final product besides other tables.

I look forward to your presentation on 11/27.

Excellent work! HM

Abstract

Technology transfer is the transition of technology or intellectual property from one person or entity to another person or entity. Improving the pace and level of the transfer of technology derived from federally-funded research and development (R&D) is a priority for the public policy of the United States of America. As such, understanding the drivers of technology transfer outcomes and improving methods for measuring those outcomes is an important topic for study. The potential benefits of developing models that describe the technology transfer process and facilitate better understanding of the factors associated with successful technology transfer include better management of technological innovation, more effective prioritization of high potential technologies for development, and more efficient allocation of federal and private sector resources. This study used binary logistic regression analysis to explore an alternative approach to measuring and predicting technology transfer outcomes using citations received by patents as an indication of technology transfer.

Keywords: technology transfer, university technology transfer, technology commercialization, federally funded research and development, patents, patent citations, logistic regression

**Introduction**

This study continues the investigation of how technology transfer success can be defined and measured that I began in Assignments 01, 02, and 03 for SOC 6100 in the Fall 2018 semester. Improving the transfer of technology derived from federally funded research and development (R&D) to the private sector is a priority for the public policy of the United States of America (OMB, 2018). In fact, increasing the return on investment from federally funded R&D has been a top priority for the U.S. government going back to the Bush administration of the early 2000s (OMB, 2002) and interest in this topic can be traced as far back as the 1940s (Bush, 1949).

In this study, I conducted a binary logistic regression analysis to investigate the relationship between the number of citations received by patents and several patent citation data variables. The purpose of this study was to better understand the drivers of technology transfer outcomes and explore alternative approaches to measuring them. Most previous studies of the subject define technology transfer in terms of transactional outputs and outcomes such as the executing license agreements, securing sponsored research, or forming new business ventures. This study, I seek to explore whether a non-transactional indication of knowledge transfer can be a useful measurement of technology transfer outputs and outcomes. As with the previous analyses, I used patents issued by the United States Patent and Trademark Office (USPTO) as a proxy for units of technology and the number of citations a given U.S. patent receives from other U.S. patents as a measure of technology transfer.

**Literature Review**

A cursory review of the literature revealed several previous studies that applied logistic regression analysis to investigate technology transfer or trends in patent data. Appio, Martini & Fantoni (2017) used a series of logistic regression models to explore the role of scientific and technological diversity in developing impactful bioinformatics inventions as measured by forward citation distribution. They found that different degrees of knowledge diversity were associated with different degrees of impact but combinations of scientific and technological knowledge diversity did not always lead to impactful inventions.

Kirkman (2013) used multinomial logistic regression to study how the innovativeness, proactiveness, and risk taking propensity of biotechnology firms influenced their selection of technology transfer modes. Kirkman specifically limited the modes of technology transfer in the study to licensing, sponsored research, and consulting agreements, which are all transactional based exchanges.

Yoshikane (2013) studied the citation frequency of patents using several types of regression analysis including multiple linear, logistic, and binomial. Yoshikane found that the number of classifications for a patent were associated with citation frequency.

**Research Questions and Theoretical Perspective**

The purpose of this study is to explore how technology transfer success can be defined and measured. While most other research conducted in this area conceptualize technology transfer as the transactional exchange of legally recognized intellectual property through a formal license, I consider an alternative conceptualization of technology transfer as a non-transactional exchange of knowledge. To explore this conceptualization, I use patents as a proxy for technology and citations of one patent by other patents as an indication of technology transfer. Specifically, I investigate the following questions:

1. Can patents that receive more than the median number of citations be effectively discriminated from patents that receive the median number of citations or less based on specific patent data variables?
2. Is there a significant association between whether a patent receives more than the median number of citations received by all patents with the year a patent was granted, the number of claims made by a patent, the originality of the patent, and the generality of the patent?

Based on previous analysis, I theorize that the number of claims made by a patent and the generality of the patent will both have a positive association with the probability that a patent receives more than the median number of citations. Moreover, I anticipate that the association for the number of claims made by a patent will be slight. While the number of claims made by a patent will have a positive association, more claims does not create significantly more opportunities for that patent to be cited because a significant portion of the claims of a patent are often dependent claims, which means that they are dependent on other claims in the patent. I suspect that the more general a patent, the more opportunities it has to be cited by other patents across multiple fields. I expect the year a patent was granted and the originality of the patent to be negatively associated with the probability that a patent receives more than the median number of citations from other patents. I suspect that the older a patent is the less relevant it becomes because of the half-life of knowledge. Also, it’s possible that the more original a patent the more difficult it is for other inventors and innovators to conceive applications of the technology in their fields.

**Data and Methods**

**Data Sources**

This study uses a subset of 2,000 observations taken from patent data obtained from the National Bureau of Economic Research (NBER) website. The source data contains both original and constructed variables. The data file included all utility patents granted in the U.S. from January 1, 1963 to December 30, 1999 listed in the Technology Assessment and Forecast (TAF) database of the USPTO. The source file contained data on 2,923,922 patents across 23 variables.

Table 1 details the original USPTO variables of the source data and explanations of their meanings. PATENT indicates the number assigned by the USPTO to the allowed patent. GYEAR is the year the USPTO allowed the patent. APPYEAR is the year the patent application was submitted to the USPTO. GDATE is the number of weeks elapsed since January 1, 1960 to the date the USPTO allowed the patent. COUNTRY is the country of citizenship for the first inventor listed on the patent application. POSTATE is the state of residency for the first inventor listed on the patent application. ASSIGNEE indicates to whom the patent is assigned and is unique to each assignee. ASSCODE indicates the type of assignee. CLAIMS is the number of independent and dependent claims listed on the patent. NCLASS indicates the broad classification for the patent.

Table 1

Original USPTO Variables of Source Data

| Variable | Variable Type | Extended Name | Description |
| --- | --- | --- | --- |
| PATENT | Numeric  Nominal | Patent Number | The number assigned to the allowed patent by the USPTO.  Takes on values integer values between 3070801 and 6009554. |
| GYEAR | Numeric  Interval | Grant Year | The year the USPTO allowed the patent.  Takes on integer values between 1963 – 1999. |
| GDATE | Numeric  Interval | Grant Date | The date the USPTO allowed the patent expressed in terms of the number of weeks elapsed since  January 1, 1960.  Takes on integer values between 156 and 2,028. |
| APPYEAR | Numeric  Interval | Application Year | The year the patent application was submitted to the USPTO.  Takes on integer values between 1963 – 1999. |
| COUNTRY | Character  Nominal | Country of First Inventor | The country of citizenship for the first inventor listed on the patent application.  Takes on values of two character string data. |
| POSTATE | Character  Nominal | State of First Inventor (US) | The state of residency for the first inventor listed on the patent application if the country of citizenship is the United States of America.  Takes on values of two character string data. |
| ASSIGNEE | Numeric  Nominal | Assignee Identifier | Unique identifier for the assignee of the patent.  Takes on values from 10950 to 99550. |
| ASSCODE | Numeric  Nominal | Assignee Code | A one character code categorizing the type of assignee.  Takes on values from 1 to 7. |
| CLAIMS | Numeric  Interval | Number of Claims | Number of independent and dependent claims on the patent.  Takes on integer values from 1 to . |
| NCLASS | Numeric  Nominal | Main Patent Class | A code that categorizes the patent into one of several broad classifications.  Takes on integer values from 1 to 800. |

Table 2 provides information about the source data constructed variables and explanations of their meanings. CAT is a higher-level classification of the main patent class. SUBCAT is a sub-category of the main patent class. CMADE indicates the number of citations made by the patent. CRECEIVE indicates the number of citations in other patents that reference the patent. RATIOCIT is the ratio of the number of citations made by all patents granted since 1963 to the total number of citations made by the patent. GENERAL is a measure of how broad the influence of a patent spans across fields. ORIGINAL is a measure of the originality of the patent. FWDAPLAG measures forward citations lag. BCKGTLAG measures backward citations lag. SELFCTUB is the upper bound for the share of citations the patent makes to other patents assigned to the same assignee (i.e., self-citations made). SELFCTLB is the lower bound for the share of citations the patent makes to other patents assigned to the same assignee. SECUPBD is the upper bound for the share of citations the patent receives from other patents assigned to the same assignee (i.e., self-citations received). SECDLWBD is the lower bound for the share of citations the patent receives from other patents assigned to the same assignee.

Table 2

Source Data Constructed Variables

| Variable | Variable Type | Extended Name | Description |
| --- | --- | --- | --- |
| CAT | Numeric  Nominal | Technological Category | A higher-level classification of the Main Patent Class.  Takes on integer values from 1 to 6. |
| SUBCAT | Numeric  Nominal | Technological Sub-category | The sub-category of the primary technological category to which the patent is assigned.  Takes on integer values from 1 to 69. |
| CMADE | Numeric  Interval | Number of Citations Made | The number of citations made by the patent.  Takes on integer values from 1 to . |
| CRECEIVE | Numeric  Interval | No. of Citations Received | The number of citations in other patents that reference the patent.  Takes on integer values from 1 to . |
| RATIOCIT | Numeric  Ratio | Percent of Citations Made to Patents Granted Since 1963 | The ratio of the number of citations made by all patents granted since 1963 to the total number of citations made by the particular patent.  Takes on continuous values between 0 and 1. |
| GENERAL | Numeric  Ratio | Measure of Generality | A measure of how broad the influence of a patent spans across fields as determined by the number of different fields of all patents that cite the patent of interest.  Calculated as the following:  Generalityi = 1 - , where *sij* denotes the percentage of citations received by patent *i* that belong to patent class *j*, out of *ni* patent classes.  Takes on continuous values between 0 and 1. |
| ORIGINAL | Numeric  Ratio | Measure of Originality | A measure of the originality of a patent as determined by the number of different fields for all patents cited by the patent of interest.  Calculated as the following:  Originalityi = 1 - , where *sij* denotes the percentage of citations made by patent *i* that belong to patent class *j*, out of *ni* patent classes.  Takes on continuous values between 0 and 1. |
| FWDAPLAG | Numeric  Ratio | Mean Forward Citation Lag | The mean time difference between the application or grant date of the patent and that of the other patents citing this patent.  Takes on continuous values between 0 and 1. |
| BCKGTLAG | Numeric  Ratio | Mean Backward Citation Lag | The mean time difference between the application or grant date of the patent and that of the patents it cites.  Takes on continuous values between 0 and 1. |
| SELFCTUB | Numeric  Ratio | Share of Self-Citations Made – Upper Bound | The number of citations made by the patent to other patents with the same assignee divided by the total number of citations made by all patents with assignee codes.  Takes on continuous values between 0 and 1. |
| SELFCTLB | Numeric  Ratio | Share of Self-Citations Made – Lower Bound | The number of citations made by the patent to other patents with the same assignee divided by the total number of citations made by all patents.  Takes on continuous values between 0 and 1. |
| SECUPBD | Numeric  Ratio | Share of Self-Citations Received – Upper Bound | The number of citations received by the patent from other patents with the same assignee divided by the total number of citations received by all patents with assignee codes.  Takes on continuous values between 0 and 1. |
| SECDLWBD | Numeric  Ratio | Share of Self-Citations Received – Lower Bound | The number of citations received by the patent from other patents with the same assignee divided by the total number of citations received by all patents.  Takes on continuous values between 0 and 1. |

**Data Selection and Modification**

For this study, I only used the GYEAR, CLAIMS, CRECEIVE, GENERAL, and ORIGINAL variables in the analysis. Based on the results of previous analyses, I made several modifications to the data and incorporated several previous observations into the logistic regression model.

I removed the following variables because of high multicollinearity: APPYEAR, BCKGTLAG, FWDAPLAG, SELFCTLB, and SECDLWBD. I created a new variable called CRECBINARY, which I used as the dependent variable (DV) of interest in the analysis. I created the CRECBINARY variable using the Transform > Recode into Different Variables function of IBM SPSS Statistics 25. The CRECBINARY variable is a dichotomous variable calculated from the CRECEIVE variable using a median split of the data. The median of the CRECEIVE data was 2 citations received. I coded cases that had 2 or fewer citations received as 1. I coded cases with 3 or more citations received as 2. For the analysis, SPSS Statistics 25 internally coded CRECBINARY as 0 for cases where the number of citations received was less than or equal to 2 and as 1 for cases with 3 or more citations received. Table 3 lists the final variables that I used in the analysis. All of the IVs used in the analysis were continuous interval or ratio variables.

Table 3

Variables Used in Analysis

| Variable | Variable Type | Extended Name | Description |
| --- | --- | --- | --- |
| GYEAR | Numeric  Interval | Grant Year | The year the USPTO allowed the patent.  Takes on integer values between 1963 – 1999. |
| CLAIMS | Numeric  Interval | Number of Claims | Number of independent and dependent claims on the patent.  Takes on integer values from 1 to . |
| CRECBINARY | Numeric  Nominal | No. of Citations Received | 1 indicates 0-2 citations  2 indicates 3 or more  Takes on an integer value of 1 or 2. |
| GENERAL | Numeric  Ratio | Measure of Generality | A measure of how broad the influence of a patent spans across fields as determined by the number of different fields of all patents that cite the patent of interest.  Takes on continuous values between 0 and 1. |
| ORIGINAL | Numeric  Ratio | Measure of Originality | A measure of the originality of a patent as determined by the number of different fields for all patents cited by the patent of interest.  Takes on continuous values between 0 and 1. |

**Analysis and Results**

I used IBM SPSS Statistics 25 to analyze the data. I used the Analyze > Regression > Binary Logistic function to perform a binary logistic regression analysis using CRECBINARY as the DV and CLAIMS, GYEAR, GENERAL, and ORIGINAL as the independent variables (IVs). I used the Enter method for the regression. For the Logistic Regression Options I selected Correlations of estimates, Hosmer-Lemeshow goodness-of-fit, Iteration history, and Confidence interval for exp(B) of 95 percent. I left the Classification cutoff at the 0.5 default, set the maximum iterations to 30, and included the constant in the model. The complete IBM SPSS Statistics 25 output file for the analysis is shown in Appendix A.

Table 1 shows details for the equation variables of the regression analysis. All IVs were significant at the 0.05 level. The CLAIMS, GYEAR, and GENERAL variables were significant at the 0.001 level.

Table 1

Equation Variables

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| Lower | Upper |
| Step 1a | CLAIMS | .015 | .005 | 10.124 | 1 | .001 | 1.015 | 1.006 | 1.024 |
| GYEAR | -.529 | .057 | 86.717 | 1 | .000 | .589 | .527 | .659 |
| GENERAL | 4.733 | .235 | 407.101 | 1 | .000 | 113.641 | 71.756 | 179.973 |
| ORIGINAL | -.627 | .213 | 8.651 | 1 | .003 | .534 | .352 | .811 |
| Constant | 1054.771 | 113.408 | 86.502 | 1 | .000 | . |  |  |
| a. Variable(s) entered on step 1: CLAIMS, GYEAR, GENERAL, ORIGINAL. | | | | | | | | | |

The GYEAR and ORIGINAL variables were negatively associated with the DV as expected. Increases in these two variables were associated with a reduced probability that the patent received more than 2 citations from other patents. The GENERAL variable had the strongest association with the DV. For a one unit increase in the GENERAL variable, the patent was 113.6 times more likely to have been cited by 3 or more other patents.

The binary logistic regression analysis generated the following equation for the logit of the probability that a patent received more than 2 citations:

Logit [(CRECBINARY=2)] = 1054.771 + 0.015(CLAIMS) – 0.529(GYEAR) +

4.733(GENERAL) – 0.627(ORIGINAL)

,where CRECBINARY was coded as 1 for patents that receive 2 citations for fewer and as 2 for patents that received more than 2 citations. This can be re-written to provide the equation for determining the estimated probability that a patent received more than 2 citations as follows:

= Logit-1() =

Table 2 shows the results of the Omnibus test of the model coefficients and Table 3 shows the model summary statistics. The result of the Omnibus test indicates that including the CLAIMS, GYEAR, GENERAL, and ORIGINAL variables improved the model fit. One or more of these IVs predict the dependent variable. The p-value was less than 0.001, which was significant. The -2 Log likelihood was reduced from 2,623.109 to 1,906.867, which was a decrease of 716.242. The Nagelkerke R2 was 0.415, which indicates that 41.5 percent of the probability that a patent received more than 2 citations was explained by the IVs included in the model.

Table 2

Omnibus Test of Model Coefficients

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Chi-square | df | Sig. |
| Step 1 | Step | 716.242 | 4 | .000 |
| Block | 716.242 | 4 | .000 |
| Model | 716.242 | 4 | .000 |

Table 3

Model Summary

|  |  |  |  |
| --- | --- | --- | --- |
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 1906.867a | .306 | .415 |
| a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001. | | | |

Table 4 shows the results of the Hosmer-Lemeshow Test and Table 5 shows the contingency table for the Hosmer-Lemeshow Test. The chi-square value was 23.671, which does not seem very small. The p-value was 0.003 which is not much above the 0.001 significance level. These results suggest a lack of fit for the model.

Table 4

Hosmer-Lemeshow Test

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Chi-square | df | Sig. |
| 1 | 23.671 | 8 | .003 |

Table 5

Contingency Table for Hosmer-Lemeshow Test

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | CRECBINARY = 1.00 | | CRECBINARY = 2.00 | | Total |
| Observed | Expected | Observed | Expected |
| Step 1 | 1 | 183 | 181.295 | 13 | 14.705 | 196 |
| 2 | 177 | 175.199 | 19 | 20.801 | 196 |
| 3 | 177 | 168.176 | 19 | 27.824 | 196 |
| 4 | 158 | 159.520 | 38 | 36.480 | 196 |
| 5 | 149 | 148.155 | 47 | 47.845 | 196 |
| 6 | 130 | 133.784 | 66 | 62.216 | 196 |
| 7 | 74 | 100.801 | 122 | 95.199 | 196 |
| 8 | 70 | 64.895 | 126 | 131.105 | 196 |
| 9 | 55 | 43.237 | 141 | 152.763 | 196 |
| 10 | 25 | 22.938 | 173 | 175.062 | 198 |

**Discussion**

**Policy Implications**

The analysis provides insight into a topic that is of considerable interest to policymakers. It provides information to help both industry professionals and policymakers better understand the drivers of the technology transfer outcomes and identify possible factors that should be considered when forming public policy regarding technology transfer. The analysis suggests that considering non-transactional measures of knowledge transfer may be feasible. This could significantly affect the objectives of policymakers with regard to technology transfer. As such, this study may influence how policymakers think about technology transfer and how they formulate public policy to increase the transfer of federally-funded research to the private sector.

**Limitations of the Analysis**

As with any research project or study, this analysis has limitations. Since this analysis was focused on patent data for a five year period from 1995 to 1999, findings based on the data may not be relevant to time frames before or after this period. Additionally, there is a truncation effect in the data. Patents issued in the earliest part of the study period have the potential of receiving citations from patents over a longer period than patents issued in the latter part of the study period.

**Future Study**

There are several opportunities to improve upon and extend the analysis presented in this paper. To begin, it might prove useful to secure more recent data and to examine a subset of data buffered by at least 5 years of data on both sides of the period of study to minimize truncation effects. Removing outliers from the data may improve the goodness-of-fit of the model. Measuring only the number of dependent claims in a patent rather than all claims may help to better isolate the association between the number of claims and the number of citations received. Introducing classifications as an indication of the diversity of a technology as well as the category and subcategory of patents into the analysis to determine if the type of technology is associated with technology transfer outcomes might also be useful.

**Conclusion**

In this study, I have continued to explore an alternative conceptualization of technology transfer and an approach to measuring technology transfer based on patent citations received, which represents a non-transactional based modality of technology transfer. Using patent data, I conducted a binary logistic regression analysis to estimate the probability that a patent will receive more than 2 citations from other patents based on the year the patent was granted, the number of claims contained in the patent, and measures of the patent’s originality and generality. The resulting model indicated that the generality of a patent had the strongest association with whether or not the patent received more than 2 citations from other patents. The study results were also consistent previous analyses that indicated an inverse relationship between the year a patent was granted and the originality of a patent with the probability that the patent received more than 2 citations from other patents.

References

Appio, F. P., Martini, A., & Fantoni, G. (2017). The light and shade of knowledge recombination: Insights from a general-purpose technology. *Technological Forecasting & Social Change*, *125*, 154–165. https://doi.org/10.1016/j.techfore.2017.07.018

Bush, V. (1945). *Science, the endless frontier*. A report to the President: Washington, U.S. Government printing office, 1945.

Kirkman, D. M. (2013). Selecting University Technology Transfer Modes: An Examination of Biotechnology Firms’ Entrepreneurial Orientation. *Journal of Technology Management & Innovation, Vol 8, Iss 2, Pp 189-208 (2013)*, (2), 189. https://doi.org/10.4067/S0718-27242013000200016

Office of Management and Budget [OMB]. (2002). *The President's Management Agenda*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a394421.pdf

Office of Management and Budget [OMB]. (2018). *The President's Management Agenda*. Retrieved from https://www.whitehouse.gov/wp-content/uploads/2018/03/Presidents-Management-Agenda.pdf

Yoshikane, F. (2013). Multiple regression analysis of a patent’s citation frequency and quantitative characteristics: the case of Japanese patents. *Scientometrics*, *96*(1), 365–379. https://doi.org/10.1007/s11192-013-0953-4

Appendix A. IBM SPSS Statistics 25 Output for the Logistic Regression Analysis

**Frequencies**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 13-NOV-2018 18:52:26 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment04\Data\DataClean\Townes\_SOC6100\_Assignment04\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 2000 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| Cases Used | Statistics are based on all cases with valid data. |
| Syntax | | FREQUENCIES VARIABLES=CRECEIVE  /STATISTICS=RANGE MINIMUM MAXIMUM STDDEV MEAN MEDIAN  /FORMAT=NOTABLE  /ORDER=ANALYSIS. |
| Resources | Processor Time | 00:00:00.03 |
| Elapsed Time | 00:00:00.02 |

|  |  |  |
| --- | --- | --- |
| **Statistics** | | |
| CRECEIVE | | |
| N | Valid | 2000 |
| Missing | 0 |
| Mean | | 3.18 |
| Median | | 2.00 |
| Std. Deviation | | 4.309 |
| Range | | 111 |
| Minimum | | 1 |
| Maximum | | 112 |

**Logistic Regression**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 13-NOV-2018 18:57:25 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment04\Data\DataClean\Townes\_SOC6100\_Assignment04\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 2000 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing |
| Syntax | | LOGISTIC REGRESSION VARIABLES CRECBINARY  /METHOD=ENTER CLAIMS GYEAR GENERAL ORIGINAL  /PRINT=GOODFIT CORR ITER(1) CI(95)  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(30) CUT(0.5). |
| Resources | Processor Time | 00:00:00.06 |
| Elapsed Time | 00:00:00.06 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Case Processing Summary** | | | |
| Unweighted Casesa | | N | Percent |
| Selected Cases | Included in Analysis | 1962 | 98.1 |
| Missing Cases | 38 | 1.9 |
| Total | 2000 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 2000 | 100.0 |

|  |
| --- |
| a. If weight is in effect, see classification table for the total number of cases. |

|  |  |
| --- | --- |
| **Dependent Variable Encoding** | |
| Original Value | Internal Value |
| 1.00 | 0 |
| 2.00 | 1 |

**Block 0: Beginning Block**

|  |  |  |  |
| --- | --- | --- | --- |
| **Iteration Historya,b,c** | | | |
| Iteration | | -2 Log likelihood | Coefficients |
| Constant |
| Step 0 | 1 | 2623.135 | -.442 |
| 2 | 2623.109 | -.450 |
| 3 | 2623.109 | -.450 |

|  |
| --- |
| a. Constant is included in the model. |
| b. Initial -2 Log Likelihood: 2623.109 |
| c. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Tablea,b** | | | | | |
|  | Observed | | Predicted | | |
| CRECBINARY | | Percentage Correct |
| 1.00 | 2.00 |
| Step 0 | CRECBINARY | 1.00 | 1198 | 0 | 100.0 |
| 2.00 | 764 | 0 | .0 |
| Overall Percentage | |  |  | 61.1 |

|  |
| --- |
| a. Constant is included in the model. |
| b. The cut value is .500 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables in the Equation** | | | | | | | |
|  | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step 0 | Constant | -.450 | .046 | 94.399 | 1 | .000 | .638 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables not in the Equation** | | | | | |
|  | | | Score | df | Sig. |
| Step 0 | Variables | CLAIMS | 12.745 | 1 | .000 |
| GYEAR | 181.722 | 1 | .000 |
| GENERAL | 581.334 | 1 | .000 |
| ORIGINAL | 1.814 | 1 | .178 |
| Overall Statistics | | 656.563 | 4 | .000 |

**Block 1: Method = Enter**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Iteration Historya,b,c,d** | | | | | | | |
| Iteration | | -2 Log likelihood | Coefficients | | | | |
| Constant | CLAIMS | GYEAR | GENERAL | ORIGINAL |
| Step 1 | 1 | 1940.456 | 666.380 | .010 | -.334 | 3.807 | -.418 |
| 2 | 1907.566 | 988.799 | .014 | -.496 | 4.609 | -.593 |
| 3 | 1906.867 | 1052.933 | .015 | -.528 | 4.730 | -.626 |
| 4 | 1906.867 | 1054.769 | .015 | -.529 | 4.733 | -.627 |
| 5 | 1906.867 | 1054.771 | .015 | -.529 | 4.733 | -.627 |
| 6 | 1906.867 | 1054.771 | .015 | -.529 | 4.733 | -.627 |

|  |
| --- |
| a. Method: Enter |
| b. Constant is included in the model. |
| c. Initial -2 Log Likelihood: 2623.109 |
| d. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Omnibus Tests of Model Coefficients** | | | | |
|  | | Chi-square | df | Sig. |
| Step 1 | Step | 716.242 | 4 | .000 |
| Block | 716.242 | 4 | .000 |
| Model | 716.242 | 4 | .000 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Summary** | | | |
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 1906.867a | .306 | .415 |

|  |
| --- |
| a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Hosmer and Lemeshow Test** | | | |
| Step | Chi-square | df | Sig. |
| 1 | 23.671 | 8 | .003 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Contingency Table for Hosmer and Lemeshow Test** | | | | | | |
|  | | CRECBINARY = 1.00 | | CRECBINARY = 2.00 | | Total |
| Observed | Expected | Observed | Expected |
| Step 1 | 1 | 183 | 181.295 | 13 | 14.705 | 196 |
| 2 | 177 | 175.199 | 19 | 20.801 | 196 |
| 3 | 177 | 168.176 | 19 | 27.824 | 196 |
| 4 | 158 | 159.520 | 38 | 36.480 | 196 |
| 5 | 149 | 148.155 | 47 | 47.845 | 196 |
| 6 | 130 | 133.784 | 66 | 62.216 | 196 |
| 7 | 74 | 100.801 | 122 | 95.199 | 196 |
| 8 | 70 | 64.895 | 126 | 131.105 | 196 |
| 9 | 55 | 43.237 | 141 | 152.763 | 196 |
| 10 | 25 | 22.938 | 173 | 175.062 | 198 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Tablea** | | | | | |
|  | Observed | | Predicted | | |
| CRECBINARY | | Percentage Correct |
| 1.00 | 2.00 |
| Step 1 | CRECBINARY | 1.00 | 1018 | 180 | 85.0 |
| 2.00 | 261 | 503 | 65.8 |
| Overall Percentage | |  |  | 77.5 |

|  |
| --- |
| a. The cut value is .500 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables in the Equation** | | | | | | | | |
|  | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) |
| Lower |
| Step 1a | CLAIMS | .015 | .005 | 10.124 | 1 | .001 | 1.015 | 1.006 |
| GYEAR | -.529 | .057 | 86.717 | 1 | .000 | .589 | .527 |
| GENERAL | 4.733 | .235 | 407.101 | 1 | .000 | 113.641 | 71.756 |
| ORIGINAL | -.627 | .213 | 8.651 | 1 | .003 | .534 | .352 |
| Constant | 1054.771 | 113.408 | 86.502 | 1 | .000 | . |  |

|  |  |  |
| --- | --- | --- |
| **Variables in the Equation** | | |
|  | | 95% C.I.for EXP(B) |
| Upper |
| Step 1a | CLAIMS | 1.024 |
| GYEAR | .659 |
| GENERAL | 179.973 |
| ORIGINAL | .811 |
| Constant |  |

|  |
| --- |
| a. Variable(s) entered on step 1: CLAIMS, GYEAR, GENERAL, ORIGINAL. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Correlation Matrix** | | | | | | |
|  | | Constant | CLAIMS | GYEAR | GENERAL | ORIGINAL |
| Step 1 | Constant | 1.000 | .078 | -1.000 | .017 | .026 |
| CLAIMS | .078 | 1.000 | -.079 | .022 | -.096 |
| GYEAR | -1.000 | -.079 | 1.000 | -.018 | -.027 |
| GENERAL | .017 | .022 | -.018 | 1.000 | -.227 |
| ORIGINAL | .026 | -.096 | -.027 | -.227 | 1.000 |