Improving the Validity of Methods for Assessing Technology Transfer Results:   
Using Patent Data to Explore Alternative Measures of Technology Transfer Outcomes

Malcolm S. Townes

Saint Louis University

Declarations of interest: none.

No sponsors were involved in designing this study.

No third-party financial support was used to conduct this research or prepare this article.

Corresponding author details:

Malcolm S. Townes

7530 Wayne Avenue, University City, MO 63130-1320 USA

malcolm.townes@slu.edu

Abstract

Technology transfer can be broadly conceived as the conveyance from one person or entity to another person or entity of a capability to perform a useful task, achieve a beneficial accomplishment, or reap the benefits thereof. 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. Previous approaches to assessing technology transfer results may lack construct validity and content validity because generally they focus exclusively on transactional-based indicators. Consequently, any conclusions drawn from studies based on such approaches are likely to be inadequate at best or misleading at worst. This study used various types of regression analyses of patent data to explore alternative approaches to measuring technology transfer outcomes. The results of this study suggest that it is feasible to consider alternative measures of technology transfer that capture aspects not reflected in measures used in previous studies. The study results revealed that although the number of claims of a patent was associated with positive technology transfer outcomes as measured by the number of citations a patent receives from other patents, the strength of the relationship was much less than expected. The generality of the patent was positively associated with technology transfer outcomes and had the strongest influence by far. While the originality of a patent also had a fairly strong association with outcomes, this association was negative.

Keywords: technology transfer, university technology transfer, technology commercialization, federally funded research and development, patents, patent citations, standard multiple regression, hierarchical regression, path model analysis, logistic regression

**Introduction**

Technology transfer has been a topic of interest to the United States federal government since at least the end of World War II (Bush, 1945). While there is no official definition of technology transfer, I propose that it can be broadly conceived as the conveyance from one person or entity to another person or entity of a capability to perform a useful task, achieve a beneficial accomplishment, or reap the benefits thereof.

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).

In fiscal year 2018, the U.S. federal budget for total R&D was greater than US$142.9 billion (American Association for the Advancement of Science [AAAS], 2018a). In 2016, American universities received roughly US$32.7 billion from the federal government for research and development support (AAAS, 2018b). This amounts to nearly a quarter of the federal R&D budget. With total federal outlays of over US$3.9 trillion, the amount directed to R&D is less than 3.7 percent of total government spending (Congressional Budget Office [CBO], 2018). One might consider this trivial in the grand scheme of things but the amount is significant in absolute terms given that it is greater than the gross domestic product (GDP) of at least 110 countries (United Nations [UN], 2017). Moreover, there are other important problems of national interest to which the government could direct those dollars such as road repairs, alleviating hunger, and addressing issues with inequity in the court system. As Table 1 shows, federal R&D expenditures is equivalent to roughly 20 percent of the federal budget deficit and exceeds federal spending on transportation, the Supplemental Nutrition Assistance Program (SNAP), and law courts (U.S. Spending, n.d.).

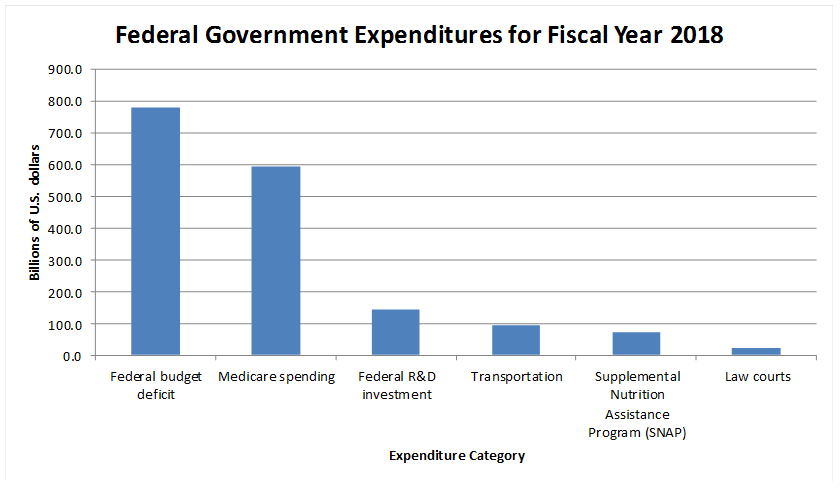


Figure 1. Federal Government Expenditures for Fiscal Year 2018

There are several potential benefits to developing predictive models to better describe technology transfer processes and understand the factors associated with positive technology transfer outcomes. Such knowledge would be useful for managing technological innovation and efficiently identifying high potential technologies for further development (Choi, Jang, Jun & Park, 2015).

However, there are important questions in the area of technology transfer that remain unanswered or underexplored including (1) how technology transfer success should be defined, (2) how technology transfer outcomes should be measured, and (3) what are the drivers of desired technology transfer outcomes.

I propose that the typical approaches to defining and measuring technology transfer may lack construct validity and content validity. Construct validity is the extent to which variables used to measure key concepts and observed outcomes of policies and programs as well as their cause-and-effect linkages adequately represents those concepts, outcomes, and linkages. Content validity is the extent to which a measure captures all the relevant aspects of a construct (McDavid, Huse, & Hawthorn, 2013).

Table 1 depicts a conceptualization of technology transfer. Most previous studies of the subject define technology transfer in terms of transactional outputs and outcomes based on the exchange of legally recognized and protected forms of intellectual property, such as license agreements executed, royalty income received, sponsored research secured, or established new business ventures. I propose that the way most previous studies define and measure technology transfer does not adequately capture what society should consider as technology transfer, which can manifest in ways beyond transactional-based outcomes. As such, any conclusions drawn from such studies are likely to be inadequate at best or misleading at worst. In this study, I seek to explore whether non-transactional indicators of knowledge transfer can be incorporated as useful measures of technology transfer outputs and outcomes to achieve increased levels of construct validity and content validity.

Table 1

Conceptualization of Technology Transfer

Asset-based

Information-based

Service-based

Patent rights

Trade secrets

Copyright

Patent documents

Patent applications

Journal articles

Presentations

Consulting services

Contract services

Sponsored research

License agreements

License income

New ventures

Acquisitions

Citations

Downloads

Readings

Event attendance

Outside interest disclosures

Contracts

Service invoices

SR agreements

Category

Embodiment

Measureable Indicators

Consider a thought experiment in which there are multiple parallel universes and in each of them a researcher uses funding from the federal government to develop a new technology that converts non-potable water into potable water inexpensively using readily available components. All aspects of each universe are the same except as follows. In universe A, the institution where the researcher works patents the technology but never licenses it. However, the technology is cited by another researcher who used insights from the first researchers work to develop a different technology for the same application that is subsequently used by 10 million households. In universe B, the institution where the research works patents the technology and licenses it to an existing business that develops a commercial product based on the technology, which is then used by 10 million households, generates $100 million in revenue for the business, and provides $1 million in royalty income for the researcher’s institution. In universe C, the researcher’s institution elects not to patent the technology. Instead the researcher publishes a peer reviewed journal article that describes the technology. The article is subsequently read by product developers at a company that use what they learn to inform the development of proprietary technology for the company. They cite the researcher’s published journal article in their patent application which is subsequently allowed by the federal government. Based on its own patented technology, the company goes on to develop a commercial product that is used by 10 million households and generates $100 million in revenue for it. In universe D, the researcher’s institution also does not patent the technology. The researcher submits a manuscript describing the technology to several journals but it is not selected for publication. The researcher decides to post the information about the technology on a website that he creates. Subsequently, 10 million households used that information to build products based on the technology for their own personal use. Table 2 summarizes the essential details of the thought experiment.

Table 2

Thought Experiment

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Universe A | Universe B | Universe C | Universe D |
| Embodiment of Knowledge | Patent application  Allowed patent | Patent application  Allowed patent | Journal article | Web post |
| License agreement executed by researcher’s institution | No | Yes | No | No |
| License income to researcher’s institution | $0.00 | $1 million | $0.00 | $0.00 |
| Research cited by others | Yes | No | Yes | No |
| Number of households that benefit from the technology | 10 million | 10 million | 10 million | 10 million |

Each universe represents a valid case of technology transfer. In each one, the same number of households benefit from the technology developed from the federally-funded research. However, based on most approaches currently used to assess technology transfer results, only the results of Universe B would be captured in the metrics because of the existence of an executed license agreement and licensing income. In some cases, the results of Universe A may be counted as successful technology transfer because of the allowed patent, which is in fact an output of the technology transfer activities but not an outcome that represents the results the activities are intended to achieve. The results of Universes C and D generally would not be captured as successful technology transfer outcomes under most current assessment approaches even though the same number of households eventually benefit from the researcher’s technology as in the other universes.

The incongruity of what is captured in the technology transfer metrics across the various universes in the thought experiment actually occurs regularly with the current approaches and methods for assessing technology transfer outcomes. As a result, the common perception among practitioners and policymakers is that the transfer of technology developed from federally-funded research is less than what probably actually occurs.

**Research Questions**

The purpose of this study is to explore how technology transfer outcomes can be more accurately defined and measured as well as better understand the relationship between technology transfer outcomes and several factors believed to be associated with technology transfer. While most other research conducted in this area conceptualize technology transfer as the transactional exchange of legally recognized and protected intellectual property through a formal contractual agreement, I consider an alternative conceptualization of technology transfer as a non-transactional exchange of knowledge. Specifically, I’m interested in the following research questions:

1. Can the construct validity and content validity of the approaches typically used for assessing technology transfer performance be improved by incorporating measures that capture other aspects of technology transfer not found in measures used in previous studies?
2. What is the nature of the relationship between independent variables thought to be associated with technology transfer and technology transfer outcomes as captured by alternative measures that reflect other aspects of technology transfer not found in measures used in previous studies?

In this study, I explored these questions by using 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 outcomes.

To conduct this study, I used various types of regression analysis to explore possible alternative non-transactional based approaches to defining and measuring technology transfer outcomes. Regression analysis is a type of multivariable technique that can be used to statistically analyze relationships among a set of factors that are distinctly measureable for observational units. In particular, regression analyses are useful for describing the extent, direction, and strength of the relationship between several independent variables and a dependent variable (Klienbaum, Kupper, Nizam, & Rosenberg, 2013). By using various types of regression analyses, I hope to demonstrate the feasibility of employing alternative non-transactional measures of technology transfer outcomes that capture aspects of the technology transfer construct not adequately reflected in the approaches of previous studies. Moreover, I hope to use various types of regression analyses to better understand the associations between various factors thought to be drivers of the technology transfer process and technology transfer outcomes as captured by the alternative non-transactional measures.

I conducted a standard multiple regression analysis to describe the extent, direction, and strength of associations between several independent variables used to measure factors believed to be associated with technology transfer outcomes and an alternative non-transactional measure of technology transfer outcomes. I conducted a hierarchical regression analysis to assess the incremental change in a base model describing the extent, direction, and strength of associations between the independent variables and an alternative measure of technology transfer outcomes when certain additional independent variables are taken into consideration. I conducted a path model analysis to investigate possible causality between the independent variables and an alternative non-transactional measure of technology transfer outcomes. Finally, I conducted a binary logistic regression analysis to investigate the relationship between the independent variables and the probability of a desired technology transfer outcome as defined by a binary alternative non-transactional measure.

**Literature Review**

**Measures of Technology Transfer Outcomes**

The topic of technology transfer has been studied quite a bit. Most of the research seems to focus on the transfer of federally-funded technologies developed at universities, so called university technology transfer (UTT). The literature on technology transfer tends to fall into a few broad categories. These include studies focused on best practices, studies focused on performance, and case studies of specific organizations or technology transfer incidents. Most of these studies define technology transfer in terms of a transactional exchange of legally recognized and protected intellectual property rights. Markman, Gianiodis, & Phan (2009) used monetization as an indicator of technology transfer performance. Their study used income received from technology transfer activities and the number of new business ventures formed to commercialize technologies as dependent variables. Kim, Anderson, & Diam (2008) used licenses executed, licensing income, and new business ventures formed to commercialize technologies as dependent variables.

Ho, Liu, Lu & Huang (2014) investigated the efficiency of UTT in different stages of the technology transfer process. Their approach included patent applications and patents allowed among the intermediate input factors in a two-stage, networked-based DEA model to study their relationships with licensing and new business venture activity.

Choi, Jang, Jun, & Park (2015) developed a predictive model of technology transfer outcomes. They did not explicitly define what constituted technology transfer but their model seems to use some measure of whether or not a patent resulted in a commercial product as an indication of technology transfer.

A review of the literature did reveal some instances where researchers seemed to conceptualize technology transfer more broadly than transactional exchanges. Anderson, Daim & Lavoie (2007) used an approach based on data envelopment analysis (DEA) to evaluate UTT productivity. Their study included patent applications and patents allowed as outputs. Ji, Lim, & Park (2016) used patent citation data to identify potential cases of technology transfer. Park, Yoon & Kim (2013) used a function-based patent analysis to identify potential opportunities to apply technologies in various fields. Sharma (2017) conducted a survey of patent citation analysis and presented a methodology for generating patent citation networks.

These are but a few examples from the literature but they are fairly representative of the approach taken in previous studies of the subject. They demonstrate how by and large previous analyses conceived of technology transfer only in terms of transactional exchanges of legally recognized and protected intellectual property.

**Variables Associated with Technology Transfer Outcomes**

Previous studies have sought to explain the relationship between various variables and technology transfer outcomes as they defined them. Markman, Gianiodis, & Phan (2009) found statistically significant positive associations with technology transfer outcomes for several variables including the size of an institution’s technology licensing office, the degree of autonomy of the institution’s technology licensing office, whether the institution was private, and the presence of a business incubator at the institution. They found statistically significant negative associations with technology transfer outcomes for other variables including the age of the technology licensing office, emphasis on sponsored research, emphasis on licensing agreements, and the level of financial incentives provided to researchers. Their study also found numerous variables that were not significantly associated with technology transfer outcomes such as the number of professional schools at an institution; whether or not the institution has a medical school; the size of the research and development budget of the institution; whether or not an institution was designated as a doctoral/research university; the institutions proximity to “high-tech” infrastructure, research capacity, and entrepreneurial activity; and the amount of venture capital activity of the local area in which the institution is located.

Cummings and Teng (2003) found that an organization’s understanding of R&D, an organization’s level of technology transfer expertise, and the range of shared basic knowledge were all associated with desirable technology transfer outcomes (as cited in Choi, Jang, Jun, & Park, 2015).

In reviewing the recent literature, I found a few studies that used path analysis or structural equation models (SEM) to examine drivers and processes of technology transfer. Vagnani & Volpe (2017) integrated SEM with a meta-analysis to study the associations between various attributes of innovation, manager behaviors, and organizational decisions regarding the adoption of innovation. They found that innovation attributes have indirect effects on the innovation adoption decisions of organizations via the behavioral preferences of managers. Raut, Priyadarshinee, Garda & Jha (2018) incorporated SEM in a hybrid three-stage approach to analyze factors that influence the adoption of cloud computing technology by private organizations in India. Yan & Yu (2016) used a path-based method to examine the structure of time-dependent, discipline-level citation networks.

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 positively associated with citation frequency.

**Data and Methods**

**Data Sources**

This study used 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 (National Bureau of Economic Research, 2018).

Table 3 details the original USPTO variables of the source data and explanations of their meanings (Hall, Jaffe, & Trajtenberg, 2001). 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 3

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 4 provides information about the source data constructed variables and explanations of their meanings (Hall, Jaffe, & Trajtenberg, 2001). 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 4

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**

I used RStudio to create a subset of the raw data. After importing the raw data, I filtered it for grant years between and including 1995 through 1999. I then filtered that data for patents that had at least one (1) claim. I subsequently filtered that data for patents with at least one (1) citation received. This generated a subset of 253,328 observations.

I inspected this intermediate data set using the miss\_var\_summary function to check for missing data by variables to ensure that there was no missing data in the CRECEIVE and CLAIMS variables. I then used the miss\_case\_summary function to determine how many observations had missing data in the other variables. I used the get\_dupes function to verify that there were no duplicates in the PATENT variable thus ensuring that it could be used as a unique identifier. I then checked for duplicates across all variables to ensure that each observation was unique.

I used the sample function to select a random sample of 2,000 observations from the subset of 253,328 observations using the set.seed function with a seed value of 1972. I then saved this sample data as a .csv file.

**Standard multiple regression analysis.**

For the standard multiple regression analysis, I did not make any other modifications to the data beyond the initial data selection and cleaning. I used CRECEIVE as the dependent variable (DV) of interest.

**Hierarchical regression analysis.**

Based on the results of the standard multiple regression analysis, I made several modifications to the model. I removed the following variables because of high multicollinearity with other variables: APPYEAR, BCKGTLAG, FWDAPLAG, SELFCTLB, and SECDLWBD. Additionally, I used a log transformation on the original DV (CRECEIVE) because the data for this variable was very skewed to the right (i.e., positively skewed) based on a visual inspection of a histogram for the data. The transformed DV is logCRECEIVE, which is what I used in the hierarchical regression analysis.

Finally, I created a product term called CLAIMSORIGINAL using the CLAIMS variable and ORIGINAL variable to test for possible interaction between these two variables. To create the logCRECEIVE and CLAIMSORIGINAL variable, I used RStudio. I imported the cleaned data subset of 2,000 observations and used the mutate function to calculate values of logCRECEIVE as the logarithm of the CRECEIVE variable. I also used the mutate function to calculate values of CLAIMSORIGNAL as the product of the CLAIMS and ORIGINAL variables. I then saved this modified data subset as a .csv file.

**Path model analysis.**

For the path model analysis, I only used the GYEAR, CLAIMS, CRECEIVE, RATIOCIT, GENERAL, and ORIGINAL variables. Based on the results of the standard multiple regression and hierarchical regression analyses, I made several modifications to the data and incorporated several previous observations into the initial theoretical path analysis model. I removed the following variables because of high multicollinearity: APPYEAR, BCKGTLAG, FWDAPLAG, SELFCTLB, and SECDLWBD. Based on a scatter plot of the CRECEIVE variable against the CLAIMS variable, I removed observations with CLAIMS greater than 90 claims and CRECEIVE greater than 40 citations received as outliers. This resulted in 42 outlier observations being removed from the analysis for a final sample count of 1,958 observations. Additionally, I created a new variable (CRECEIVEln) using the Transform > Compute Variable function of SPSS Statistics 25. The CRECEIVEln variable is the natural logarithm transformation of the CRECEIVE variable. I performed a transformation on the CRECEIVE variable because the data was skewed to the right (i.e., positively skewed) based on a visual inspection of a histogram. I chose a natural logarithm transformation because it appeared to bring out potential linear relationships between the CRECEIVE variable and primary independent variable (IV) of interest (CLAIMS) better than other transformations that I considered, which included base 10 logarithm and reciprocal transformations. The CRECEIVEln variable is what I used as the DV of interest in the path model analysis.

**Binary logistic regression analysis.**

For the binary logistic regression analysis, I only used the GYEAR, CLAIMS, CRECEIVE, GENERAL, and ORIGINAL variables in the analysis. Based on the results of standard multiple regression, hierarchical regression, and path model analyses, I made several modifications to the data and incorporated several previous observations into the logistic regression model. I removed the APPYEAR, BCKGTLAG, FWDAPLAG, SELFCTLB, and SECDLWBD variables because of high multicollinearity. I created a new variable called CRECBINARY, which I used as the 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 and 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 5 lists the final variables that I used in the binary logistic regression analysis. All of the IVs used in the analysis were continuous interval or ratio variables.

Table 5

Variables Used in Binary Logistic Regression 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. |

**Theoretical Perspective and Anticipated Results**

**Standard Multiple Regression Analysis**

In the standard multiple regression analysis, I used patents as a proxy for units of technology. I proposed that there is a positive linear relationship between the transfer of a technology as measured by the numbers of citations received by a patent and measures of the patent’s originality, generality, citation lag time, application year, and grant year. I theorized that certain IVs are associated with the DV as follows:

CRECEIVE = β0 + β1(GRYEAR) + β2(APPYEAR) + β3(CLAIMS) + β4(GENERAL) + β5(ORIGINAL) + β6(BCKGTLAG) + β7(FWDAPLAG) + β8(SELFCTUB) +

Β9(SELFCTLB) + β10(SECDLWBD) + β11(SECDUPBD)+ ε

The null hypothesis was that the number of citations a patent receives is NOT related the IVs. That is, the regression coefficient for all IVs is equal to zero:

H0: β1 = β2 … = β11 = 0

The alternative hypothesis was that the number of citations a patent receives is related to at least one of the IVs. Stated another way, the regression coefficient for at least one IV is NOT equal to zero:

HA: β ≠ 0 for at least one independent variable

**Hierarchical Regression Analysis**

I used IBM SPSS Statistics 25 to perform the hierarchical regression analysis. I verified that the correct data type and data variable type was applied to each variable in the variable view tab. I assigned the logCRECEIVE as the DV.

For each model I used the enter method. The first model in the hierarchical regression analysis included the CLAIMS, CMADE, GENERAL, GYEAR, ORIGINAL, and RATIOCIT variables as IVs. The claims of a patent define the scope of the subject that it asserts to be novel, nonobvious, and useful. I theorized that the more claims that are included in a patent (CLAIMS) the more opportunities it has to be cited another patents. I suspected that the number of citations made by a patent (CMADE) would indicate its relevance to the field of interest and thus the more citations made the higher the likelihood it will be relevant to other technologies in the field. I surmised that the more general the nature of a patent (GENERAL) the more likely it is to be relevant to a wide variety of fields. I expected the older a patent was (GYEAR) the less likely it was to be relevant to other technologies because of half-life effects. Machlup (1962) is credited with proposing the concept of half-life of knowledge, which can be thought of as the time it takes for half of the knowledge in a field to be rendered irrelevant. Machlup proposed that although patents provided a number of years of protection and exclusivity (17 years at the time) obsolescence reduced the practical duration of this protection to no more than a few years.

Based on the results of the previous standard multiple regression analysis, I suspected that patents that rate high on originality (ORIGINAL) would probably not be relevant to technologies because they likely make use of new paradigms that are incompatible with the popular approaches to innovation within a field at that point in time. I included the ratio of citations made by all patents granted since 1963 to the total number of citations made by a particular patent (RATIOCIT) in the model because I theorized that this is an indication of overall technology transfer activity. I expected that the greater the overall level of technology transfer activity, the more likely that any given patent will be cited by other patents.

In addition to the IVs included in the first model, the second model included the SECDUPBD, SELFCTUB, and CLAIMSORIGINAL variables to test whether the addition of these variables improved the fit of the regression model. The SECDUPBD and SELFCTUB variables account for the number of citations received and made by a patent, respectively, to other patents with the same assignee (typically an organization). I anticipated there is an incentive for an organization to leverage technologies assigned to it as much as possible, thus increasing the chances of technology transfer within the organization and thus the number of citations a patent receives. I included the CLAIMSORIGINAL product term because I suspected interaction effects between the CLAIMS variable and the ORIGINAL variable. I suspected that the relationship between the DV and the CLAIMS variable varies as a function of the ORIGINAL variable. I surmised that the higher the originality of a technology as represented by a patent, the more interest that it may receive from other innovators and thus the more likely that the patent will be cited in the patents of other innovators. If there is a CLAIMS by ORIGINAL interaction, then these two variables do not operate independently of one another in their influence on the DV and therefore cannot be considered in isolation of one another.

**Path Model Analysis**

Figure 2 shows the theoretical path model that I developed from logical consideration of the relationships among the variables. The theoretical path model uses the CRECEIVEln variable as the primary DV of interest. Sub-model 1 used GENERAL as the DV and ORIGINAL as the IV. It is likely that the originality of a patent will influence whether or not it will have broad applicability. The more original the patent, the more likely that other innovators in various fields will identify applications of the technology over time. Patents that rank low in originality are likely to be specific or specialized to a narrower range of applications within closely related fields.



Figure 2. Logically Derived Theoretical Path Model

Sub-model 2 used CLAIMS as the DV and ORIGINAL, GENERAL, GYEAR, and RATIOCIT as the IVs. The claims of a patent define the scope of the subject that it asserts to be novel, nonobvious, and useful. I proposed that patents that rank higher in originality are likely to generate more claims because they stake out new innovation territory. Patents the rank higher in generality are likely to generate more claims because the scope of their applicability. In general, patents are likely to have more claims as the grant year increases because of the temporal nature of advances in sciences and the cumulative effects of scientific knowledge. I suspected that as the ratio of the number of citations made by all patents granted since 1960 to the total number of citations made by a particular patent increases the number of claims for a patent will increase because of the general increase in scientific knowledge.

Sub-model 3 used CRECEIVEln as the DV and ORIGINAL, GENERAL, CLAIMS, GYEAR, and RATIOCIT as the IVs. Previous analyses indicated an inverse relationship between the originality of a patent and the number of citations it received. This may be because the full capabilities of highly original patents are less readily apparent than patents that rank lower on originality. Patents that rank high in generality may receive higher numbers of citations because the broader scope of their applicability creates more opportunities to be cited. Likewise, patents that have more claims probably have more opportunities to be cited than patents with fewer claims. I suspected that patents in general are likely to receive more citations over time because scientific knowledge accumulates and spreads over time. I suspected that as the ratio of the number of citations made by all patents granted since 1960 to the total number of citations made by a particular patent increases the number of citations that patent receives will likely increase because of network effects.

**Logistic Regression Analysis**

I used logistic regression analysis to investigate whether patents that are likely to perform better than average as measured by patent citations could be discriminated from all other patents. I anticipated that the probability that a patent would receive more than the median number of citations would be associated with specific patent data variables including 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 theorized 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 anticipated that the association for the number of claims made by a patent was slight based on the results of the standard multiple regression, hierarchical regression, and path model analyses. While the number of claims made by a patent will have a positive association, more claims does not necessarily create significantly more opportunities for that patent to be cited possibly 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 suspected that the more general a patent, the more opportunities it has to be cited by other patents across multiple fields.

Based on the results of the standard multiple regression, hierarchical regression, and path model analyses, I expected 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 suspected that the older a patent is the less relevant it becomes possibly 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.

**Analysis and Results**

**Standard Multiple Regression Analysis**

For the standard multiple regression, I analyzed the sample data using IBM SPSS Statistics 25. The complete output file for the analysis is shown in Appendix A. I began by verifying that the correct data type and variable type was applied to each variable in the variable view tab. For this study, I assigned CRECEIVE as the DV. I prepared descriptive statistics for the DV using the Analyze > Descriptive Statistics > Frequencies function. The specific statistics calculated for the DV included mean, standard error of mean, median, mode, minimum, maximum, and range, standard deviation, variance, skewness, and kurtosis. I created a histogram with the normal distribution curve superimposed to visually examine the data.

I then used the Graphs > Chart Builder function to create a scatter plot of the DV against the CLAIMS variable, which is one IV of interest. I later used the scatter plot to visually examine whether there was a potential relationship between the two variables.

Finally, I used the Analyze > Regression > Linear function to conduct a multiple regression analysis. The CRECEIVE variable remained the DV. The IVs included GRYEAR, APPYEAR, CLAIMS, GENERAL, ORIGINAL, BCKGTLAG, FWDAPLAG, SELFCTUB, SELFCTLB, SECDLWBD, and SECDUPBD. Although I suspected that the variables NCLASS, CAT, and SUBCAT were all associated the number of citations a patent receives, I specifically excluded them from the analysis because they are nominal variables and would thus violate the assumptions of the linear regression. The options I selected for the linear regression statistics included model fit, R squared change, and part and partial correlations; estimates, a confidence level of 95 percent, and covariance matrix for the regression coefficients; and Durbin-Watson and casewise diagnotics for outliers beyond 3 standard deviations for the residuals.

Table 6 shows descriptive statistics for the dependent variable CRECEIVE. The values of the DV ranged from 1 to 112 citations received. The mean value for the DV was 3.18 while the median value was 2.00 suggesting that the sample distribution is skewed. The standard deviation was 4.309, which seems relatively wide.

Table 6

Descriptive Statistics for CRECEIVE

|  |  |  |
| --- | --- | --- |
| N | Valid | 2000 |
| Missing | 0 |
| Mean | | 3.18 |
| Std. Error of Mean | | .096 |
| Median | | 2.00 |
| Mode | | 1 |
| Std. Deviation | | 4.309 |
| Variance | | 18.567 |
| Skewness | | 10.292 |
| Std. Error of Skewness | | .055 |
| Kurtosis | | 214.022 |
| Std. Error of Kurtosis | | .109 |
| Range | | 111 |
| Minimum | | 1 |
| Maximum | | 112 |

Figure 3 shows a histogram of the DV. Visual inspection of the data confirms that the sample distribution is in fact positively skewed to the right.

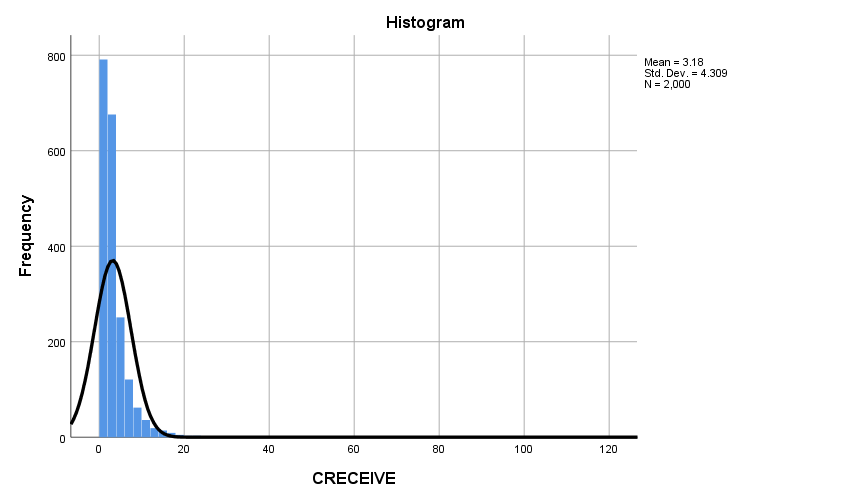


Figure 3. Histogram of CRECEIVE Values

Figure 4 shows a scatter plot of the DV against the independent variable CLAIMS. I chose this particular IV because conceptually one might expect a positive linear relationship between the number of citations a patent receives and the number of claims in the patent. More patent claims might be indicative of broad applicability of the technology or simply more opportunities for the patent to be cited. However, this does not appear to be the case. I discerned no obvious pattern upon visual inspection of the scatter plot.

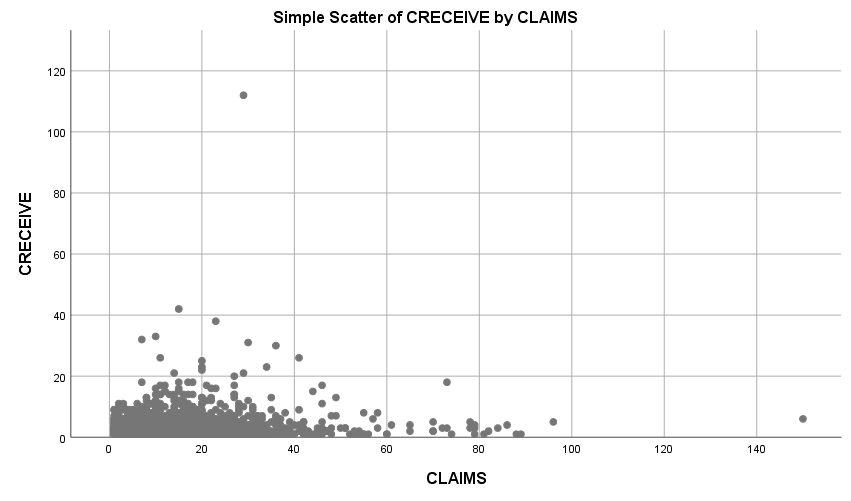


Figure 4. Scatter Plot of CRECEIVE Against CLAIMS

Table 7 shows the model summary for the regression. The value of the correlation coefficient, R, is 0.443 suggesting the amount of change in the DV (CRECEIVE) is only moderately determined by the IVs included in the model. The Adjusted R-Square value indicates that 19.1 percent of the variation in the number of citations a patent receives can be explained by the IVs included in the model.

Table 7

Regression Model Summary

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | | Durbin-Watson |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .443a | .196 | .191 | 4.098 | .196 | 36.647 | 11 | 1655 | .000 | 1.970 |
| a. Predictors: (Constant), SELFCTUB, CLAIMS, GYEAR, ORIGINAL, BCKGTLAG, SECDUPBD, GENERAL, FWDAPLAG, APPYEAR, SELFCTLB, SECDLWBD | | | | | | | | | | |
| b. Dependent Variable: CRECEIVE | | | | | | | | | | |

Table 8 shows the analysis of variance (ANOVA) results. The significance value of the regression is 0.000, which is less than 0.05 to which the confidence level was set. This suggests that the probability that the variation in the DV variable explained by the IVs in the model is due to chance is low. The F statistic of 33.617 indicates that at least one variable in the regression was significant.

Table 8

ANOVA Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 6769.213 | 11 | 615.383 | 36.647 | .000b |
| Residual | 27790.817 | 1655 | 16.792 |  |  |
| Total | 34560.030 | 1666 |  |  |  |
| a. Dependent Variable: CRECEIVE | | | | | | |
| b. Predictors: (Constant), SELFCTUB, CLAIMS, GYEAR, ORIGINAL, BCKGTLAG, SECDUPBD, GENERAL, FWDAPLAG, APPYEAR, SELFCTLB, SECDLWBD | | | | | | |

Based on the results of the analysis, I rejected the null hypothesis because at least one of the regression coefficients was not equal to zero and statistically significant. This suggests that the fit of the model was improved by including at least one of the IVs compared with a model containing only the Y-intercept and no IVs.

Table 9 shows the results for the regression coefficients as well as collinearity statistics. The regression produced the following equation:

CRECEIVE = 1509.628 – 0.281(APPYEAR) – 0.064(BCKGTLAG) + 0.027(CLAIMS) – 0.034(FWDAPLAG) + 5.826(GENERAL) – 0.474(GYEAR) – 1.001(ORIGINAL) – 3.184(SECDLWBD) + 3.09(SECDUPBD) – 1.653(SELFCTLB) + 1.188(SELCTUB)

The GENERAL, SECDLWBD, and SECDUPBD variables had the greatest impact on the value of the DV. Somewhat surprisingly, the ORIGINAL variable, which is a measure of the originality of the patent, had a smaller impact on the value of the DV. Moreover, it was inversely related to the value of the DV, which was not initially expected. The significance value for the variables FWDAPLG, SECDLWBD, SECDUPBD, SEFLCTLB, and SELFCTUB were all greater than 0.05, which indicates that these variables were not significant. While the significance value for the CLAIMS variable was less than 0.05, the impact of this IV on the DV was smaller than expected. This seems to confirm what was observed in the scatter plot shown in Figure 4 above.

Table 9

Regression Coefficients and Collinearity Statistics

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | | Unstandardized Coefficients | | Stdz. Coef. | t | Sig. | 95.0% Confidence Interval for B | | Correlations | | | Collinearity Statistics | |
| B | Std. Error | Beta | Lower Bound | Upper Bound | Zero-order | Partial | Part | Tol. | VIF |
|  | (Constant) | 1509.628 | 254.147 |  | 5.940 | .000 | 1011.146 | 2008.111 |  |  |  |  |  |
| APPYEAR | -.281 | .167 | -.078 | -1.683 | .092 | -.608 | .046 | -.239 | -.041 | -.037 | .226 | 4.417 |
| BCKGTLAG | -.064 | .013 | -.114 | -4.805 | .000 | -.091 | -.038 | -.124 | -.117 | -.106 | .865 | 1.157 |
| CLAIMS | .027 | .008 | .074 | 3.319 | .001 | .011 | .043 | .090 | .081 | .073 | .976 | 1.025 |
| FWDAPLAG | -.034 | .141 | -.008 | -.244 | .807 | -.311 | .242 | .143 | -.006 | -.005 | .409 | 2.447 |
| GENERAL | 5.826 | .408 | .336 | 14.264 | .000 | 5.024 | 6.627 | .385 | .331 | .314 | .877 | 1.140 |
| GYEAR | -.474 | .147 | -.112 | -3.232 | .001 | -.762 | -.187 | -.243 | -.079 | -.071 | .406 | 2.461 |
| ORIGINAL | -1.001 | .374 | -.061 | -2.676 | .008 | -1.735 | -.267 | .007 | -.066 | -.059 | .938 | 1.066 |
| SECDLWBD | -3.184 | 2.330 | -.234 | -1.367 | .172 | -7.753 | 1.385 | -.039 | -.034 | -.030 | .017 | 60.138 |
| SECDUPBD | 3.049 | 2.274 | .229 | 1.341 | .180 | -1.411 | 7.509 | -.033 | .033 | .030 | .017 | 59.861 |
| SELFCTLB | -1.653 | 2.138 | -.078 | -.773 | .440 | -5.847 | 2.541 | -.007 | -.019 | -.017 | .048 | 21.015 |
| SELFCTUB | 1.188 | 1.968 | .060 | .604 | .546 | -2.672 | 5.047 | -.010 | .015 | .013 | .049 | 20.547 |
| a. Dependent Variable: CRECEIVE | | | | | | | | | | | | | |

The results suggest that the variables APPYEAR, FWDAPLAG, and GYEAR have a high degree of multicollinearity based on the collinearity statistic tolerance value. Likewise, the variables SECDLWBD and SECDUPBD have a high degree of multicollinearity. The same is true of the SELFCTLB and SELFCTUB variables. This suggests that these groups of variables are likely measuring the same characteristic.

Table 10 shows the casewise diagnostics for the regression. There were 16 patents in which the DV had values beyond 3 standard deviations. Case number 230 was the most extreme having a value that was greater than 25 standard deviations from the mean. Ten other observations had values that were greater than 4 standard deviations from the mean. This is likely related to the skewness in the sample distribution.

Table 10

Casewise Diagnostics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Case Number | Std. Residual | CRECEIVE | Predicted Value | Residual |
| 30 | 3.911 | 23 | 6.98 | 16.024 |
| 153 | 6.131 | 32 | 6.88 | 25.125 |
| 202 | 8.840 | 42 | 5.78 | 36.224 |
| 230 | 25.144 | 112 | 8.96 | 103.036 |
| 243 | 4.366 | 23 | 5.11 | 17.892 |
| 405 | 3.015 | 18 | 5.65 | 12.355 |
| 446 | 4.977 | 25 | 4.61 | 20.393 |
| 457 | 6.329 | 31 | 5.07 | 25.934 |
| 585 | 3.081 | 20 | 7.37 | 12.626 |
| 759 | 3.309 | 21 | 7.44 | 13.560 |
| 879 | 4.839 | 26 | 6.17 | 19.830 |
| 1063 | 8.040 | 38 | 5.05 | 32.947 |
| 1116 | 4.651 | 22 | 2.94 | 19.059 |
| 1220 | 6.530 | 33 | 6.24 | 26.759 |
| 1680 | 5.189 | 26 | 4.74 | 21.264 |
| 1786 | 3.470 | 21 | 6.78 | 14.219 |
| a. Dependent Variable: CRECEIVE | | | | |

**Hierarchical Regression Analysis**

Table 11 below shows the results of the hierarchical regression analysis of the patent data. I used IBM SPSS Statistics 25 to perform the hierarchical regression. I used the Analyze > Regression > Linear function to conduct a hierarchical regression analysis. The options I selected included model fit, R square change, part and partial correlations, and collinearity diagnostics for the regression statistics; estimates, a confidence level of 95 percent, and covariance matrix for the regression coefficients; and Durbin-Watson, casewise diagnostics for outliers beyond 3 standard deviations for the residuals. The complete output file for the analysis is shown in Appendix B.

Table 11

Results of hierarchical regression analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Unstandardized | | | | Standardized | | | |
| Variable | **Model 1** |  | **Model 2** |  | **Model 1** |  | **Model 2** |  |
| Constant | 330.669 | \*\*\* | 331.411 | \*\*\* | 0.000 | \*\*\* | 0.000 | \*\*\* |
| CLAIMS | 0.005 | \*\*\* | 0.008 | \*\*\* | 0.077 | \*\*\* | 0.127 | \*\*\* |
| CMADE | 0.001 |  | 0.001 |  | 0.021 |  | 0.022 |  |
| GENERAL | 1.744 | \*\*\* | 1.752 | \*\*\* | 0.569 | \*\*\* | 0.572 | \*\*\* |
| GYEAR | -0.166 | \*\*\* | -0.166 | \*\*\* | -0.221 | \*\*\* | -0.222 | \*\*\* |
| ORIGINAL | -0.301 | \*\*\* | -0.195 | \* | -0.104 | \*\*\* | -0.067 | \* |
| RATIOCIT | 0.555 | \*\*\* | 0.557 | \*\*\* | 0.081 | \*\*\* | 0.082 | \*\*\* |
| SECDUPBD |  |  | 0.053 |  |  |  | 0.023 |  |
| SELFCTUB |  |  | -0.033 |  |  |  | -0.010 |  |
| CLAIMSORIGINAL |  |  | -0.007 |  |  |  | -0.073 |  |
|  |  |  |  |  |  |  |  |  |
| Adj. R2 | 0.436 |  | 0.437 |  | 0.436 |  | 0.437 |  |
| Adj. R2 change | 0.436 |  | 0.001 |  | 0.436 |  | 0.001 |  |
| F | 215.741 | \* | 144.442 | \* | 215.741 | \* | 144.442 | \* |
| F change | 215.741 | \* | 1.474 |  | 215.741 | \* | 1.474 |  |
| Δ F | 215.741 |  | 71.299 |  | 215.741 |  | 71.299 |  |

\* p < 0.05; \*\* p < 0.01; \*\*\* p < 0.001

The multiple regression model using logCRECEIVE as the DV and CLAIMS, CMADE, GENERAL, GYEAR, ORIGINAL, RATIOCIT as IVs (Model 1) showed a good model fit (F=215.741, p<0.05) indicating that at least one of the IVs was significant. The adjusted R2 for the model was 0.436 indicating that the IVs in the model explained 43.6 percent of the DV.

The multiple regression model using logCRECEIVE as the DV and the IVs from Model 1 plus SECDUPBD, SELFCTUB, and CLAIMSORIGINAL as additional IVs (Model 2) also showed a good model fit (F=144.442, p<0.05). However, the addition of the SECDUPBD, SELFCTUB, and CLAIMSORIGINAL variables did not incrementally improve the model to any significant extent. The adjusted R2 for Model 2 was 0.437 indicating the IVs in the model explained 43.7 percent of the DV. None of the additional variables included in Model 2 were significant (i.e., p>0.05 for all three variables).

**Path Model Analysis**

I used IBM SPSS Statistics 25 to analyze the theoretical path model. I used the Analyze > Regression > Linear function to prepare regression analyses for each sub-model. The options I selected included model fit, R square change, part and partial correlations, and collinearity diagnostics for the regression statistics; estimates, a confidence level of 95 percent, and covariance matrix for the regression coefficients; and Durbin-Watson, casewise diagnostics for outliers beyond 3 standard deviations for the residuals. For each model I used the enter method. The complete SPSS Statistics 25 output file for the analysis is shown in Appendix C.

Figure 5 shows theoretical path model with standardized coefficients and p-values from the various regression analyses.



Figure 5. Path Model with Standardized Coefficients and P-Values

The associations between GENERAL and ORIGINAL and between CLAIMS and ORIGINAL, GENERAL, and GYEAR were all significant. GENERAL, ORIGINAL, and GYEAR all had indirect effects on CRECEIVEln through CLAIMS. I initially assumed that RATIOCIT would have an indirect effect on CRECEIVEln through the CLAIMS variable, but the p-value of the association was just above the threshold for significance at the 0.05 level. Interestingly, GYEAR had an inverse relationship with CRECEIVEln, which was counter to my original assumption about the relationship between these two variables. However, this may have been due to truncation effects in the data (Hall, Jaffe, & Trajtenberg, 2001).

Figure 6 shows the final path model with standardized coefficients and p-values. I removed the RATIOCIT variable from the final model because it technically was not significant and doing so simplified the model without greatly decreasing the R2 value. The p-value was 0.055, which was just above the threshold for significance at the 0.05 level. After removing the RATIOCIT variable, I re-calculated the linear regression for CLAIMS as the DV with ORIGINAL, GENERAL, and GYEAR as the IVs. However, the standardized coefficients changed only slightly and did not change the indirect effects calculations for these variables.

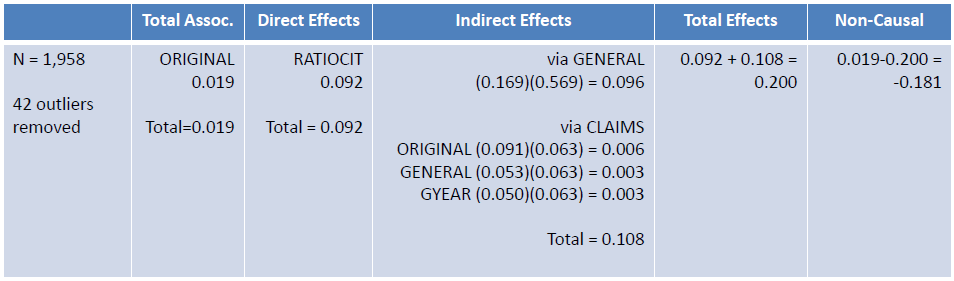


Figure 6. Final Path Model with Standardized Coefficients and P-Values

Table 12 summarizes the direct, indirect, and total effects for the final path model. Total effects were 0.200 while non-causal effects were -0.181. The indirect effects on CRECEIVEln through the CLAIMS variables were relatively miniscule.

Table 12

Path Model Direct and Indirect Effects



**Binary Logistic Regression Analysis**

I used IBM SPSS Statistics 25 to perform the binary logistic regression analysis. 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 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 D.

Table 13 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 13

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 based on the results of the standard multiple regression, hierarchical regression, and path model analyses. 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)

CRECBINARY was coded as 1 for patents that received 2 citations for fewer and as 2 for patents that received more than 2 citations. This equation 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 14 shows the results of the Omnibus test of the model coefficients and Table 15 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 14

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 15

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 16 shows the results of the Hosmer-Lemeshow Test and Table 17 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. However, it’s important to note that the Hosmer-Lemeshow Test is sensitive to the number of groupings used in the analysis as well as the sample size used (Lai & Liu, 2018). This may account for the discrepancy between the results of the Hosemer-Lemeshow Test and the Omnibus Test.

Table 16

Hosmer-Lemeshow Test

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

Table 17

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**

**Findings**

The results of this study suggest that it is feasible to consider alternative measures of technology transfer that capture aspects not reflected in measures used in previous studies. Doing so could significantly improve the construct validity and content validity of approaches used to assess technology transfer performance.

The study provides additional insight about the nature of the relationship between technology transfer outcomes and factors believed to drive the technology transfer process. While the number of claims of a patent was positively associated with desired technology transfer outcomes as measured by the number of citations a patent receives from other patents, the strength of the relationship was much less than initially expected. The study demonstrated that by far the generality of the patent had the strongest positive association with technology transfer outcomes. The study revealed that the originality of a patent also had a fairly strong association with outcomes but this association was negative, which was somewhat surprising. Likewise, it showed that the year a patent was granted also had a negative association with the number of citations it received from other patents, but this may be due to truncation effects in the data (Hall, Jaffe & Trajtenberg, 2001).

**Policy Implications**

The analysis provides insight into a topic that is of considerable interest to policymakers. It provides information to help both industry practitioners and policymakers better understand the drivers of the technology transfer outcomes and identify possible factors that should be considered when developing best practices and forming public policy regarding technology transfer. The analysis suggests that it may be feasible to consider alternative measures of technology transfer that capture aspects not reflected in measures used in previous studies. This could significantly affect how technology transfer performance is currently perceived as well as the objectives of policymakers with regard to technology transfer policy. 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 the study 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 (Hall, Jaffe & Trajtenber, 2001). 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. 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 study period to minimize truncation effects. Removing outliers from the data may improve the model. Measuring only the number of independent 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 broadness of applicability 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. Additionally, including technology readiness level (TRL) as an independent variable may provide useful insights about what drives technology transfer outcomes. The TRL scale provides an indication of the level of development of a technology. Repeating the standard multiple regression and hierarchical regression analyses using the natural logarithm of the number of citations receive as the dependent variable would provide a common basis of comparison for the various types of analyses. Repeating the binary logistic regression using a mean split instead of a median split might prove more insightful given the positive skew in the data. Also, comparing various path models could help identify the strongest causal linkages between the independent variables used in the model.

**Conclusion**

In this study, I have explored 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 various regression analyses to describe the nature of the relationship between various independent variables and alternative measures of technology transfer outcomes that reflect aspects not captured in measures used in previous studies. The study results suggest that it may be feasible to consider alternative measures of technology transfer that capture aspects not reflected in measures used in previous approaches. It also provides additional insights into the nature of the relationship between independent variables thought to be associated with technology transfer and technology transfer outcomes as captured by the alternative measures.

References

American Association for the Advancement of Science [AAAS]. (2018a). *Defense, Nondefense, and Total R&D, 1976-2018* [Data file]. Retrieved from https://www.aaas.org/page/historical-trends-federal-rd

American Association for the Advancement of Science [AAAS]. (2018b). *Federal Support for Universities by Agency, 1990-2016 (Obligations)* [Data file]. Retrieved from https://www.aaas.org/page/rd-colleges-and-universities

Anderson, T. R., Daim, T. U., & Lavoie, F. F. (2007). Measuring the efficiency of university technology transfer. *Technovation*, 27(5), 306-318. doi:10.1016/j.technovation.2006.10.003

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.

Choi, J., Jang, D., Jun, S., & Park, S. (2015). A Predictive Model of Technology Transfer Using Patent Analysis. *Sustainability* (2071-1050), 7(12), 16175

Congressional Budget Office [CBO]. (2018). Historical Budget Data [Data file]. *The Budget and Economic Outlook: 2018 to 2028* Retrieved from https://www.cbo.gov/about/products/budget-economic-data#2

Hall, B. H., Jaffe, A. B. and Trajtenberg, M. (2001). "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." *NBER Working Paper 8498*. Retrieved from http://www.nber.org/patents/

Ho, M. H. C., Liu, J. S., Huang, C. C., & Lu, W. M. (2014). A new perspective to explore the technology transfer efficiencies in US universities. *Journal of Technology Transfer*, 39(2), 247-275. doi:10.1007/s10961-013-9298-7

Ji, I., Lim, H., & Park, T.-Y. (2016). Exploring Potential Users of Patents for Technology Transfer: Utilizing Patent Citation Data. *Procedia Computer Science*, 91, 211–220. https://doi.org/10.1016/j.procs.2016.07.059

Kim, J., Anderson, T., & Daim, T. (2008). Assessing university technology transfer: A measure of efficiency patterns. *International Journal of Innovation and Technology Management, 5*(4), 495-526. doi:10.1142/S0219877008001497

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

Kleinbaum, D. G., Kupper, L. L., Nizam, A., & Rosenberg, E. S. (2013). *Applied regression analysis and other multivariable methods*. Boston, MA : Cengage Learning.

Lai, X., & Liu, L. (2018). A simple test procedure in standardizing the power of Hosmer-Lemeshow test in large data sets. *Journal of Statistical Computation & Simulation*, *88*(13), 2463–2472. https://doi.org/10.1080/00949655.2018.1467912

Machlup, F. (1962). *The production and distribution of knowledge in the United States*. Princeton, NJ: Princeton University Press.

Markman, G. D., Gianiodis, P. T., & Phan, P. H. (2009). Supply-Side Innovation and Technology Commercialization. *Journal of Management Studies*, 46(4), 625-649. doi:10.1111/j.1467-6486.2009.00835.x

McDavid, J.C., Huse, I., & Hawthron, L.R.L. (2013). *Program evaluation and performance measurement: An introduction to practice* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

National Bureau of Economic Research. (2018). Patent data, including constructed variables [data file]. Retrieved from http://www.nber.org/patents/

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

Park, H., Yoon, J., & Kim, K. (2013). Using function-based patent analysis to identify potential application areas of technology for technology transfer. *Expert Systems With Applications*, *40*, 5260–5265. https://doi.org/10.1016/j.eswa.2013.03.033

Raut, R. D., Priyadarshinee, P., Gardas, B. B., & Jha, M. K. (2018). Analyzing the factors influencing cloud computing adoption using three stage hybrid SEM-ANN-ISM (SEANIS) approach. *Technological Forecasting & Social Change*, 134, 98–123. https://doi.org/10.1016/j.techfore.2018.05.020

Sharma, P., & Tripathi, R. C. (2017). Patent citation: A technique for measuring the knowledge flow of information and innovation. *World Patent Information*, 51, 31–42. https://doi.org/10.1016/j.wpi.2017.11.002

United Nations. (2017). *GDP and its breakdown at current prices in U.S. dollars* [Data file]. Retrieved from https://unstats.un.org/unsd/snaama/dnllist.asp

U.S. Spending. (n.d.) USGovernmentSpending.com. Retrieved from https://www.usgovernmentspending.com/year\_spending\_2018USbn\_20bs2n\_4041\_605#usgs302

Vagnani, G., & Volpe, L. (2017). Innovation attributes and managers’ decisions about the adoption of innovations in organizations: A meta-analytical review. *International Journal of Innovation Studies*, 1, 107–133. https://doi.org/10.1016/j.ijis.2017.10.001

Yan, E., & Yu, Q. (2016). Using path-based approaches to examine the dynamic structure of discipline-level citation networks: 1997-2011. *Journal of the Association for Information Science & Technology*, 67(8), 1943–1955. https://doi.org/10.1002/asi.23516

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 25 Output for Standard Multiple Regression Analysis

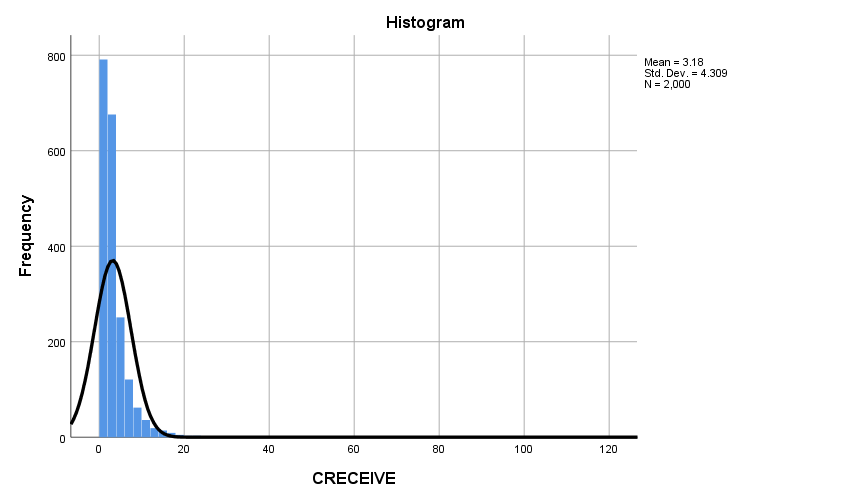
**Frequencies**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 20-SEP-2018 10:36:53 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment01\Results\Results1995to1999\Townes\_SOC6100\_Assignment01\_Data\_Sample.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=STDDEV VARIANCE RANGE MINIMUM MAXIMUM SEMEAN MEAN MEDIAN MODE SKEWNESS SESKEW  KURTOSIS SEKURT  /HISTOGRAM NORMAL  /ORDER=ANALYSIS. |
| Resources | Processor Time | 00:00:03.32 |
| Elapsed Time | 00:00:01.47 |

[DataSet1] D:\SOC6100\Assignments\Assignment01\Results\Results1995to1999\Townes\_SOC6100\_Assignment01\_Data\_Sample.sav

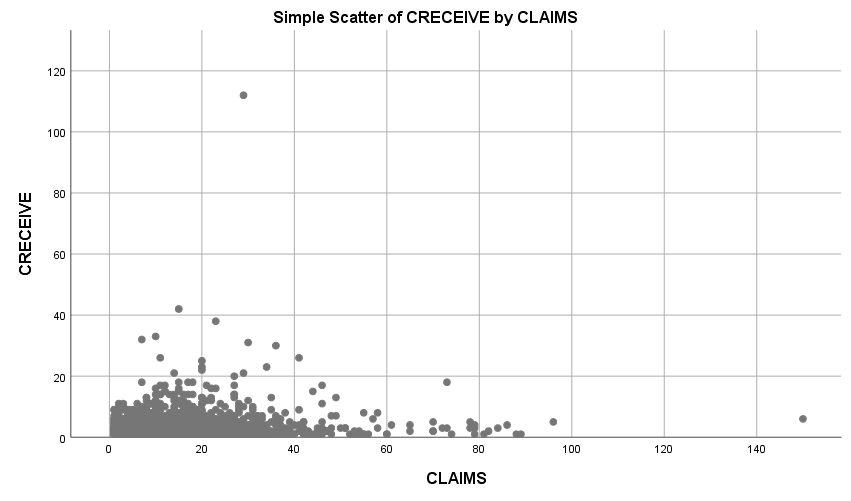
|  |  |  |
| --- | --- | --- |
| **Statistics** | | |
| CRECEIVE | | |
| N | Valid | 2000 |
| Missing | 0 |
| Mean | | 3.18 |
| Std. Error of Mean | | .096 |
| Median | | 2.00 |
| Mode | | 1 |
| Std. Deviation | | 4.309 |
| Variance | | 18.567 |
| Skewness | | 10.292 |
| Std. Error of Skewness | | .055 |
| Kurtosis | | 214.022 |
| Std. Error of Kurtosis | | .109 |
| Range | | 111 |
| Minimum | | 1 |
| Maximum | | 112 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CRECEIVE** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 1 | 791 | 39.6 | 39.6 | 39.6 |
| 2 | 433 | 21.7 | 21.7 | 61.2 |
| 3 | 243 | 12.2 | 12.2 | 73.4 |
| 4 | 146 | 7.3 | 7.3 | 80.7 |
| 5 | 105 | 5.3 | 5.3 | 85.9 |
| 6 | 75 | 3.8 | 3.8 | 89.6 |
| 7 | 46 | 2.3 | 2.3 | 92.0 |
| 8 | 36 | 1.8 | 1.8 | 93.8 |
| 9 | 26 | 1.3 | 1.3 | 95.1 |
| 10 | 22 | 1.1 | 1.1 | 96.2 |
| 11 | 14 | .7 | .7 | 96.9 |
| 12 | 10 | .5 | .5 | 97.4 |
| 13 | 9 | .4 | .4 | 97.8 |
| 14 | 10 | .5 | .5 | 98.3 |
| 15 | 4 | .2 | .2 | 98.5 |
| 16 | 4 | .2 | .2 | 98.7 |
| 17 | 5 | .3 | .3 | 99.0 |
| 18 | 5 | .3 | .3 | 99.2 |
| 20 | 1 | .1 | .1 | 99.3 |
| 21 | 2 | .1 | .1 | 99.4 |
| 22 | 1 | .1 | .1 | 99.4 |
| 23 | 2 | .1 | .1 | 99.5 |
| 25 | 1 | .1 | .1 | 99.6 |
| 26 | 2 | .1 | .1 | 99.7 |
| 30 | 1 | .1 | .1 | 99.7 |
| 31 | 1 | .1 | .1 | 99.8 |
| 32 | 1 | .1 | .1 | 99.8 |
| 33 | 1 | .1 | .1 | 99.9 |
| 38 | 1 | .1 | .1 | 99.9 |
| 42 | 1 | .1 | .1 | 100.0 |
| 112 | 1 | .1 | .1 | 100.0 |
| Total | 2000 | 100.0 | 100.0 |  |



**GGraph**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 20-SEP-2018 10:39:40 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment01\Results\Results1995to1999\Townes\_SOC6100\_Assignment01\_Data\_Sample.sav |
| Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 2000 |
| Syntax | | GGRAPH  /GRAPHDATASET NAME="graphdataset" VARIABLES=CLAIMS CRECEIVE MISSING=LISTWISE REPORTMISSING=NO  /GRAPHSPEC SOURCE=INLINE  /FITLINE TOTAL=NO.  BEGIN GPL  SOURCE: s=userSource(id("graphdataset"))  DATA: CLAIMS=col(source(s), name("CLAIMS"))  DATA: CRECEIVE=col(source(s), name("CRECEIVE"))  GUIDE: axis(dim(1), label("CLAIMS"))  GUIDE: axis(dim(2), label("CRECEIVE"))  GUIDE: text.title(label("Simple Scatter of CRECEIVE by CLAIMS"))  ELEMENT: point(position(CLAIMS\*CRECEIVE))  END GPL. |
| Resources | Processor Time | 00:00:00.87 |
| Elapsed Time | 00:00:00.45 |



**Regression**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 19-SEP-2018 20:40:56 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment01\Results\Results1995to1999\Townes\_SOC6100\_Assignment01\_Data\_Sample.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 cases with no missing values for any variable used. |
| Syntax | | REGRESSION  /MISSING LISTWISE  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL CHANGE ZPP  /CRITERIA=PIN(.05) POUT(.10)  /NOORIGIN  /DEPENDENT CRECEIVE  /METHOD=ENTER APPYEAR BCKGTLAG CLAIMS FWDAPLAG GENERAL GYEAR ORIGINAL SECDLWBD SECDUPBD SELFCTLB  SELFCTUB  /RESIDUALS DURBIN  /CASEWISE PLOT(ZRESID) OUTLIERS(3). |
| Resources | Processor Time | 00:00:00.02 |
| Elapsed Time | 00:00:00.01 |
| Memory Required | 10800 bytes |
| Additional Memory Required for Residual Plots | 0 bytes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | SELFCTUB, CLAIMS, GYEAR, ORIGINAL, BCKGTLAG, SECDUPBD, GENERAL, FWDAPLAG, APPYEAR, SELFCTLB, SECDLWBDb | . | Enter |

|  |
| --- |
| a. Dependent Variable: CRECEIVE |
| b. All requested variables entered. |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | |
| R Square Change | F Change | df1 |
| 1 | .443a | .196 | .191 | 4.098 | .196 | 36.647 | 11 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Summaryb** | | | |
| Model | Change Statistics | | |
| df2 | Sig. F Change |  |
| 1 | 1655 | .000 | 1.970 |

|  |
| --- |
| a. Predictors: (Constant), SELFCTUB, CLAIMS, GYEAR, ORIGINAL, BCKGTLAG, SECDUPBD, GENERAL, FWDAPLAG, APPYEAR, SELFCTLB, SECDLWBD |
| b. Dependent Variable: CRECEIVE |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 6769.213 | 11 | 615.383 | 36.647 | .000b |
| Residual | 27790.817 | 1655 | 16.792 |  |  |
| Total | 34560.030 | 1666 |  |  |  |

|  |
| --- |
| a. Dependent Variable: CRECEIVE |
| b. Predictors: (Constant), SELFCTUB, CLAIMS, GYEAR, ORIGINAL, BCKGTLAG, SECDUPBD, GENERAL, FWDAPLAG, APPYEAR, SELFCTLB, SECDLWBD |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 1509.628 | 254.147 |  | 5.940 | .000 |
| APPYEAR | -.281 | .167 | -.078 | -1.683 | .092 |
| BCKGTLAG | -.064 | .013 | -.114 | -4.805 | .000 |
| CLAIMS | .027 | .008 | .074 | 3.319 | .001 |
| FWDAPLAG | -.034 | .141 | -.008 | -.244 | .807 |
| GENERAL | 5.826 | .408 | .336 | 14.264 | .000 |
| GYEAR | -.474 | .147 | -.112 | -3.232 | .001 |
| ORIGINAL | -1.001 | .374 | -.061 | -2.676 | .008 |
| SECDLWBD | -3.184 | 2.330 | -.234 | -1.367 | .172 |
| SECDUPBD | 3.049 | 2.274 | .229 | 1.341 | .180 |
| SELFCTLB | -1.653 | 2.138 | -.078 | -.773 | .440 |
| SELFCTUB | 1.188 | 1.968 | .060 | .604 | .546 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | |
| Model | | 95.0% Confidence Interval for B | | Correlations | | | Collinearity Statistics |
| Lower Bound | Upper Bound | Zero-order | Partial | Part | Tolerance |
| 1 | (Constant) | 1011.146 | 2008.111 |  |  |  |  |
| APPYEAR | -.608 | .046 | -.239 | -.041 | -.037 | .226 |
| BCKGTLAG | -.091 | -.038 | -.124 | -.117 | -.106 | .865 |
| CLAIMS | .011 | .043 | .090 | .081 | .073 | .976 |
| FWDAPLAG | -.311 | .242 | .143 | -.006 | -.005 | .409 |
| GENERAL | 5.024 | 6.627 | .385 | .331 | .314 | .877 |
| GYEAR | -.762 | -.187 | -.243 | -.079 | -.071 | .406 |
| ORIGINAL | -1.735 | -.267 | .007 | -.066 | -.059 | .938 |
| SECDLWBD | -7.753 | 1.385 | -.039 | -.034 | -.030 | .017 |
| SECDUPBD | -1.411 | 7.509 | -.033 | .033 | .030 | .017 |
| SELFCTLB | -5.847 | 2.541 | -.007 | -.019 | -.017 | .048 |
| SELFCTUB | -2.672 | 5.047 | -.010 | .015 | .013 | .049 |

|  |  |  |
| --- | --- | --- |
| **Coefficientsa** | | |
| Model | | Collinearity Statistics |
| VIF |
| 1 | (Constant) |  |
| APPYEAR | 4.417 |
| BCKGTLAG | 1.157 |
| CLAIMS | 1.025 |
| FWDAPLAG | 2.447 |
| GENERAL | 1.140 |
| GYEAR | 2.461 |
| ORIGINAL | 1.066 |
| SECDLWBD | 60.138 |
| SECDUPBD | 59.861 |
| SELFCTLB | 21.015 |
| SELFCTUB | 20.547 |

|  |
| --- |
| a. Dependent Variable: CRECEIVE |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | | | | | |
| Model | | | SELFCTUB | CLAIMS | GYEAR | ORIGINAL | BCKGTLAG |
| 1 | Correlations | SELFCTUB | 1.000 | .013 | -.021 | .011 | -.257 |
| CLAIMS | .013 | 1.000 | -.089 | -.088 | .003 |
| GYEAR | -.021 | -.089 | 1.000 | -.091 | .072 |
| ORIGINAL | .011 | -.088 | -.091 | 1.000 | -.080 |
| BCKGTLAG | -.257 | .003 | .072 | -.080 | 1.000 |
| SECDUPBD | -.038 | -.031 | .053 | .029 | -.102 |
| GENERAL | -.019 | -.043 | .095 | -.194 | .114 |
| FWDAPLAG | .005 | -.012 | -.237 | .019 | -.101 |
| APPYEAR | .009 | .057 | -.677 | .057 | -.064 |
| SELFCTLB | -.974 | -.014 | .033 | -.003 | .288 |
| SECDLWBD | .038 | .030 | -.051 | -.028 | .100 |
| Covariances | SELFCTUB | 3.872 | .000 | -.006 | .008 | -.007 |
| CLAIMS | .000 | 6.632E-5 | .000 | .000 | 3.690E-7 |
| GYEAR | -.006 | .000 | .022 | -.005 | .000 |
| ORIGINAL | .008 | .000 | -.005 | .140 | .000 |
| BCKGTLAG | -.007 | 3.690E-7 | .000 | .000 | .000 |
| SECDUPBD | -.171 | -.001 | .018 | .024 | -.003 |
| GENERAL | -.015 | .000 | .006 | -.030 | .001 |
| FWDAPLAG | .001 | -1.401E-5 | -.005 | .001 | .000 |
| APPYEAR | .003 | 7.754E-5 | -.017 | .004 | .000 |
| SELFCTLB | -4.097 | .000 | .010 | -.003 | .008 |
| SECDLWBD | .176 | .001 | -.017 | -.025 | .003 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | | | | |
| Model | | | SECDUPBD | GENERAL | FWDAPLAG | APPYEAR |
| 1 | Correlations | SELFCTUB | -.038 | -.019 | .005 | .009 |
| CLAIMS | -.031 | -.043 | -.012 | .057 |
| GYEAR | .053 | .095 | -.237 | -.677 |
| ORIGINAL | .029 | -.194 | .019 | .057 |
| BCKGTLAG | -.102 | .114 | -.101 | -.064 |
| SECDUPBD | 1.000 | -.023 | -.037 | -.013 |
| GENERAL | -.023 | 1.000 | .041 | .093 |
| FWDAPLAG | -.037 | .041 | 1.000 | .688 |
| APPYEAR | -.013 | .093 | .688 | 1.000 |
| SELFCTLB | .041 | .023 | -.008 | -.017 |
| SECDLWBD | -.991 | .028 | .042 | .010 |
| Covariances | SELFCTUB | -.171 | -.015 | .001 | .003 |
| CLAIMS | -.001 | .000 | -1.401E-5 | 7.754E-5 |
| GYEAR | .018 | .006 | -.005 | -.017 |
| ORIGINAL | .024 | -.030 | .001 | .004 |
| BCKGTLAG | -.003 | .001 | .000 | .000 |
| SECDUPBD | 5.170 | -.022 | -.012 | -.005 |
| GENERAL | -.022 | .167 | .002 | .006 |
| FWDAPLAG | -.012 | .002 | .020 | .016 |
| APPYEAR | -.005 | .006 | .016 | .028 |
| SELFCTLB | .201 | .020 | -.003 | -.006 |
| SECDLWBD | -5.250 | .027 | .014 | .004 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | | |
| Model | | | SELFCTLB | SECDLWBD |
| 1 | Correlations | SELFCTUB | -.974 | .038 |
| CLAIMS | -.014 | .030 |
| GYEAR | .033 | -.051 |
| ORIGINAL | -.003 | -.028 |
| BCKGTLAG | .288 | .100 |
| SECDUPBD | .041 | -.991 |
| GENERAL | .023 | .028 |
| FWDAPLAG | -.008 | .042 |
| APPYEAR | -.017 | .010 |
| SELFCTLB | 1.000 | -.048 |
| SECDLWBD | -.048 | 1.000 |
| Covariances | SELFCTUB | -4.097 | .176 |
| CLAIMS | .000 | .001 |
| GYEAR | .010 | -.017 |
| ORIGINAL | -.003 | -.025 |
| BCKGTLAG | .008 | .003 |
| SECDUPBD | .201 | -5.250 |
| GENERAL | .020 | .027 |
| FWDAPLAG | -.003 | .014 |
| APPYEAR | -.006 | .004 |
| SELFCTLB | 4.571 | -.237 |
| SECDLWBD | -.237 | 5.427 |

|  |
| --- |
| a. Dependent Variable: CRECEIVE |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
| (Constant) | APPYEAR | BCKGTLAG | CLAIMS |
| 1 | 1 | 7.500 | 1.000 | .00 | .00 | .00 | .00 |
| 2 | 1.607 | 2.161 | .00 | .00 | .01 | .01 |
| 3 | 1.189 | 2.512 | .00 | .00 | .00 | .00 |
| 4 | .598 | 3.540 | .00 | .00 | .08 | .00 |
| 5 | .380 | 4.444 | .00 | .00 | .23 | .70 |
| 6 | .290 | 5.083 | .00 | .00 | .14 | .12 |
| 7 | .259 | 5.381 | .00 | .00 | .42 | .13 |
| 8 | .153 | 7.009 | .00 | .00 | .01 | .02 |
| 9 | .018 | 20.291 | .00 | .00 | .09 | .00 |
| 10 | .006 | 34.531 | .00 | .00 | .01 | .00 |
| 11 | 1.112E-7 | 8214.054 | .86 | .03 | .00 | .00 |
| 12 | 5.913E-8 | 11262.073 | .14 | .97 | .00 | .00 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Variance Proportions | | | | | |
| FWDAPLAG | GENERAL | GYEAR | ORIGINAL | SECDLWBD | SECDUPBD |
| 1 | 1 | .00 | .00 | .00 | .00 | .00 | .00 |
| 2 | .00 | .01 | .00 | .01 | .00 | .00 |
| 3 | .00 | .00 | .00 | .00 | .00 | .00 |
| 4 | .00 | .74 | .00 | .00 | .00 | .00 |
| 5 | .00 | .04 | .00 | .00 | .00 | .00 |
| 6 | .01 | .10 | .00 | .76 | .00 | .00 |
| 7 | .10 | .04 | .00 | .12 | .00 | .00 |
| 8 | .32 | .01 | .00 | .10 | .00 | .00 |
| 9 | .00 | .00 | .00 | .00 | .00 | .00 |
| 10 | .00 | .00 | .00 | .00 | .99 | .99 |
| 11 | .18 | .05 | .37 | .00 | .00 | .00 |
| 12 | .39 | .00 | .63 | .00 | .00 | .00 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | |
| Model | Dimension | Variance Proportions | |
| SELFCTLB | SELFCTUB |
| 1 | 1 | .00 | .00 |
| 2 | .00 | .00 |
| 3 | .01 | .01 |
| 4 | .00 | .00 |
| 5 | .00 | .00 |
| 6 | .00 | .00 |
| 7 | .00 | .00 |
| 8 | .00 | .00 |
| 9 | .98 | .98 |
| 10 | .00 | .00 |
| 11 | .00 | .00 |
| 12 | .00 | .00 |

|  |
| --- |
| a. Dependent Variable: CRECEIVE |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Casewise Diagnosticsa** | | | | |
| Case Number | Std. Residual | CRECEIVE | Predicted Value | Residual |
| 30 | 3.911 | 23 | 6.98 | 16.024 |
| 153 | 6.131 | 32 | 6.88 | 25.125 |
| 202 | 8.840 | 42 | 5.78 | 36.224 |
| 230 | 25.144 | 112 | 8.96 | 103.036 |
| 243 | 4.366 | 23 | 5.11 | 17.892 |
| 405 | 3.015 | 18 | 5.65 | 12.355 |
| 446 | 4.977 | 25 | 4.61 | 20.393 |
| 457 | 6.329 | 31 | 5.07 | 25.934 |
| 585 | 3.081 | 20 | 7.37 | 12.626 |
| 759 | 3.309 | 21 | 7.44 | 13.560 |
| 879 | 4.839 | 26 | 6.17 | 19.830 |
| 1063 | 8.040 | 38 | 5.05 | 32.947 |
| 1116 | 4.651 | 22 | 2.94 | 19.059 |
| 1220 | 6.530 | 33 | 6.24 | 26.759 |
| 1680 | 5.189 | 26 | 4.74 | 21.264 |
| 1786 | 3.470 | 21 | 6.78 | 14.219 |

|  |
| --- |
| a. Dependent Variable: CRECEIVE |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | -2.11 | 9.70 | 3.30 | 2.016 | 1667 |
| Residual | -5.320 | 103.036 | .000 | 4.084 | 1667 |
| Std. Predicted Value | -2.682 | 3.175 | .000 | 1.000 | 1667 |
| Std. Residual | -1.298 | 25.144 | .000 | .997 | 1667 |

|  |
| --- |
| a. Dependent Variable: CRECEIVE |

Appendix B. IBM Statistics SPSS 25 Output for Hierarchical Regression Analysis

GET FILE='D:\SOC6100\Assignments\Assignment02\Results\Townes\_SOC6100\_Assignment02\_Data.sav'.

DATASET NAME DataSet1 WINDOW=FRONT.

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/MISSING LISTWISE

/STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL CHANGE ZPP

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT logCRECEIVE

/METHOD=ENTER CLAIMS CMADE GENERAL GYEAR ORIGINAL RATIOCIT

/METHOD=ENTER CLAIMSORIGINAL SECDUPBD SELFCTUB

/RESIDUALS DURBIN

/CASEWISE PLOT(ZRESID) OUTLIERS(3).

**Regression**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 10-OCT-2018 20:23:46 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment02\Results\Townes\_SOC6100\_Assignment02\_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 cases with no missing values for any variable used. |
| Syntax | | REGRESSION  /DESCRIPTIVES MEAN STDDEV CORR SIG N  /MISSING LISTWISE  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL CHANGE ZPP  /CRITERIA=PIN(.05) POUT(.10)  /NOORIGIN  /DEPENDENT logCRECEIVE  /METHOD=ENTER CLAIMS CMADE GENERAL GYEAR ORIGINAL RATIOCIT  /METHOD=ENTER CLAIMSORIGINAL SECDUPBD SELFCTUB  /RESIDUALS DURBIN  /CASEWISE PLOT(ZRESID) OUTLIERS(3). |
| Resources | Processor Time | 00:00:00.03 |
| Elapsed Time | 00:00:00.06 |
| Memory Required | 9024 bytes |
| Additional Memory Required for Residual Plots | 0 bytes |

[DataSet1] D:\SOC6100\Assignments\Assignment02\Results\Townes\_SOC6100\_Assignment02\_Data.sav

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptive Statistics** | | | |
|  | Mean | Std. Deviation | N |
| logCRECEIVE | .803931581920302 | .804113290481782 | 1667 |
| CLAIMS | 15.52 | 12.480 | 1667 |
| CMADE | 10.76 | 13.775 | 1667 |
| GENERAL | .201574 | .2624779 | 1667 |
| GYEAR | 1996.28 | 1.073 | 1667 |
| ORIGINAL | .396933 | .2769866 | 1667 |
| RATIOCIT | .955843 | .1180271 | 1667 |
| CLAIMSORIGINAL | 6.525935 | 8.5131005 | 1667 |
| SECDUPBD | .20 | .342 | 1667 |
| SELFCTUB | .14 | .231 | 1667 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Correlations** | | | | | | |
|  | | logCRECEIVE | CLAIMS | CMADE | GENERAL | GYEAR |
| Pearson Correlation | logCRECEIVE | 1.000 | .100 | .018 | .612 | -.350 |
| CLAIMS | .100 | 1.000 | .167 | .063 | .030 |
| CMADE | .018 | .167 | 1.000 | .067 | .072 |
| GENERAL | .612 | .063 | .067 | 1.000 | -.242 |
| GYEAR | -.350 | .030 | .072 | -.242 | 1.000 |
| ORIGINAL | .018 | .106 | .288 | .191 | .017 |
| RATIOCIT | .077 | .012 | -.104 | .031 | .080 |
| CLAIMSORIGINAL | .077 | .750 | .286 | .161 | .029 |
| SECDUPBD | -.015 | .006 | .013 | -.058 | .040 |
| SELFCTUB | -.006 | -.007 | -.056 | -.024 | -.012 |
| Sig. (1-tailed) | logCRECEIVE | . | .000 | .232 | .000 | .000 |
| CLAIMS | .000 | . | .000 | .005 | .114 |
| CMADE | .232 | .000 | . | .003 | .002 |
| GENERAL | .000 | .005 | .003 | . | .000 |
| GYEAR | .000 | .114 | .002 | .000 | . |
| ORIGINAL | .233 | .000 | .000 | .000 | .239 |
| RATIOCIT | .001 | .317 | .000 | .105 | .001 |
| CLAIMSORIGINAL | .001 | .000 | .000 | .000 | .118 |
| SECDUPBD | .270 | .401 | .303 | .009 | .051 |
| SELFCTUB | .400 | .385 | .012 | .168 | .312 |
| N | logCRECEIVE | 1667 | 1667 | 1667 | 1667 | 1667 |
| CLAIMS | 1667 | 1667 | 1667 | 1667 | 1667 |
| CMADE | 1667 | 1667 | 1667 | 1667 | 1667 |
| GENERAL | 1667 | 1667 | 1667 | 1667 | 1667 |
| GYEAR | 1667 | 1667 | 1667 | 1667 | 1667 |
| ORIGINAL | 1667 | 1667 | 1667 | 1667 | 1667 |
| RATIOCIT | 1667 | 1667 | 1667 | 1667 | 1667 |
| CLAIMSORIGINAL | 1667 | 1667 | 1667 | 1667 | 1667 |
| SECDUPBD | 1667 | 1667 | 1667 | 1667 | 1667 |
| SELFCTUB | 1667 | 1667 | 1667 | 1667 | 1667 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Correlations** | | | | | |
|  | | ORIGINAL | RATIOCIT | CLAIMSORIGINAL | SECDUPBD |
| Pearson Correlation | logCRECEIVE | .018 | .077 | .077 | -.015 |
| CLAIMS | .106 | .012 | .750 | .006 |
| CMADE | .288 | -.104 | .286 | .013 |
| GENERAL | .191 | .031 | .161 | -.058 |
| GYEAR | .017 | .080 | .029 | .040 |
| ORIGINAL | 1.000 | .028 | .580 | -.025 |
| RATIOCIT | .028 | 1.000 | .027 | .032 |
| CLAIMSORIGINAL | .580 | .027 | 1.000 | -.013 |
| SECDUPBD | -.025 | .032 | -.013 | 1.000 |
| SELFCTUB | -.055 | .060 | -.045 | .198 |
| Sig. (1-tailed) | logCRECEIVE | .233 | .001 | .001 | .270 |
| CLAIMS | .000 | .317 | .000 | .401 |
| CMADE | .000 | .000 | .000 | .303 |
| GENERAL | .000 | .105 | .000 | .009 |
| GYEAR | .239 | .001 | .118 | .051 |
| ORIGINAL | . | .123 | .000 | .155 |
| RATIOCIT | .123 | . | .138 | .096 |
| CLAIMSORIGINAL | .000 | .138 | . | .304 |
| SECDUPBD | .155 | .096 | .304 | . |
| SELFCTUB | .012 | .008 | .032 | .000 |
| N | logCRECEIVE | 1667 | 1667 | 1667 | 1667 |
| CLAIMS | 1667 | 1667 | 1667 | 1667 |
| CMADE | 1667 | 1667 | 1667 | 1667 |
| GENERAL | 1667 | 1667 | 1667 | 1667 |
| GYEAR | 1667 | 1667 | 1667 | 1667 |
| ORIGINAL | 1667 | 1667 | 1667 | 1667 |
| RATIOCIT | 1667 | 1667 | 1667 | 1667 |
| CLAIMSORIGINAL | 1667 | 1667 | 1667 | 1667 |
| SECDUPBD | 1667 | 1667 | 1667 | 1667 |
| SELFCTUB | 1667 | 1667 | 1667 | 1667 |

|  |  |  |
| --- | --- | --- |
| **Correlations** | | |
|  | | SELFCTUB |
| Pearson Correlation | logCRECEIVE | -.006 |
| CLAIMS | -.007 |
| CMADE | -.056 |
| GENERAL | -.024 |
| GYEAR | -.012 |
| ORIGINAL | -.055 |
| RATIOCIT | .060 |
| CLAIMSORIGINAL | -.045 |
| SECDUPBD | .198 |
| SELFCTUB | 1.000 |
| Sig. (1-tailed) | logCRECEIVE | .400 |
| CLAIMS | .385 |
| CMADE | .012 |
| GENERAL | .168 |
| GYEAR | .312 |
| ORIGINAL | .012 |
| RATIOCIT | .008 |
| CLAIMSORIGINAL | .032 |
| SECDUPBD | .000 |
| SELFCTUB | . |
| N | logCRECEIVE | 1667 |
| CLAIMS | 1667 |
| CMADE | 1667 |
| GENERAL | 1667 |
| GYEAR | 1667 |
| ORIGINAL | 1667 |
| RATIOCIT | 1667 |
| CLAIMSORIGINAL | 1667 |
| SECDUPBD | 1667 |
| SELFCTUB | 1667 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | RATIOCIT, CLAIMS, GENERAL, CMADE, GYEAR, ORIGINALb | . | Enter |
| 2 | SECDUPBD, SELFCTUB, CLAIMSORIGINALb | . | Enter |

|  |
| --- |
| a. Dependent Variable: logCRECEIVE |
| b. All requested variables entered. |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryc** | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | |
| R Square Change | F Change | df1 |
| 1 | .662a | .438 | .436 | .603832813033063 | .438 | 215.741 | 6 |
| 2 | .663b | .440 | .437 | .603574533062267 | .001 | 1.474 | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Summaryc** | | | |
| Model | Change Statistics | | |
| df2 | Sig. F Change |  |
| 1 | 1660 | .000 |  |
| 2 | 1657 | .220 | 1.991 |

|  |
| --- |
| a. Predictors: (Constant), RATIOCIT, CLAIMS, GENERAL, CMADE, GYEAR, ORIGINAL |
| b. Predictors: (Constant), RATIOCIT, CLAIMS, GENERAL, CMADE, GYEAR, ORIGINAL, SECDUPBD, SELFCTUB, CLAIMSORIGINAL |
| c. Dependent Variable: logCRECEIVE |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 471.973 | 6 | 78.662 | 215.741 | .000b |
| Residual | 605.259 | 1660 | .365 |  |  |
| Total | 1077.233 | 1666 |  |  |  |
| 2 | Regression | 473.584 | 9 | 52.620 | 144.442 | .000c |
| Residual | 603.649 | 1657 | .364 |  |  |
| Total | 1077.233 | 1666 |  |  |  |

|  |
| --- |
| a. Dependent Variable: logCRECEIVE |
| b. Predictors: (Constant), RATIOCIT, CLAIMS, GENERAL, CMADE, GYEAR, ORIGINAL |
| c. Predictors: (Constant), RATIOCIT, CLAIMS, GENERAL, CMADE, GYEAR, ORIGINAL, SECDUPBD, SELFCTUB, CLAIMSORIGINAL |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 330.699 | 28.660 |  | 11.539 | .000 |
| CLAIMS | .005 | .001 | .077 | 4.096 | .000 |
| CMADE | .001 | .001 | .021 | 1.079 | .281 |
| GENERAL | 1.744 | .059 | .569 | 29.316 | .000 |
| GYEAR | -.166 | .014 | -.221 | -11.536 | .000 |
| ORIGINAL | -.301 | .057 | -.104 | -5.285 | .000 |
| RATIOCIT | .555 | .127 | .081 | 4.371 | .000 |
| 2 | (Constant) | 331.411 | 28.663 |  | 11.562 | .000 |
| CLAIMS | .008 | .002 | .127 | 3.623 | .000 |
| CMADE | .001 | .001 | .022 | 1.135 | .257 |
| GENERAL | 1.752 | .060 | .572 | 29.392 | .000 |
| GYEAR | -.166 | .014 | -.222 | -11.562 | .000 |
| ORIGINAL | -.195 | .084 | -.067 | -2.326 | .020 |
| RATIOCIT | .557 | .127 | .082 | 4.378 | .000 |
| CLAIMSORIGINAL | -.007 | .004 | -.073 | -1.703 | .089 |
| SECDUPBD | .053 | .044 | .023 | 1.204 | .229 |
| SELFCTUB | -.033 | .066 | -.010 | -.510 | .610 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | 95.0% Confidence Interval for B | | Correlations | | |
| Lower Bound | Upper Bound | Zero-order | Partial | Part |
| 1 | (Constant) | 274.485 | 386.914 |  |  |  |
| CLAIMS | .003 | .007 | .100 | .100 | .075 |
| CMADE | -.001 | .003 | .018 | .026 | .020 |
| GENERAL | 1.627 | 1.861 | .612 | .584 | .539 |
| GYEAR | -.194 | -.138 | -.350 | -.272 | -.212 |
| ORIGINAL | -.412 | -.189 | .018 | -.129 | -.097 |
| RATIOCIT | .306 | .804 | .077 | .107 | .080 |
| 2 | (Constant) | 275.191 | 387.631 |  |  |  |
| CLAIMS | .004 | .013 | .100 | .089 | .067 |
| CMADE | -.001 | .004 | .018 | .028 | .021 |
| GENERAL | 1.635 | 1.868 | .612 | .585 | .541 |
| GYEAR | -.194 | -.138 | -.350 | -.273 | -.213 |
| ORIGINAL | -.360 | -.031 | .018 | -.057 | -.043 |
| RATIOCIT | .307 | .806 | .077 | .107 | .081 |
| CLAIMSORIGINAL | -.015 | .001 | .077 | -.042 | -.031 |
| SECDUPBD | -.034 | .140 | -.015 | .030 | .022 |
| SELFCTUB | -.162 | .095 | -.006 | -.013 | -.009 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Coefficientsa** | | | |
| Model | | Collinearity Statistics | |
| Tolerance | VIF |
| 1 | (Constant) |  |  |
| CLAIMS | .965 | 1.036 |
| CMADE | .879 | 1.137 |
| GENERAL | .898 | 1.114 |
| GYEAR | .922 | 1.085 |
| ORIGINAL | .881 | 1.135 |
| RATIOCIT | .975 | 1.026 |
| 2 | (Constant) |  |  |
| CLAIMS | .275 | 3.643 |
| CMADE | .875 | 1.143 |
| GENERAL | .894 | 1.119 |
| GYEAR | .921 | 1.086 |
| ORIGINAL | .405 | 2.470 |
| RATIOCIT | .971 | 1.030 |
| CLAIMSORIGINAL | .183 | 5.458 |
| SECDUPBD | .956 | 1.046 |
| SELFCTUB | .953 | 1.050 |

|  |
| --- |
| a. Dependent Variable: logCRECEIVE |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Excluded Variablesa** | | | | | | | |
| Model | | Beta In | t | Sig. | Partial Correlation | Collinearity Statistics | |
| Tolerance | VIF |
| 1 | CLAIMSORIGINAL | -.073b | -1.702 | .089 | -.042 | .183 | 5.454 |
| SECDUPBD | .021b | 1.140 | .254 | .028 | .994 | 1.006 |
| SELFCTUB | -.004b | -.234 | .815 | -.006 | .992 | 1.009 |

|  |  |  |
| --- | --- | --- |
| **Excluded Variablesa** | | |
| Model | | Collinearity Statistics |
| Minimum Tolerance |
| 1 | CLAIMSORIGINAL | .183 |
| SECDUPBD | .879 |
| SELFCTUB | .878 |

|  |
| --- |
| a. Dependent Variable: logCRECEIVE |
| b. Predictors in the Model: (Constant), RATIOCIT, CLAIMS, GENERAL, CMADE, GYEAR, ORIGINAL |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | | | | | |
| Model | | | RATIOCIT | CLAIMS | GENERAL | CMADE | GYEAR |
| 1 | Correlations | RATIOCIT | 1.000 | -.022 | -.050 | .127 | -.098 |
| CLAIMS | -.022 | 1.000 | -.047 | -.142 | -.027 |
| GENERAL | -.050 | -.047 | 1.000 | -.031 | .256 |
| CMADE | .127 | -.142 | -.031 | 1.000 | -.083 |
| GYEAR | -.098 | -.027 | .256 | -.083 | 1.000 |
| ORIGINAL | -.050 | -.050 | -.179 | -.270 | -.036 |
| Covariances | RATIOCIT | .016 | -3.372E-6 | .000 | 1.842E-5 | .000 |
| CLAIMS | -3.372E-6 | 1.456E-6 | -3.392E-6 | -1.965E-7 | -4.738E-7 |
| GENERAL | .000 | -3.392E-6 | .004 | -2.085E-6 | .000 |
| CMADE | 1.842E-5 | -1.965E-7 | -2.085E-6 | 1.312E-6 | -1.358E-6 |
| GYEAR | .000 | -4.738E-7 | .000 | -1.358E-6 | .000 |
| ORIGINAL | .000 | -3.459E-6 | -.001 | -1.758E-5 | -2.969E-5 |
| 2 | Correlations | RATIOCIT | 1.000 | .003 | -.052 | .126 | -.098 |
| CLAIMS | .003 | 1.000 | .014 | -.025 | -.010 |
| GENERAL | -.052 | .014 | 1.000 | -.028 | .255 |
| CMADE | .126 | -.025 | -.028 | 1.000 | -.081 |
| GYEAR | -.098 | -.010 | .255 | -.081 | 1.000 |
| ORIGINAL | -.023 | .603 | -.086 | -.138 | -.020 |
| SECDUPBD | -.023 | -.006 | .045 | -.030 | -.027 |
| SELFCTUB | -.053 | -.023 | .008 | .037 | .022 |
| CLAIMSORIGINAL | -.017 | -.846 | -.046 | -.059 | -.006 |
| Covariances | RATIOCIT | .016 | 7.629E-7 | .000 | 1.836E-5 | .000 |
| CLAIMS | 7.629E-7 | 5.115E-6 | 1.853E-6 | -6.583E-8 | -3.202E-7 |
| GENERAL | .000 | 1.853E-6 | .004 | -1.945E-6 | .000 |
| CMADE | 1.836E-5 | -6.583E-8 | -1.945E-6 | 1.318E-6 | -1.329E-6 |
| GYEAR | .000 | -3.202E-7 | .000 | -1.329E-6 | .000 |
| ORIGINAL | .000 | .000 | .000 | -1.331E-5 | -2.452E-5 |
| SECDUPBD | .000 | -6.508E-7 | .000 | -1.502E-6 | -1.690E-5 |
| SELFCTUB | .000 | -3.441E-6 | 3.128E-5 | 2.748E-6 | 2.100E-5 |
| CLAIMSORIGINAL | -8.638E-6 | -7.764E-6 | -1.117E-5 | -2.766E-7 | -3.220E-7 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | | | | |
| Model | | | ORIGINAL | SECDUPBD | SELFCTUB | CLAIMSORIGINAL |
| 1 | Correlations | RATIOCIT | -.050 |  |  |  |
| CLAIMS | -.050 |  |  |  |
| GENERAL | -.179 |  |  |  |
| CMADE | -.270 |  |  |  |
| GYEAR | -.036 |  |  |  |
| ORIGINAL | 1.000 |  |  |  |
| Covariances | RATIOCIT | .000 |  |  |  |
| CLAIMS | -3.459E-6 |  |  |  |
| GENERAL | -.001 |  |  |  |
| CMADE | -1.758E-5 |  |  |  |
| GYEAR | -2.969E-5 |  |  |  |
| ORIGINAL | .003 |  |  |  |
| 2 | Correlations | RATIOCIT | -.023 | -.023 | -.053 | -.017 |
| CLAIMS | .603 | -.006 | -.023 | -.846 |
| GENERAL | -.086 | .045 | .008 | -.046 |
| CMADE | -.138 | -.030 | .037 | -.059 |
| GYEAR | -.020 | -.027 | .022 | -.006 |
| ORIGINAL | 1.000 | .007 | .006 | -.735 |
| SECDUPBD | .007 | 1.000 | -.196 | .004 |
| SELFCTUB | .006 | -.196 | 1.000 | .026 |
| CLAIMSORIGINAL | -.735 | .004 | .026 | 1.000 |
| Covariances | RATIOCIT | .000 | .000 | .000 | -8.638E-6 |
| CLAIMS | .000 | -6.508E-7 | -3.441E-6 | -7.764E-6 |
| GENERAL | .000 | .000 | 3.128E-5 | -1.117E-5 |
| CMADE | -1.331E-5 | -1.502E-6 | 2.748E-6 | -2.766E-7 |
| GYEAR | -2.452E-5 | -1.690E-5 | 2.100E-5 | -3.220E-7 |
| ORIGINAL | .007 | 2.728E-5 | 3.425E-5 | .000 |
| SECDUPBD | 2.728E-5 | .002 | -.001 | 7.006E-7 |
| SELFCTUB | 3.425E-5 | -.001 | .004 | 6.794E-6 |
| CLAIMSORIGINAL | .000 | 7.006E-7 | 6.794E-6 | 1.647E-5 |

|  |
| --- |
| a. Dependent Variable: logCRECEIVE |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
| (Constant) | CLAIMS | CMADE | GENERAL |
| 1 | 1 | 5.307 | 1.000 | .00 | .01 | .01 | .01 |
| 2 | .590 | 3.000 | .00 | .01 | .34 | .59 |
| 3 | .527 | 3.174 | .00 | .03 | .49 | .29 |
| 4 | .331 | 4.004 | .00 | .87 | .00 | .03 |
| 5 | .236 | 4.745 | .00 | .08 | .13 | .02 |
| 6 | .010 | 23.218 | .00 | .00 | .02 | .00 |
| 7 | 1.331E-7 | 6314.809 | 1.00 | .00 | .01 | .07 |
| 2 | 1 | 6.410 | 1.000 | .00 | .00 | .01 | .01 |
| 2 | .995 | 2.538 | .00 | .00 | .03 | .01 |
| 3 | .660 | 3.116 | .00 | .01 | .05 | .36 |
| 4 | .591 | 3.292 | .00 | .01 | .03 | .06 |
| 5 | .527 | 3.489 | .00 | .02 | .55 | .12 |
| 6 | .476 | 3.668 | .00 | .00 | .23 | .33 |
| 7 | .287 | 4.723 | .00 | .11 | .08 | .04 |
| 8 | .044 | 12.051 | .00 | .83 | .00 | .00 |
| 9 | .010 | 25.677 | .00 | .01 | .02 | .00 |
| 10 | 1.329E-7 | 6943.382 | 1.00 | .00 | .01 | .07 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Variance Proportions | | | | | |
| GYEAR | ORIGINAL | RATIOCIT | CLAIMSORIGINAL | SECDUPBD | SELFCTUB |
| 1 | 1 | .00 | .01 | .00 |  |  |  |
| 2 | .00 | .00 | .00 |  |  |  |
| 3 | .00 | .00 | .00 |  |  |  |
| 4 | .00 | .12 | .00 |  |  |  |
| 5 | .00 | .87 | .00 |  |  |  |
| 6 | .00 | .00 | .98 |  |  |  |
| 7 | 1.00 | .00 | .01 |  |  |  |
| 2 | 1 | .00 | .00 | .00 | .00 | .01 | .01 |
| 2 | .00 | .00 | .00 | .02 | .23 | .26 |
| 3 | .00 | .00 | .00 | .02 | .26 | .02 |
| 4 | .00 | .00 | .00 | .02 | .32 | .54 |
| 5 | .00 | .00 | .00 | .03 | .13 | .02 |
| 6 | .00 | .00 | .00 | .01 | .06 | .15 |
| 7 | .00 | .28 | .00 | .01 | .00 | .00 |
| 8 | .00 | .70 | .02 | .88 | .00 | .00 |
| 9 | .00 | .01 | .97 | .01 | .00 | .00 |
| 10 | 1.00 | .00 | .01 | .00 | .00 | .00 |

|  |
| --- |
| a. Dependent Variable: logCRECEIVE |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Casewise Diagnosticsa** | | | | |
| Case Number | Std. Residual | logCRECEIVE | Predicted Value | Residual |
| 202 | 3.893 | 3.737669618283370 | 1.388148576505667 | 2.349521041777704 |
| 230 | 4.030 | 4.718498871295090 | 2.285905306556838 | 2.432593564738252 |
| 243 | 3.429 | 3.135494215929150 | 1.065928600534789 | 2.069565615394361 |
| 269 | 3.349 | 2.639057329615260 | .617947331006890 | 2.021109998608371 |
| 446 | 3.445 | 3.218875824868200 | 1.139558356212943 | 2.079317468655257 |
| 457 | 4.013 | 3.433987204485150 | 1.011865850748473 | 2.422121353736677 |
| 832 | 3.125 | 2.639057329615260 | .752986076687250 | 1.886071252928010 |
| 859 | 3.563 | 2.564949357461540 | .414341646752613 | 2.150607710708927 |
| 1058 | 3.116 | 1.945910149055310 | .065267694600394 | 1.880642454454916 |
| 1063 | 4.049 | 3.637586159726390 | 1.193910852275735 | 2.443675307450655 |
| 1116 | 3.836 | 3.091042453358320 | .775945771776405 | 2.315096681581915 |
| 1220 | 3.299 | 3.496507561466480 | 1.505310400540486 | 1.991197160925994 |
| 1379 | 3.098 | 2.484906649788000 | .615250002856623 | 1.869656646931377 |
| 1680 | 3.725 | 3.258096538021480 | 1.009941980234544 | 2.248154557786936 |
| 1744 | 3.100 | 2.772588722239780 | .901372162285195 | 1.871216559954585 |

|  |
| --- |
| a. Dependent Variable: logCRECEIVE |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | -.220429137349129 | 2.285905361175537 | .803931581920265 | .533164126838409 | 1667 |
| Residual | -1.385528206825256 | 2.443675279617310 | .000000000000043 | .601942021944027 | 1667 |
| Std. Predicted Value | -1.921 | 2.780 | .000 | 1.000 | 1667 |
| Std. Residual | -2.296 | 4.049 | .000 | .997 | 1667 |

|  |
| --- |
| a. Dependent Variable: logCRECEIVE |

Appendix C. IBM Statistics SPSS 25 Output for Path Model Analysis

**Frequencies**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 28-OCT-2018 16:39:40 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_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 CRECEIVEln CRECEIVElog10 CRECEIVErecip  /STATISTICS=STDDEV RANGE MINIMUM MAXIMUM SEMEAN MEAN MEDIAN SKEWNESS SESKEW KURTOSIS SEKURT  /HISTOGRAM NORMAL  /ORDER=ANALYSIS. |
| Resources | Processor Time | 00:00:07.36 |
| Elapsed Time | 00:00:03.49 |

[DataSet1] D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Statistics** | | | | | |
|  | | CRECEIVE | CRECEIVEln | CRECEIVElog10 | CRECEIVErecip |
| N | Valid | 2000 | 2000 | 2000 | 2000 |
| Missing | 0 | 0 | 0 | 0 |
| Mean | | 3.18 | .7788 | .3382 | .589921152058347 |
| Std. Error of Mean | | .096 | .01770 | .00769 | .007877009960755 |
| Median | | 2.00 | .6931 | .3010 | .500000000000000 |
| Std. Deviation | | 4.309 | .79147 | .34373 | .352270594633808 |
| Skewness | | 10.292 | .797 | .797 | .095 |
| Std. Error of Skewness | | .055 | .055 | .055 | .055 |
| Kurtosis | | 214.022 | .090 | .090 | -1.630 |
| Std. Error of Kurtosis | | .109 | .109 | .109 | .109 |
| Range | | 111 | 4.72 | 2.05 | .99107142857142860 |
| Minimum | | 1 | .00 | .00 | .00892857142857143 |
| Maximum | | 112 | 4.72 | 2.05 | 1.00000000000000000 |

**Frequency Table**

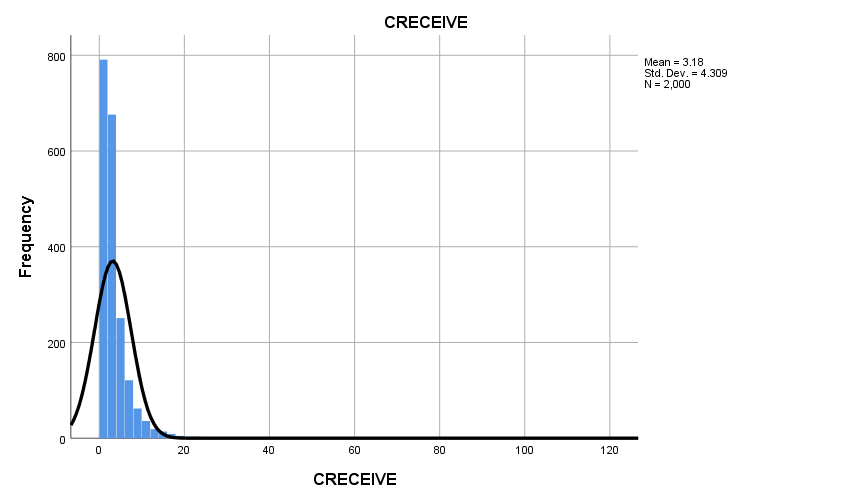
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CRECEIVE** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 1 | 791 | 39.6 | 39.6 | 39.6 |
| 2 | 433 | 21.7 | 21.7 | 61.2 |
| 3 | 243 | 12.2 | 12.2 | 73.4 |
| 4 | 146 | 7.3 | 7.3 | 80.7 |
| 5 | 105 | 5.3 | 5.3 | 85.9 |
| 6 | 75 | 3.8 | 3.8 | 89.6 |
| 7 | 46 | 2.3 | 2.3 | 92.0 |
| 8 | 36 | 1.8 | 1.8 | 93.8 |
| 9 | 26 | 1.3 | 1.3 | 95.1 |
| 10 | 22 | 1.1 | 1.1 | 96.2 |
| 11 | 14 | .7 | .7 | 96.9 |
| 12 | 10 | .5 | .5 | 97.4 |
| 13 | 9 | .4 | .4 | 97.8 |
| 14 | 10 | .5 | .5 | 98.3 |
| 15 | 4 | .2 | .2 | 98.5 |
| 16 | 4 | .2 | .2 | 98.7 |
| 17 | 5 | .3 | .3 | 99.0 |
| 18 | 5 | .3 | .3 | 99.2 |
| 20 | 1 | .1 | .1 | 99.3 |
| 21 | 2 | .1 | .1 | 99.4 |
| 22 | 1 | .1 | .1 | 99.4 |
| 23 | 2 | .1 | .1 | 99.5 |
| 25 | 1 | .1 | .1 | 99.6 |
| 26 | 2 | .1 | .1 | 99.7 |
| 30 | 1 | .1 | .1 | 99.7 |
| 31 | 1 | .1 | .1 | 99.8 |
| 32 | 1 | .1 | .1 | 99.8 |
| 33 | 1 | .1 | .1 | 99.9 |
| 38 | 1 | .1 | .1 | 99.9 |
| 42 | 1 | .1 | .1 | 100.0 |
| 112 | 1 | .1 | .1 | 100.0 |
| Total | 2000 | 100.0 | 100.0 |  |

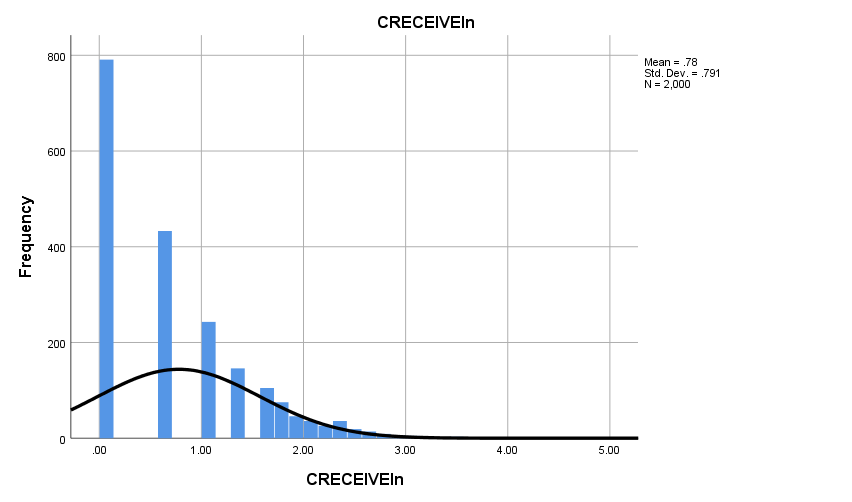
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CRECEIVEln** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | .00 | 791 | 39.6 | 39.6 | 39.6 |
| .69 | 433 | 21.7 | 21.7 | 61.2 |
| 1.10 | 243 | 12.2 | 12.2 | 73.4 |
| 1.39 | 146 | 7.3 | 7.3 | 80.7 |
| 1.61 | 105 | 5.3 | 5.3 | 85.9 |
| 1.79 | 75 | 3.8 | 3.8 | 89.6 |
| 1.95 | 46 | 2.3 | 2.3 | 92.0 |
| 2.08 | 36 | 1.8 | 1.8 | 93.8 |
| 2.20 | 26 | 1.3 | 1.3 | 95.1 |
| 2.30 | 22 | 1.1 | 1.1 | 96.2 |
| 2.40 | 14 | .7 | .7 | 96.9 |
| 2.48 | 10 | .5 | .5 | 97.4 |
| 2.56 | 9 | .4 | .4 | 97.8 |
| 2.64 | 10 | .5 | .5 | 98.3 |
| 2.71 | 4 | .2 | .2 | 98.5 |
| 2.77 | 4 | .2 | .2 | 98.7 |
| 2.83 | 5 | .3 | .3 | 99.0 |
| 2.89 | 5 | .3 | .3 | 99.2 |
| 3.00 | 1 | .1 | .1 | 99.3 |
| 3.04 | 2 | .1 | .1 | 99.4 |
| 3.09 | 1 | .1 | .1 | 99.4 |
| 3.14 | 2 | .1 | .1 | 99.5 |
| 3.22 | 1 | .1 | .1 | 99.6 |
| 3.26 | 2 | .1 | .1 | 99.7 |
| 3.40 | 1 | .1 | .1 | 99.7 |
| 3.43 | 1 | .1 | .1 | 99.8 |
| 3.47 | 1 | .1 | .1 | 99.8 |
| 3.50 | 1 | .1 | .1 | 99.9 |
| 3.64 | 1 | .1 | .1 | 99.9 |
| 3.74 | 1 | .1 | .1 | 100.0 |
| 4.72 | 1 | .1 | .1 | 100.0 |
| Total | 2000 | 100.0 | 100.0 |  |

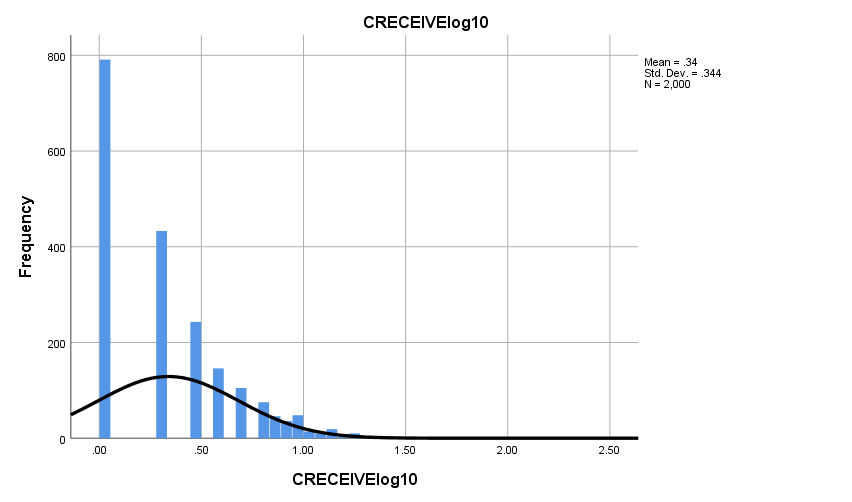
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CRECEIVElog10** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | .00 | 791 | 39.6 | 39.6 | 39.6 |
| .30 | 433 | 21.7 | 21.7 | 61.2 |
| .48 | 243 | 12.2 | 12.2 | 73.4 |
| .60 | 146 | 7.3 | 7.3 | 80.7 |
| .70 | 105 | 5.3 | 5.3 | 85.9 |
| .78 | 75 | 3.8 | 3.8 | 89.6 |
| .85 | 46 | 2.3 | 2.3 | 92.0 |
| .90 | 36 | 1.8 | 1.8 | 93.8 |
| .95 | 26 | 1.3 | 1.3 | 95.1 |
| 1.00 | 22 | 1.1 | 1.1 | 96.2 |
| 1.04 | 14 | .7 | .7 | 96.9 |
| 1.08 | 10 | .5 | .5 | 97.4 |
| 1.11 | 9 | .4 | .4 | 97.8 |
| 1.15 | 10 | .5 | .5 | 98.3 |
| 1.18 | 4 | .2 | .2 | 98.5 |
| 1.20 | 4 | .2 | .2 | 98.7 |
| 1.23 | 5 | .3 | .3 | 99.0 |
| 1.26 | 5 | .3 | .3 | 99.2 |
| 1.30 | 1 | .1 | .1 | 99.3 |
| 1.32 | 2 | .1 | .1 | 99.4 |
| 1.34 | 1 | .1 | .1 | 99.4 |
| 1.36 | 2 | .1 | .1 | 99.5 |
| 1.40 | 1 | .1 | .1 | 99.6 |
| 1.41 | 2 | .1 | .1 | 99.7 |
| 1.48 | 1 | .1 | .1 | 99.7 |
| 1.49 | 1 | .1 | .1 | 99.8 |
| 1.51 | 1 | .1 | .1 | 99.8 |
| 1.52 | 1 | .1 | .1 | 99.9 |
| 1.58 | 1 | .1 | .1 | 99.9 |
| 1.62 | 1 | .1 | .1 | 100.0 |
| 2.05 | 1 | .1 | .1 | 100.0 |
| Total | 2000 | 100.0 | 100.0 |  |

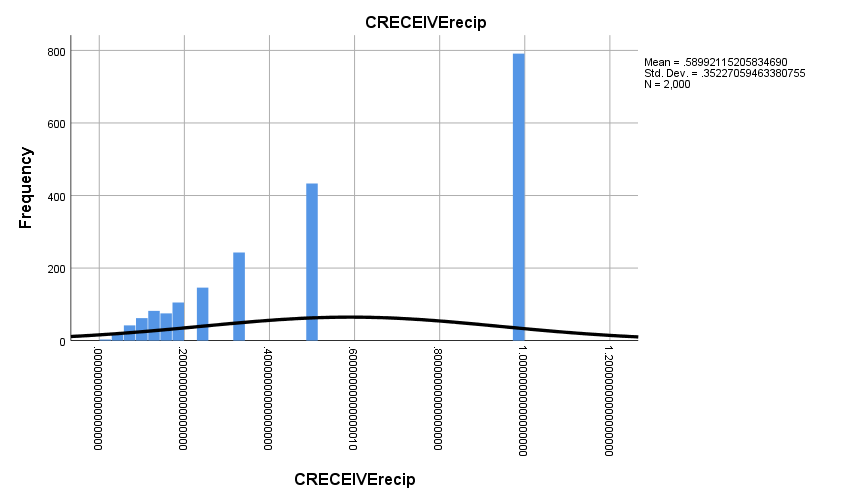
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CRECEIVErecip** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | .00892857142857143 | 1 | .1 | .1 | .1 |
| .02380952380952380 | 1 | .1 | .1 | .1 |
| .02631578947368420 | 1 | .1 | .1 | .2 |
| .03030303030303030 | 1 | .1 | .1 | .2 |
| .03125000000000000 | 1 | .1 | .1 | .3 |
| .03225806451612900 | 1 | .1 | .1 | .3 |
| .03333333333333330 | 1 | .1 | .1 | .4 |
| .03846153846153850 | 2 | .1 | .1 | .4 |
| .04000000000000000 | 1 | .1 | .1 | .5 |
| .04347826086956520 | 2 | .1 | .1 | .6 |
| .04545454545454550 | 1 | .1 | .1 | .7 |
| .04761904761904760 | 2 | .1 | .1 | .8 |
| .05000000000000000 | 1 | .1 | .1 | .8 |
| .05555555555555560 | 5 | .3 | .3 | 1.1 |
| .05882352941176470 | 5 | .3 | .3 | 1.3 |
| .06250000000000000 | 4 | .2 | .2 | 1.5 |
| .06666666666666670 | 4 | .2 | .2 | 1.7 |
| .07142857142857140 | 10 | .5 | .5 | 2.2 |
| .07692307692307690 | 9 | .4 | .4 | 2.7 |
| .08333333333333330 | 10 | .5 | .5 | 3.2 |
| .09090909090909090 | 14 | .7 | .7 | 3.9 |
| .10000000000000000 | 22 | 1.1 | 1.1 | 5.0 |
| .11111111111111100 | 26 | 1.3 | 1.3 | 6.3 |
| .12500000000000000 | 36 | 1.8 | 1.8 | 8.1 |
| .14285714285714300 | 46 | 2.3 | 2.3 | 10.4 |
| .16666666666666700 | 75 | 3.8 | 3.8 | 14.1 |
| .20000000000000000 | 105 | 5.3 | 5.3 | 19.4 |
| .25000000000000000 | 146 | 7.3 | 7.3 | 26.7 |
| .33333333333333300 | 243 | 12.2 | 12.2 | 38.8 |
| .50000000000000000 | 433 | 21.7 | 21.7 | 60.5 |
| 1.00000000000000000 | 791 | 39.6 | 39.6 | 100.0 |
| Total | 2000 | 100.0 | 100.0 |  |

**Histogram**



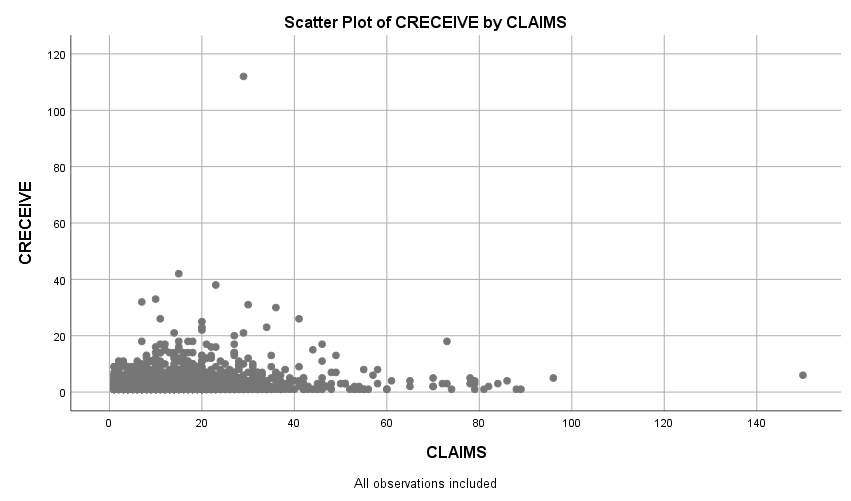






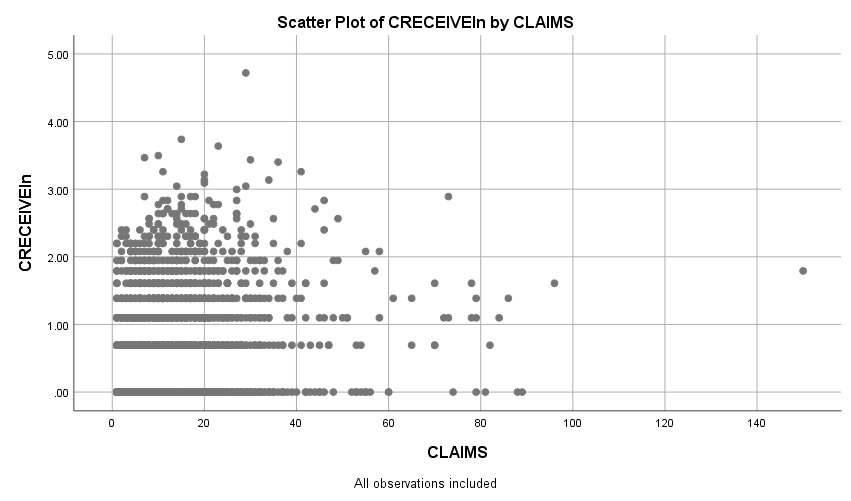
**Graph**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 28-OCT-2018 16:50:51 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 2000 |
| Syntax | | GRAPH  /SCATTERPLOT(BIVAR)=CLAIMS WITH CRECEIVE  /MISSING=LISTWISE  /TITLE='Scatter Plot of CRECEIVE by CLAIMS'  /FOOTNOTE='All observations included'. |
| Resources | Processor Time | 00:00:01.69 |
| Elapsed Time | 00:00:00.88 |



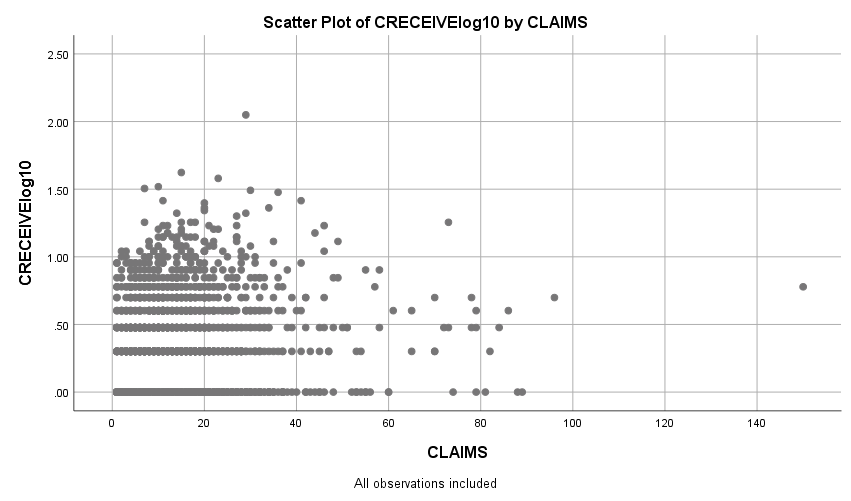
**Graph**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 28-OCT-2018 16:52:07 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 2000 |
| Syntax | | GRAPH  /SCATTERPLOT(BIVAR)=CLAIMS WITH CRECEIVEln  /MISSING=LISTWISE  /TITLE='Scatter Plot of CRECEIVEln by CLAIMS'  /FOOTNOTE='All observations included'. |
| Resources | Processor Time | 00:00:00.97 |
| Elapsed Time | 00:00:00.70 |



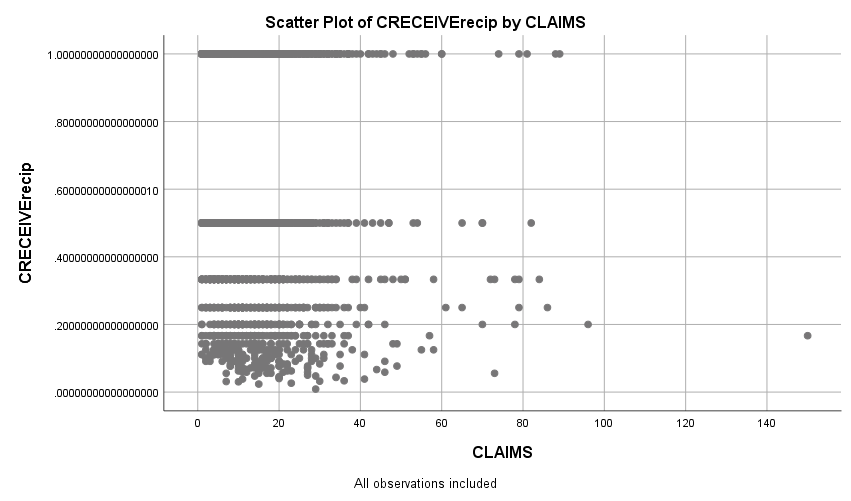
**Graph**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 28-OCT-2018 16:52:57 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 2000 |
| Syntax | | GRAPH  /SCATTERPLOT(BIVAR)=CLAIMS WITH CRECEIVElog10  /MISSING=LISTWISE  /TITLE='Scatter Plot of CRECEIVElog10 by CLAIMS'  /FOOTNOTE='All observations included'. |
| Resources | Processor Time | 00:00:01.19 |
| Elapsed Time | 00:00:00.60 |



**Graph**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 28-OCT-2018 16:53:51 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 2000 |
| Syntax | | GRAPH  /SCATTERPLOT(BIVAR)=CLAIMS WITH CRECEIVErecip  /MISSING=LISTWISE  /TITLE='Scatter Plot of CRECEIVErecip by CLAIMS'  /FOOTNOTE='All observations included'. |
| Resources | Processor Time | 00:00:01.11 |
| Elapsed Time | 00:00:00.59 |



**Regression**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 28-OCT-2018 19:27:42 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | CRECEIVE < 40 AND CLAIMS < 90 (FILTER) |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 1996 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| Cases Used | Statistics are based on cases with no missing values for any variable used. |
| Syntax | | REGRESSION  /DESCRIPTIVES MEAN STDDEV CORR SIG N  /MISSING LISTWISE  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL CHANGE ZPP  /CRITERIA=PIN(.05) POUT(.10)  /NOORIGIN  /DEPENDENT GENERAL  /METHOD=ENTER ORIGINAL  /SCATTERPLOT=(\*ZRESID ,\*ZPRED)  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)  /CASEWISE PLOT(ZRESID) OUTLIERS(3). |
| Resources | Processor Time | 00:00:08.63 |
| Elapsed Time | 00:00:04.53 |
| Memory Required | 3488 bytes |
| Additional Memory Required for Residual Plots | 680 bytes |

[DataSet1] D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptive Statistics** | | | |
|  | Mean | Std. Deviation | N |
| GENERAL | .194597 | .2589553 | 1958 |
| ORIGINAL | .398282 | .2739932 | 1958 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Correlations** | | | |
|  | | GENERAL | ORIGINAL |
| Pearson Correlation | GENERAL | 1.000 | .169 |
| ORIGINAL | .169 | 1.000 |
| Sig. (1-tailed) | GENERAL | . | .000 |
| ORIGINAL | .000 | . |
| N | GENERAL | 1958 | 1958 |
| ORIGINAL | 1958 | 1958 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | ORIGINALb | . | Enter |

|  |
| --- |
| a. Dependent Variable: GENERAL |
| b. All requested variables entered. |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | |
| R Square Change | F Change | df1 |
| 1 | .169a | .029 | .028 | .2552980 | .029 | 57.473 | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Summaryb** | | | |
| Model | Change Statistics | | |
| df2 | Sig. F Change |  |
| 1 | 1956 | .000 | 2.053 |

|  |
| --- |
| a. Predictors: (Constant), ORIGINAL |
| b. Dependent Variable: GENERAL |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 3.746 | 1 | 3.746 | 57.473 | .000b |
| Residual | 127.486 | 1956 | .065 |  |  |
| Total | 131.232 | 1957 |  |  |  |

|  |
| --- |
| a. Dependent Variable: GENERAL |
| b. Predictors: (Constant), ORIGINAL |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | .131 | .010 |  | 12.867 | .000 |
| ORIGINAL | .160 | .021 | .169 | 7.581 | .000 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | |
| Model | | 95.0% Confidence Interval for B | | Correlations | | | Collinearity Statistics |
| Lower Bound | Upper Bound | Zero-order | Partial | Part | Tolerance |
| 1 | (Constant) | .111 | .151 |  |  |  |  |
| ORIGINAL | .118 | .201 | .169 | .169 | .169 | 1.000 |

|  |  |  |
| --- | --- | --- |
| **Coefficientsa** | | |
| Model | | Collinearity Statistics |
| VIF |
| 1 | (Constant) |  |
| ORIGINAL | 1.000 |

|  |
| --- |
| a. Dependent Variable: GENERAL |

|  |  |  |  |
| --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | |
| Model | | | ORIGINAL |
| 1 | Correlations | ORIGINAL | 1.000 |
| Covariances | ORIGINAL | .000 |

|  |
| --- |
| a. Dependent Variable: GENERAL |

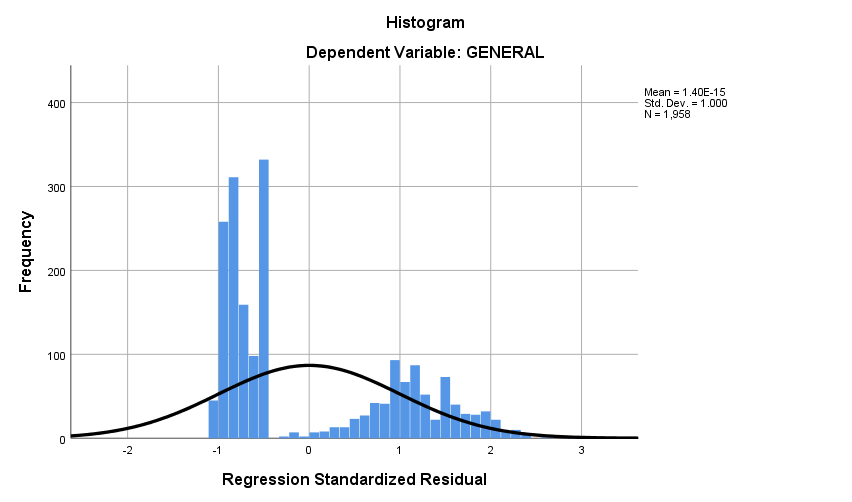
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | |
| (Constant) | ORIGINAL |
| 1 | 1 | 1.824 | 1.000 | .09 | .09 |
| 2 | .176 | 3.219 | .91 | .91 |

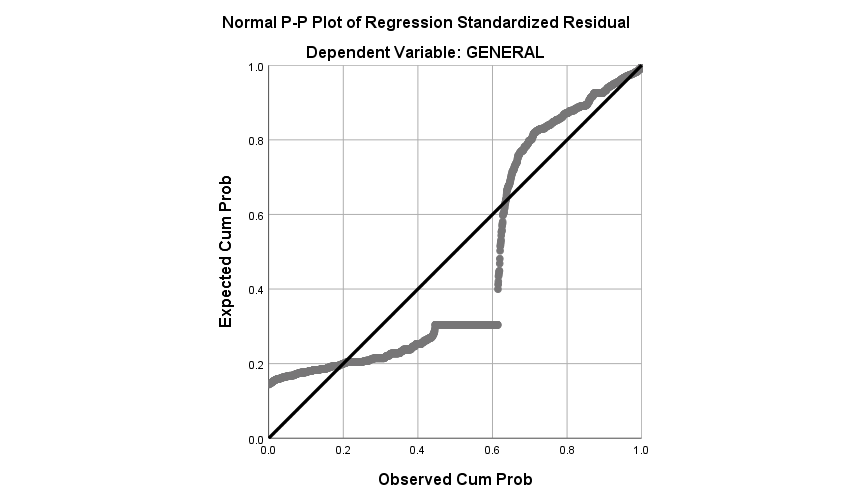
|  |
| --- |
| a. Dependent Variable: GENERAL |

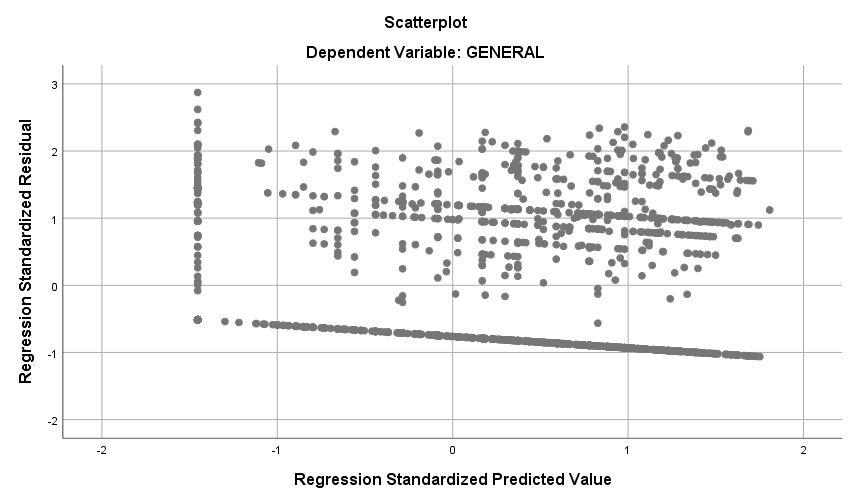
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | .131001 | .273656 | .194597 | .0437504 | 1958 |
| Residual | -.2711814 | .7331990 | .0000000 | .2552328 | 1958 |
| Std. Predicted Value | -1.454 | 1.807 | .000 | 1.000 | 1958 |
| Std. Residual | -1.062 | 2.872 | .000 | 1.000 | 1958 |

|  |
| --- |
| a. Dependent Variable: GENERAL |

**Charts**







**Regression**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 28-OCT-2018 19:32:32 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | CRECEIVE < 40 AND CLAIMS < 90 (FILTER) |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 1996 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| Cases Used | Statistics are based on cases with no missing values for any variable used. |
| Syntax | | REGRESSION  /DESCRIPTIVES MEAN STDDEV CORR SIG N  /MISSING LISTWISE  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL CHANGE ZPP  /CRITERIA=PIN(.05) POUT(.10)  /NOORIGIN  /DEPENDENT CLAIMS  /METHOD=ENTER ORIGINAL GENERAL GYEAR RATIOCIT  /SCATTERPLOT=(\*ZRESID ,\*ZPRED)  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)  /CASEWISE PLOT(ZRESID) OUTLIERS(3). |
| Resources | Processor Time | 00:00:02.46 |
| Elapsed Time | 00:00:01.57 |
| Memory Required | 5072 bytes |
| Additional Memory Required for Residual Plots | 632 bytes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptive Statistics** | | | |
|  | Mean | Std. Deviation | N |
| CLAIMS | 14.97 | 11.689 | 1958 |
| ORIGINAL | .398282 | .2739932 | 1958 |
| GENERAL | .194597 | .2589553 | 1958 |
| GYEAR | 1996.27 | 1.075 | 1958 |
| RATIOCIT | .939529 | .1404329 | 1958 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Correlations** | | | | | | |
|  | | CLAIMS | ORIGINAL | GENERAL | GYEAR | RATIOCIT |
| Pearson Correlation | CLAIMS | 1.000 | .101 | .056 | .039 | .054 |
| ORIGINAL | .101 | 1.000 | .169 | .017 | .052 |
| GENERAL | .056 | .169 | 1.000 | -.238 | .043 |
| GYEAR | .039 | .017 | -.238 | 1.000 | .083 |
| RATIOCIT | .054 | .052 | .043 | .083 | 1.000 |
| Sig. (1-tailed) | CLAIMS | . | .000 | .006 | .042 | .008 |
| ORIGINAL | .000 | . | .000 | .225 | .011 |
| GENERAL | .006 | .000 | . | .000 | .028 |
| GYEAR | .042 | .225 | .000 | . | .000 |
| RATIOCIT | .008 | .011 | .028 | .000 | . |
| N | CLAIMS | 1958 | 1958 | 1958 | 1958 | 1958 |
| ORIGINAL | 1958 | 1958 | 1958 | 1958 | 1958 |
| GENERAL | 1958 | 1958 | 1958 | 1958 | 1958 |
| GYEAR | 1958 | 1958 | 1958 | 1958 | 1958 |
| RATIOCIT | 1958 | 1958 | 1958 | 1958 | 1958 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | RATIOCIT, GENERAL, ORIGINAL, GYEARb | . | Enter |

|  |
| --- |
| a. Dependent Variable: CLAIMS |
| b. All requested variables entered. |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | |
| R Square Change | F Change | df1 |
| 1 | .127a | .016 | .014 | 11.607 | .016 | 7.952 | 4 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Summaryb** | | | |
| Model | Change Statistics | | |
| df2 | Sig. F Change |  |
| 1 | 1953 | .000 | 1.966 |

|  |
| --- |
| a. Predictors: (Constant), RATIOCIT, GENERAL, ORIGINAL, GYEAR |
| b. Dependent Variable: CLAIMS |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 4285.217 | 4 | 1071.304 | 7.952 | .000b |
| Residual | 263126.455 | 1953 | 134.729 |  |  |
| Total | 267411.672 | 1957 |  |  |  |

|  |
| --- |
| a. Dependent Variable: CLAIMS |
| b. Predictors: (Constant), RATIOCIT, GENERAL, ORIGINAL, GYEAR |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | -986.372 | 504.599 |  | -1.955 | .051 |
| ORIGINAL | 3.820 | .974 | .090 | 3.921 | .000 |
| GENERAL | 2.274 | 1.062 | .050 | 2.142 | .032 |
| GYEAR | .499 | .253 | .046 | 1.973 | .049 |
| RATIOCIT | 3.608 | 1.880 | .043 | 1.919 | .055 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | |
| Model | | 95.0% Confidence Interval for B | | Correlations | | | Collinearity Statistics |
| Lower Bound | Upper Bound | Zero-order | Partial | Part | Tolerance |
| 1 | (Constant) | -1975.981 | 3.237 |  |  |  |  |
| ORIGINAL | 1.909 | 5.730 | .101 | .088 | .088 | .966 |
| GENERAL | .192 | 4.357 | .056 | .048 | .048 | .911 |
| GYEAR | .003 | .995 | .039 | .045 | .044 | .932 |
| RATIOCIT | -.080 | 7.296 | .054 | .043 | .043 | .987 |

|  |  |  |
| --- | --- | --- |
| **Coefficientsa** | | |
| Model | | Collinearity Statistics |
| VIF |
| 1 | (Constant) |  |
| ORIGINAL | 1.035 |
| GENERAL | 1.098 |
| GYEAR | 1.073 |
| RATIOCIT | 1.013 |

|  |
| --- |
| a. Dependent Variable: CLAIMS |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | | | | |
| Model | | | RATIOCIT | GENERAL | ORIGINAL | GYEAR |
| 1 | Correlations | RATIOCIT | 1.000 | -.057 | -.040 | -.094 |
| GENERAL | -.057 | 1.000 | -.175 | .248 |
| ORIGINAL | -.040 | -.175 | 1.000 | -.056 |
| GYEAR | -.094 | .248 | -.056 | 1.000 |
| Covariances | RATIOCIT | 3.536 | -.113 | -.073 | -.044 |
| GENERAL | -.113 | 1.128 | -.181 | .067 |
| ORIGINAL | -.073 | -.181 | .949 | -.014 |
| GYEAR | -.044 | .067 | -.014 | .064 |

|  |
| --- |
| a. Dependent Variable: CLAIMS |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
| (Constant) | ORIGINAL | GENERAL | GYEAR |
| 1 | 1 | 4.172 | 1.000 | .00 | .01 | .02 | .00 |
| 2 | .558 | 2.734 | .00 | .00 | .91 | .00 |
| 3 | .255 | 4.048 | .00 | .98 | .01 | .00 |
| 4 | .015 | 16.893 | .00 | .00 | .00 | .00 |
| 5 | 1.351E-7 | 5557.627 | 1.00 | .00 | .06 | 1.00 |

|  |  |  |
| --- | --- | --- |
| **Collinearity Diagnosticsa** | | |
| Model | Dimension | Variance Proportions |
| RATIOCIT |
| 1 | 1 | .00 |
| 2 | .00 |
| 3 | .01 |
| 4 | .98 |
| 5 | .01 |

|  |
| --- |
| a. Dependent Variable: CLAIMS |

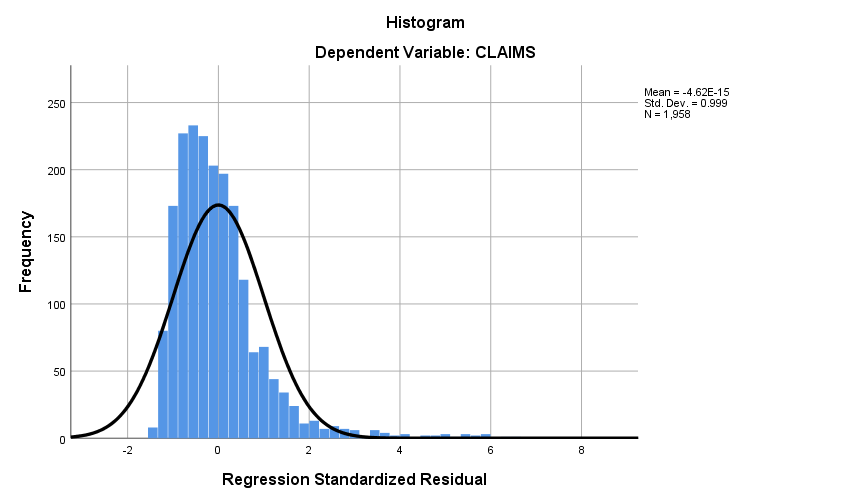
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Casewise Diagnosticsa** | | | | |
| Case Number | Std. Residual | CLAIMS | Predicted Value | Residual |
| 33 | 5.444 | 79 | 15.81 | 63.192 |
| 102 | 6.496 | 88 | 12.60 | 75.404 |
| 114 | 4.478 | 70 | 18.02 | 51.981 |
| 117 | 3.576 | 57 | 15.49 | 41.510 |
| 149 | 4.972 | 72 | 14.29 | 57.707 |
| 187 | 3.016 | 53 | 17.99 | 35.012 |
| 255 | 5.088 | 73 | 13.94 | 59.061 |
| 290 | 4.781 | 70 | 14.51 | 55.490 |
| 382 | 4.097 | 61 | 13.45 | 47.551 |
| 501 | 3.060 | 49 | 13.49 | 35.515 |
| 515 | 3.909 | 60 | 14.63 | 45.375 |
| 541 | 4.146 | 65 | 16.87 | 48.128 |
| 605 | 4.999 | 74 | 15.98 | 58.019 |
| 668 | 3.610 | 56 | 14.09 | 41.907 |
| 766 | 3.054 | 52 | 16.55 | 35.445 |
| 794 | 3.606 | 58 | 16.14 | 41.857 |
| 860 | 3.771 | 60 | 16.23 | 43.768 |
| 868 | 5.362 | 79 | 16.77 | 62.233 |
| 951 | 4.127 | 65 | 17.10 | 47.903 |
| 1010 | 3.075 | 51 | 15.30 | 35.696 |
| 1024 | 5.912 | 84 | 15.38 | 68.617 |
| 1029 | 3.382 | 54 | 14.74 | 39.258 |
| 1128 | 5.289 | 79 | 17.60 | 61.395 |
| 1160 | 5.807 | 81 | 13.59 | 67.406 |
| 1210 | 3.789 | 58 | 14.02 | 43.981 |
| 1248 | 4.772 | 73 | 17.61 | 55.390 |
| 1272 | 3.367 | 55 | 15.92 | 39.084 |
| 1336 | 3.439 | 55 | 15.08 | 39.916 |
| 1381 | 6.375 | 89 | 15.00 | 73.995 |
| 1451 | 3.266 | 51 | 13.09 | 37.905 |
| 1461 | 3.375 | 54 | 14.83 | 39.174 |
| 1480 | 5.592 | 78 | 13.09 | 64.905 |
| 1507 | 5.635 | 82 | 16.60 | 65.402 |
| 1557 | 3.104 | 53 | 16.97 | 36.029 |
| 1655 | 5.537 | 78 | 13.73 | 64.274 |
| 1671 | 3.438 | 53 | 13.09 | 39.905 |
| 1774 | 5.883 | 86 | 17.72 | 68.280 |
| 1884 | 3.488 | 55 | 14.51 | 40.492 |
| 1991 | 4.615 | 70 | 16.43 | 53.565 |

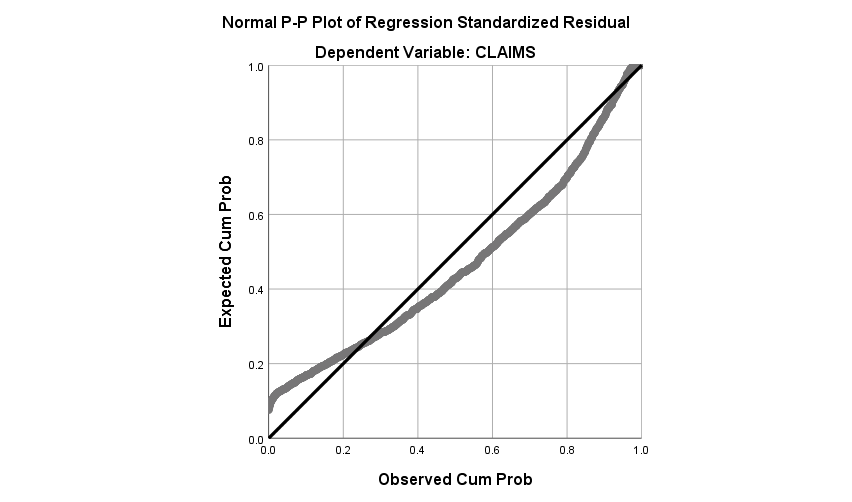
|  |
| --- |
| a. Dependent Variable: CLAIMS |

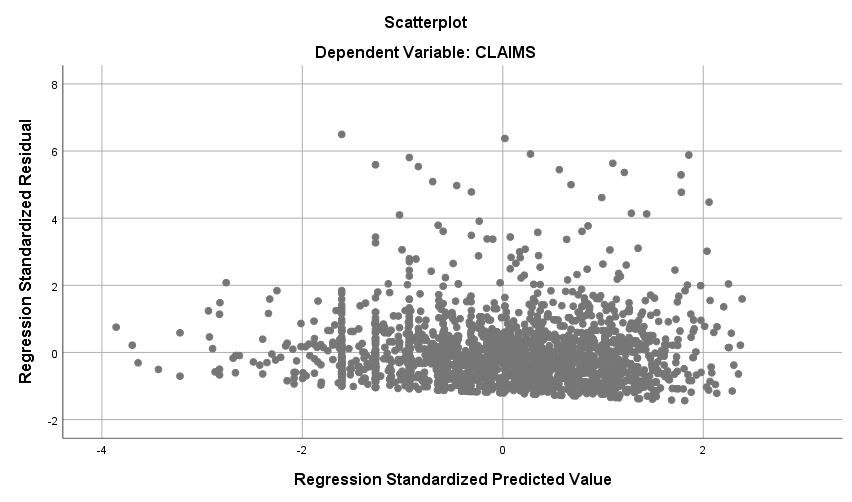
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | 9.27 | 18.51 | 14.97 | 1.480 | 1958 |
| Residual | -16.660 | 75.404 | .000 | 11.595 | 1958 |
| Std. Predicted Value | -3.858 | 2.388 | .000 | 1.000 | 1958 |
| Std. Residual | -1.435 | 6.496 | .000 | .999 | 1958 |

|  |
| --- |
| a. Dependent Variable: CLAIMS |

**Charts**







**Regression**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 28-OCT-2018 19:34:32 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | CRECEIVE < 40 AND CLAIMS < 90 (FILTER) |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 1996 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| Cases Used | Statistics are based on cases with no missing values for any variable used. |
| Syntax | | REGRESSION  /DESCRIPTIVES MEAN STDDEV CORR SIG N  /MISSING LISTWISE  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL CHANGE ZPP  /CRITERIA=PIN(.05) POUT(.10)  /NOORIGIN  /DEPENDENT CRECEIVEln  /METHOD=ENTER ORIGINAL GENERAL GYEAR RATIOCIT CLAIMS  /SCATTERPLOT=(\*ZRESID ,\*ZPRED)  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)  /CASEWISE PLOT(ZRESID) OUTLIERS(3). |
| Resources | Processor Time | 00:00:01.95 |
| Elapsed Time | 00:00:01.54 |
| Memory Required | 5728 bytes |
| Additional Memory Required for Residual Plots | 616 bytes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptive Statistics** | | | |
|  | Mean | Std. Deviation | N |
| CRECEIVEln | .7773 | .78494 | 1958 |
| ORIGINAL | .398282 | .2739932 | 1958 |
| GENERAL | .194597 | .2589553 | 1958 |
| GYEAR | 1996.27 | 1.075 | 1958 |
| RATIOCIT | .939529 | .1404329 | 1958 |
| CLAIMS | 14.97 | 11.689 | 1958 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Correlations** | | | | | | |
|  | | CRECEIVEln | ORIGINAL | GENERAL | GYEAR | RATIOCIT |
| Pearson Correlation | CRECEIVEln | 1.000 | .019 | .612 | -.339 | .098 |
| ORIGINAL | .019 | 1.000 | .169 | .017 | .052 |
| GENERAL | .612 | .169 | 1.000 | -.238 | .043 |
| GYEAR | -.339 | .017 | -.238 | 1.000 | .083 |
| RATIOCIT | .098 | .052 | .043 | .083 | 1.000 |
| CLAIMS | .083 | .101 | .056 | .039 | .054 |
| Sig. (1-tailed) | CRECEIVEln | . | .204 | .000 | .000 | .000 |
| ORIGINAL | .204 | . | .000 | .225 | .011 |
| GENERAL | .000 | .000 | . | .000 | .028 |
| GYEAR | .000 | .225 | .000 | . | .000 |
| RATIOCIT | .000 | .011 | .028 | .000 | . |
| CLAIMS | .000 | .000 | .006 | .042 | .008 |
| N | CRECEIVEln | 1958 | 1958 | 1958 | 1958 | 1958 |
| ORIGINAL | 1958 | 1958 | 1958 | 1958 | 1958 |
| GENERAL | 1958 | 1958 | 1958 | 1958 | 1958 |
| GYEAR | 1958 | 1958 | 1958 | 1958 | 1958 |
| RATIOCIT | 1958 | 1958 | 1958 | 1958 | 1958 |
| CLAIMS | 1958 | 1958 | 1958 | 1958 | 1958 |

|  |  |  |
| --- | --- | --- |
| **Correlations** | | |
|  | | CLAIMS |
| Pearson Correlation | CRECEIVEln | .083 |
| ORIGINAL | .101 |
| GENERAL | .056 |
| GYEAR | .039 |
| RATIOCIT | .054 |
| CLAIMS | 1.000 |
| Sig. (1-tailed) | CRECEIVEln | .000 |
| ORIGINAL | .000 |
| GENERAL | .006 |
| GYEAR | .042 |
| RATIOCIT | .008 |
| CLAIMS | . |
| N | CRECEIVEln | 1958 |
| ORIGINAL | 1958 |
| GENERAL | 1958 |
| GYEAR | 1958 |
| RATIOCIT | 1958 |
| CLAIMS | 1958 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | CLAIMS, GYEAR, RATIOCIT, ORIGINAL, GENERALb | . | Enter |

|  |
| --- |
| a. Dependent Variable: CRECEIVEln |
| b. All requested variables entered. |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | |
| R Square Change | F Change | df1 |
| 1 | .658a | .433 | .432 | .59178 | .433 | 298.204 | 5 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Summaryb** | | | |
| Model | Change Statistics | | |
| df2 | Sig. F Change |  |
| 1 | 1952 | .000 | 2.008 |

|  |
| --- |
| a. Predictors: (Constant), CLAIMS, GYEAR, RATIOCIT, ORIGINAL, GENERAL |
| b. Dependent Variable: CRECEIVEln |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 522.161 | 5 | 104.432 | 298.204 | .000b |
| Residual | 683.599 | 1952 | .350 |  |  |
| Total | 1205.759 | 1957 |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| a. Dependent Variable: CRECEIVEln | | | | | | |
| b. Predictors: (Constant), CLAIMS, GYEAR, RATIOCIT, ORIGINAL, GENERAL | | | | | | |
| **Coefficientsa** | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | |
| B | Std. Error | Beta |
| 1 | (Constant) | 310.279 | 25.751 |  | 12.049 | .000 | |
| ORIGINAL | -.243 | .050 | -.085 | -4.874 | .000 | |
| GENERAL | 1.723 | .054 | .569 | 31.794 | .000 | |
| GYEAR | -.155 | .013 | -.213 | -12.046 | .000 | |
| RATIOCIT | .516 | .096 | .092 | 5.382 | .000 | |
| CLAIMS | .004 | .001 | .063 | 3.650 | .000 | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | |
| Model | | 95.0% Confidence Interval for B | | Correlations | | | Collinearity Statistics |
| Lower Bound | Upper Bound | Zero-order | Partial | Part | Tolerance |
| 1 | (Constant) | 259.776 | 360.782 |  |  |  |  |
| ORIGINAL | -.341 | -.145 | .019 | -.110 | -.083 | .959 |
| GENERAL | 1.617 | 1.830 | .612 | .584 | .542 | .908 |
| GYEAR | -.181 | -.130 | -.339 | -.263 | -.205 | .930 |
| RATIOCIT | .328 | .705 | .098 | .121 | .092 | .986 |
| CLAIMS | .002 | .006 | .083 | .082 | .062 | .984 |

|  |  |  |
| --- | --- | --- |
| **Coefficientsa** | | |
| Model | | Collinearity Statistics |
| VIF |
| 1 | (Constant) |  |
| ORIGINAL | 1.043 |
| GENERAL | 1.101 |
| GYEAR | 1.075 |
| RATIOCIT | 1.015 |
| CLAIMS | 1.016 |

|  |
| --- |
| a. Dependent Variable: CRECEIVEln |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | | | | | |
| Model | | | CLAIMS | GYEAR | RATIOCIT | ORIGINAL | GENERAL |
| 1 | Correlations | CLAIMS | 1.000 | -.045 | -.043 | -.088 | -.048 |
| GYEAR | -.045 | 1.000 | -.091 | -.052 | .250 |
| RATIOCIT | -.043 | -.091 | 1.000 | -.036 | -.055 |
| ORIGINAL | -.088 | -.052 | -.036 | 1.000 | -.170 |
| GENERAL | -.048 | .250 | -.055 | -.170 | 1.000 |
| Covariances | CLAIMS | 1.331E-6 | -6.640E-7 | -4.802E-6 | -5.084E-6 | -3.027E-6 |
| GYEAR | -6.640E-7 | .000 | .000 | -3.315E-5 | .000 |
| RATIOCIT | -4.802E-6 | .000 | .009 | .000 | .000 |
| ORIGINAL | -5.084E-6 | -3.315E-5 | .000 | .002 | .000 |
| GENERAL | -3.027E-6 | .000 | .000 | .000 | .003 |

|  |
| --- |
| a. Dependent Variable: CRECEIVEln |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
| (Constant) | ORIGINAL | GENERAL | GYEAR |
| 1 | 1 | 4.843 | 1.000 | .00 | .01 | .01 | .00 |
| 2 | .574 | 2.905 | .00 | .00 | .89 | .00 |
| 3 | .321 | 3.886 | .00 | .18 | .03 | .00 |
| 4 | .248 | 4.415 | .00 | .80 | .01 | .00 |
| 5 | .015 | 18.199 | .00 | .00 | .00 | .00 |
| 6 | 1.348E-7 | 5993.179 | 1.00 | .00 | .06 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | |
| Model | Dimension | Variance Proportions | |
| RATIOCIT | CLAIMS |
| 1 | 1 | .00 | .01 |
| 2 | .00 | .03 |
| 3 | .00 | .85 |
| 4 | .01 | .10 |
| 5 | .98 | .00 |
| 6 | .01 | .00 |

|  |
| --- |
| a. Dependent Variable: CRECEIVEln |

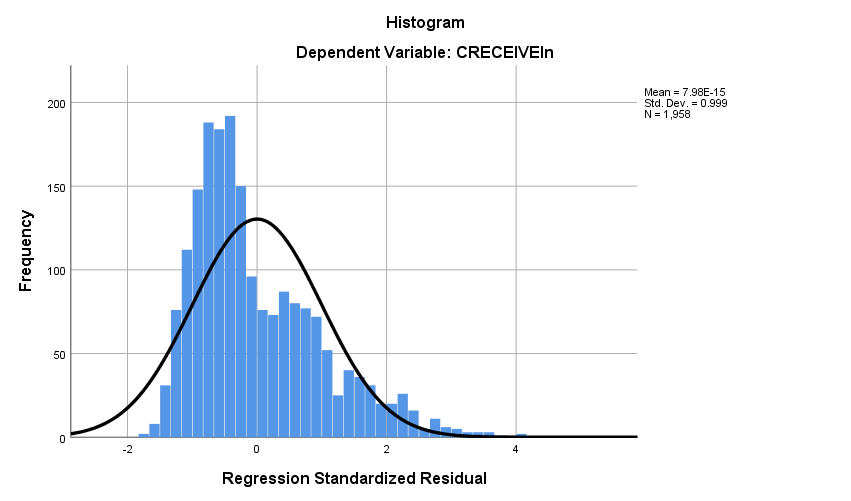
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Casewise Diagnosticsa** | | | | |
| Case Number | Std. Residual | CRECEIVEln | Predicted Value | Residual |
| 159 | 3.367 | 2.20 | .2046 | 1.99260 |
| 198 | 3.034 | 2.48 | .6894 | 1.79550 |
| 243 | 3.559 | 3.14 | 1.0295 | 2.10601 |
| 269 | 3.441 | 2.64 | .6028 | 2.03628 |
| 446 | 3.513 | 3.22 | 1.1397 | 2.07915 |
| 457 | 4.068 | 3.43 | 1.0264 | 2.40757 |
| 776 | 3.046 | 2.71 | .9052 | 1.80282 |
| 832 | 3.199 | 2.64 | .7461 | 1.89295 |
| 859 | 3.632 | 2.56 | .4157 | 2.14925 |
| 879 | 3.013 | 3.26 | 1.4748 | 1.78326 |
| 1058 | 3.130 | 1.95 | .0934 | 1.85254 |
| 1063 | 4.125 | 3.64 | 1.1964 | 2.44122 |
| 1116 | 3.895 | 3.09 | .7859 | 2.30512 |
| 1220 | 3.375 | 3.50 | 1.4990 | 1.99752 |
| 1347 | 3.126 | 2.83 | .9832 | 1.85002 |
| 1379 | 3.251 | 2.48 | .5613 | 1.92362 |
| 1680 | 3.815 | 3.26 | 1.0002 | 2.25791 |
| 1744 | 3.179 | 2.77 | .8914 | 1.88122 |

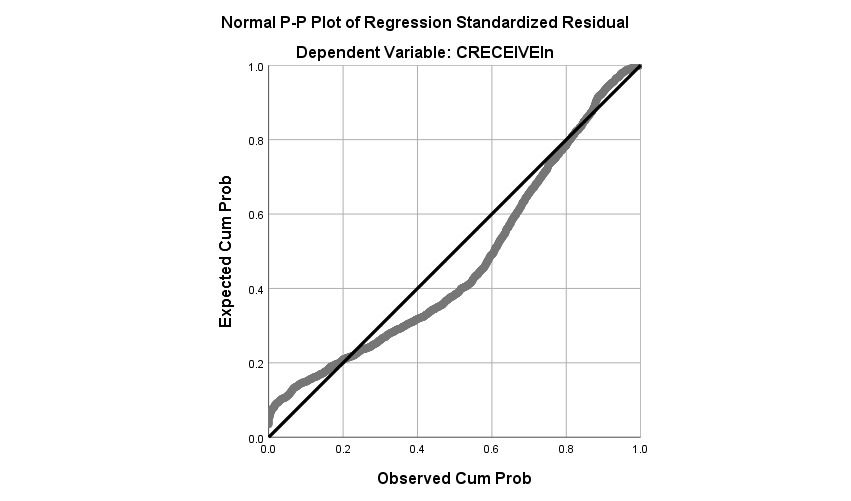
|  |
| --- |
| a. Dependent Variable: CRECEIVEln |

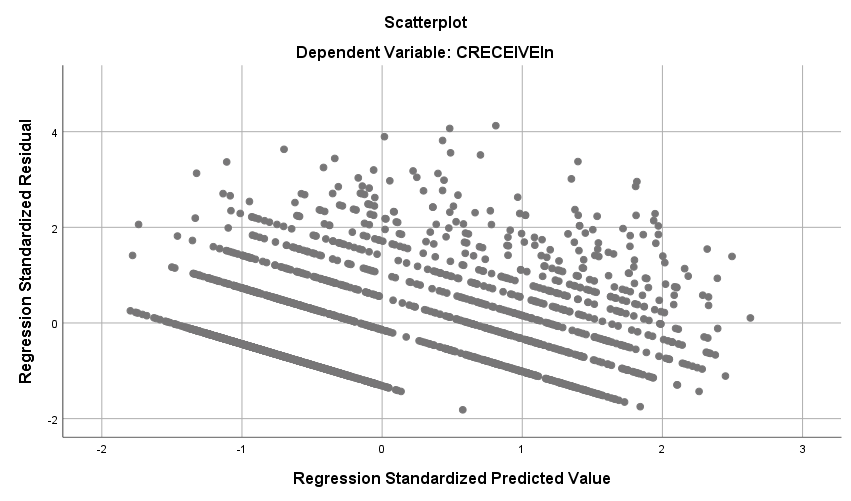
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | -.1505 | 2.1346 | .7773 | .51654 | 1958 |
| Residual | -1.07455 | 2.44122 | .00000 | .59102 | 1958 |
| Std. Predicted Value | -1.796 | 2.628 | .000 | 1.000 | 1958 |
| Std. Residual | -1.816 | 4.125 | .000 | .999 | 1958 |

|  |
| --- |
| a. Dependent Variable: CRECEIVEln |

**Charts**







**Regression**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 30-OCT-2018 18:57:59 |
| Comments | |  |
| Input | Data | D:\SOC6100\Assignments\Assignment03\Data\DataClean\Townes\_SOC6100\_Assignment03\_Data.sav |
| Active Dataset | DataSet1 |
| Filter | CRECEIVE<40 AND CLAIMS<90 (FILTER) |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 1996 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| Cases Used | Statistics are based on cases with no missing values for any variable used. |
| Syntax | | REGRESSION  /DESCRIPTIVES MEAN STDDEV CORR SIG N  /MISSING LISTWISE  /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL CHANGE ZPP  /CRITERIA=PIN(.05) POUT(.10)  /NOORIGIN  /DEPENDENT CLAIMS  /METHOD=ENTER GENERAL GYEAR ORIGINAL  /SCATTERPLOT=(\*ZRESID ,\*ZPRED)  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)  /CASEWISE PLOT(ZRESID) OUTLIERS(3). |
| Resources | Processor Time | 00:00:04.10 |
| Elapsed Time | 00:00:01.87 |
| Memory Required | 4480 bytes |
| Additional Memory Required for Residual Plots | 648 bytes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptive Statistics** | | | |
|  | Mean | Std. Deviation | N |
| CLAIMS | 14.97 | 11.689 | 1958 |
| GENERAL | .194597 | .2589553 | 1958 |
| GYEAR | 1996.27 | 1.075 | 1958 |
| ORIGINAL | .398282 | .2739932 | 1958 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Correlations** | | | | | |
|  | | CLAIMS | GENERAL | GYEAR | ORIGINAL |
| Pearson Correlation | CLAIMS | 1.000 | .056 | .039 | .101 |
| GENERAL | .056 | 1.000 | -.238 | .169 |
| GYEAR | .039 | -.238 | 1.000 | .017 |
| ORIGINAL | .101 | .169 | .017 | 1.000 |
| Sig. (1-tailed) | CLAIMS | . | .006 | .042 | .000 |
| GENERAL | .006 | . | .000 | .000 |
| GYEAR | .042 | .000 | . | .225 |
| ORIGINAL | .000 | .000 | .225 | . |
| N | CLAIMS | 1958 | 1958 | 1958 | 1958 |
| GENERAL | 1958 | 1958 | 1958 | 1958 |
| GYEAR | 1958 | 1958 | 1958 | 1958 |
| ORIGINAL | 1958 | 1958 | 1958 | 1958 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | ORIGINAL, GYEAR, GENERALb | . | Enter |

|  |
| --- |
| a. Dependent Variable: CLAIMS |
| b. All requested variables entered. |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | |
| R Square Change | F Change | df1 |
| 1 | .119a | .014 | .013 | 11.615 | .014 | 9.362 | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Summaryb** | | | |
| Model | Change Statistics | | |
| df2 | Sig. F Change |  |
| 1 | 1954 | .000 | 1.963 |

|  |
| --- |
| a. Predictors: (Constant), ORIGINAL, GYEAR, GENERAL |
| b. Dependent Variable: CLAIMS |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 3789.155 | 3 | 1263.052 | 9.362 | .000b |
| Residual | 263622.516 | 1954 | 134.914 |  |  |
| Total | 267411.672 | 1957 |  |  |  |

|  |
| --- |
| a. Dependent Variable: CLAIMS |
| b. Predictors: (Constant), ORIGINAL, GYEAR, GENERAL |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | -1073.630 | 502.890 |  | -2.135 | .033 |
| GENERAL | 2.390 | 1.061 | .053 | 2.253 | .024 |
| GYEAR | .544 | .252 | .050 | 2.161 | .031 |
| ORIGINAL | 3.894 | .974 | .091 | 3.998 | .000 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | |
| Model | | 95.0% Confidence Interval for B | | Correlations | | | Collinearity Statistics |
| Lower Bound | Upper Bound | Zero-order | Partial | Part | Tolerance |
| 1 | (Constant) | -2059.888 | -87.372 |  |  |  |  |
| GENERAL | .310 | 4.471 | .056 | .051 | .051 | .913 |
| GYEAR | .050 | 1.038 | .039 | .049 | .049 | .940 |
| ORIGINAL | 1.984 | 5.805 | .101 | .090 | .090 | .968 |

|  |  |  |
| --- | --- | --- |
| **Coefficientsa** | | |
| Model | | Collinearity Statistics |
| VIF |
| 1 | (Constant) |  |
| GENERAL | 1.095 |
| GYEAR | 1.064 |
| ORIGINAL | 1.033 |

|  |
| --- |
| a. Dependent Variable: CLAIMS |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | | | |
| Model | | | ORIGINAL | GYEAR | GENERAL |
| 1 | Correlations | ORIGINAL | 1.000 | -.060 | -.178 |
| GYEAR | -.060 | 1.000 | .244 |
| GENERAL | -.178 | .244 | 1.000 |
| Covariances | ORIGINAL | .949 | -.015 | -.184 |
| GYEAR | -.015 | .063 | .065 |
| GENERAL | -.184 | .065 | 1.125 |

|  |
| --- |
| a. Dependent Variable: CLAIMS |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
| (Constant) | GENERAL | GYEAR | ORIGINAL |
| 1 | 1 | 3.237 | 1.000 | .00 | .03 | .00 | .02 |
| 2 | .531 | 2.468 | .00 | .91 | .00 | .02 |
| 3 | .232 | 3.737 | .00 | .00 | .00 | .95 |
| 4 | 1.362E-7 | 4874.267 | 1.00 | .06 | 1.00 | .00 |

|  |
| --- |
| a. Dependent Variable: CLAIMS |

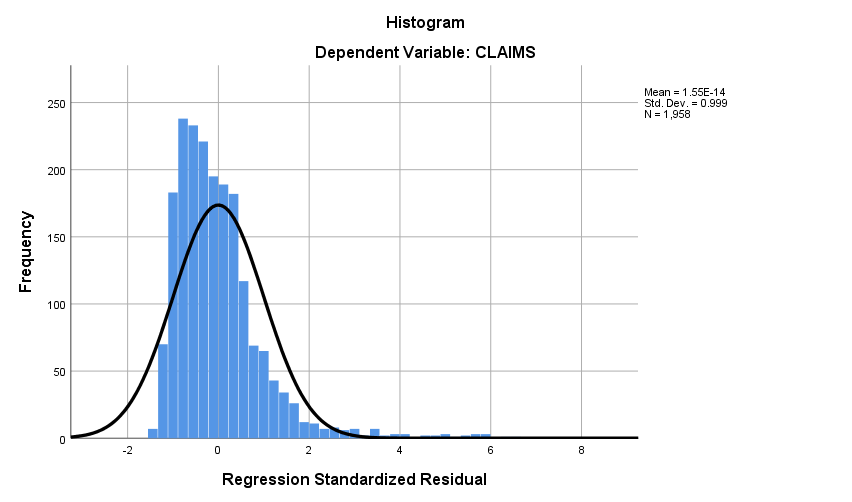
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Casewise Diagnosticsa** | | | | |
| Case Number | Std. Residual | CLAIMS | Predicted Value | Residual |
| 33 | 5.454 | 79 | 15.65 | 63.351 |
| 102 | 6.520 | 88 | 12.27 | 75.732 |
| 114 | 4.485 | 70 | 17.90 | 52.096 |
| 117 | 3.466 | 57 | 16.75 | 40.253 |
| 149 | 4.994 | 72 | 14.00 | 58.002 |
| 187 | 3.021 | 53 | 17.91 | 35.092 |
| 255 | 4.983 | 73 | 15.12 | 57.880 |
| 290 | 4.797 | 70 | 14.28 | 55.722 |
| 382 | 4.118 | 61 | 13.16 | 47.836 |
| 501 | 3.081 | 49 | 13.21 | 35.789 |
| 515 | 3.931 | 60 | 14.34 | 45.663 |
| 541 | 4.158 | 65 | 16.70 | 48.301 |
| 605 | 5.011 | 74 | 15.79 | 58.210 |
| 668 | 3.624 | 56 | 13.90 | 42.099 |
| 766 | 3.064 | 52 | 16.41 | 35.589 |
| 794 | 3.617 | 58 | 15.99 | 42.008 |
| 860 | 3.781 | 60 | 16.08 | 43.918 |
| 868 | 5.375 | 79 | 16.57 | 62.434 |
| 951 | 4.136 | 65 | 16.96 | 48.045 |
| 1010 | 3.094 | 51 | 15.06 | 35.939 |
| 1024 | 5.913 | 84 | 15.32 | 68.684 |
| 1029 | 3.404 | 54 | 14.46 | 39.544 |
| 1128 | 5.297 | 79 | 17.48 | 61.522 |
| 1160 | 5.824 | 81 | 13.36 | 67.644 |
| 1210 | 3.807 | 58 | 13.78 | 44.217 |
| 1248 | 4.764 | 73 | 17.67 | 55.331 |
| 1272 | 3.374 | 55 | 15.81 | 39.186 |
| 1336 | 3.457 | 55 | 14.84 | 40.160 |
| 1381 | 6.392 | 89 | 14.76 | 74.241 |
| 1451 | 3.288 | 51 | 12.81 | 38.188 |
| 1461 | 3.367 | 54 | 14.89 | 39.106 |
| 1480 | 5.612 | 78 | 12.81 | 65.188 |
| 1507 | 5.649 | 82 | 16.38 | 65.616 |
| 1557 | 3.114 | 53 | 16.84 | 36.165 |
| 1655 | 5.557 | 78 | 13.46 | 64.544 |
| 1671 | 3.460 | 53 | 12.81 | 40.188 |
| 1774 | 5.887 | 86 | 17.63 | 68.374 |
| 1884 | 3.418 | 55 | 15.30 | 39.696 |
| 1991 | 4.627 | 70 | 16.25 | 53.747 |
| 1994 | 3.018 | 50 | 14.95 | 35.055 |

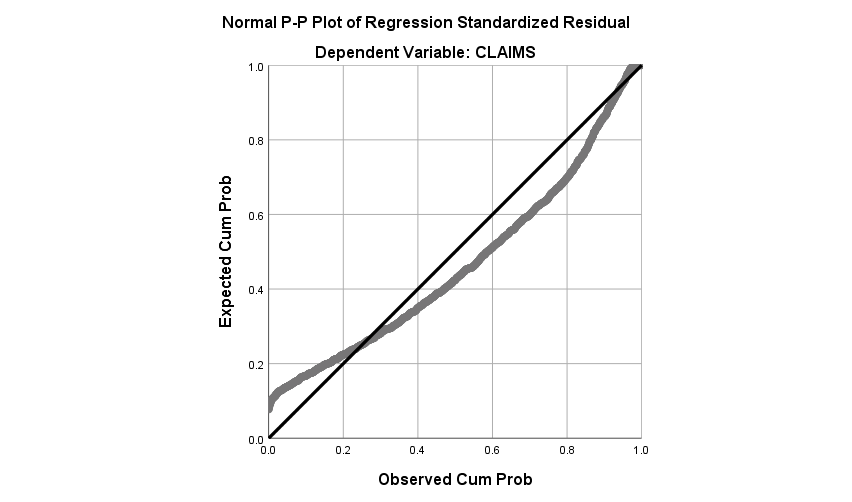
|  |
| --- |
| a. Dependent Variable: CLAIMS |

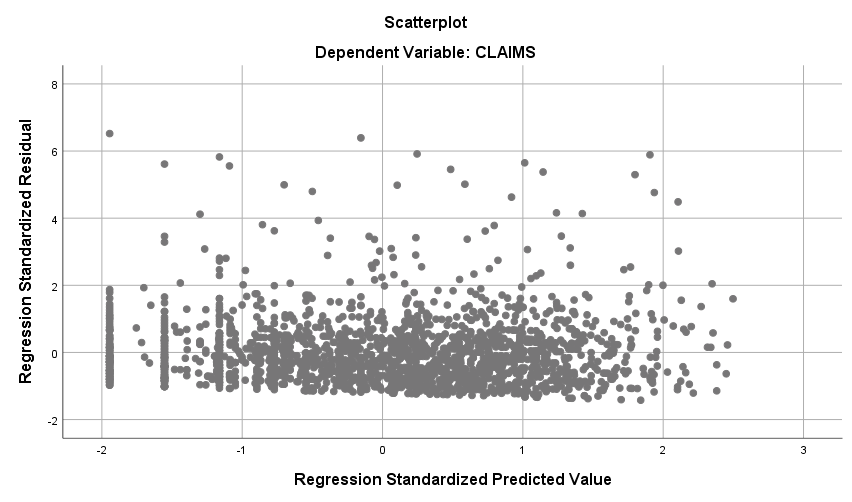
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | 12.27 | 18.45 | 14.97 | 1.391 | 1958 |
| Residual | -16.534 | 75.732 | .000 | 11.606 | 1958 |
| Std. Predicted Value | -1.945 | 2.498 | .000 | 1.000 | 1958 |
| Std. Residual | -1.423 | 6.520 | .000 | .999 | 1958 |

|  |
| --- |
| a. Dependent Variable: CLAIMS |

**Charts**







Appendix D. 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 |