5-2014

Geographic Variation of Health Care Spending on Heart Failure in Metropolitan Areas

Kevin McMillan
University of Nebraska-Lincoln, kevin.mcmillan@huskers.unl.edu

GEOGRAPHIC VARIATION OF HEALTH CARE SPENDING ON HEART FAILURE IN METROPOLITAN AREAS

By

Kevin McMillan

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Master of Arts

Major: Geography

Under the Supervision of Professor James W. Merchant

Lincoln, Nebraska

May, 2014

Geographic Variation of Health Care Spending on Heart Failure in Metropolitan Areas

Kevin McMillan, M.A.

University of Nebraska, 2014

Advisor: James Merchant

The costs of healthcare have long been a concern in the United States. It is well known that these costs vary geographically, but attempts to explain this variation have been met with limited and varied success. This is partly attributable to the fact that data available have restricted analyses to assessing the issue to using Medicare cost per beneficiary. In June, 2013, the Center for Medicare and Medicaid Services (CMS) released new Medicare data that detailed the charges and payments made to hospitals throughout the United States in 2011. In this thesis, this new dataset was used to examine costs of treatment for heart failure, a widespread and serious health concern in the U.S.. Costs were examined from two perspectives: the average Medicare payments and the average amount hospitals charged within metropolitan areas. Ordinary Least Squares (OLS) regression analysis was used in an attempt to explain geographic variation in the Medicare payments for the treatment of heart failure in Metropolitan Statistical Areas (MSAs) based on six key demographic variables identified from previous research on spending per Medicare beneficiary. Additionally, these demographic variables were applied to the average amount hospitals charged for the treatment of heart failure. These

six variables include the percent African American, the percent with a Bachelor's degree or higher, the average number of hospital discharges for heart failure, the percent female, the percent in poverty, and the percent aged sixty-five and older. Results suggest that variables with a key relationship to Medicare payments for the treatment for heart failure include the percent with a Bachelor's degree or higher and the percent aged sixty-five or older within a MSA. Key variables correlated with the average amount hospitals charged for the treatment of heart failure in a MSA include the average number of hospital discharges for heart failure, the percent female, and the percent aged sixty-five and older within a MSA.

Acknowledgements

I would like to express my appreciation for all of my committee members, Dr. James Merchant, Dr. Clark Archer, Dr. Yunwoo Nam, and Dr. Ge Lin. Their suggestions, input, advice, and encouragement were invaluable to the completion of this thesis. I would also like to thank the University of Nebraska for the opportunity to further my education in graduate school and for the opportunities presented during my time as a student, teacher, and researcher. This includes Dr. Zhenghong Tang for all of the research and projects we have worked on together, as well as many of the faculty and staff in the School of Natural Resources for their continued support. Finally, I would like to thank my friends and family for their encouragement and support during this process, without which this thesis would not be complete – especially Kristin Sorensen, Katelyn Pietz, and David Grosso. I would not have been successful without all of you helping me along the way.

Table of Contents

List of Figures and Tables

Chapter 1 - Introduction

Costs of health care have long been a concern in the United States. It has often been noted that the costs of health care vary geographically, however the causes and consequences of this variation are not well understood (Fuchs et. al. 2001; Figure 1.1). In the U.S., health care costs are often reported by Diagnosis Related Groups (DRGs), which represent groups of similar diseases/treatments used by Medicare for payment purposes. These "costs" are reported as Medicare payments determined based on a formula (detailed in Chapter 2). These payments are a reflection of the services rendered and the various cost associated with providing those services (e.g., rent, malpractice insurance, supplies, amount of work involved in providing treatment). Maps of health care costs reported by DRG show geographic disparities similar to those observed when assessing health care costs reported by costs per beneficiary (Figure 1.2; Figure 1.3). Since such maps show similar patterns of costs, a reasonable assumption is that the drivers of health care costs (e.g., demographic variables) may be similar, yet this has not been proven.

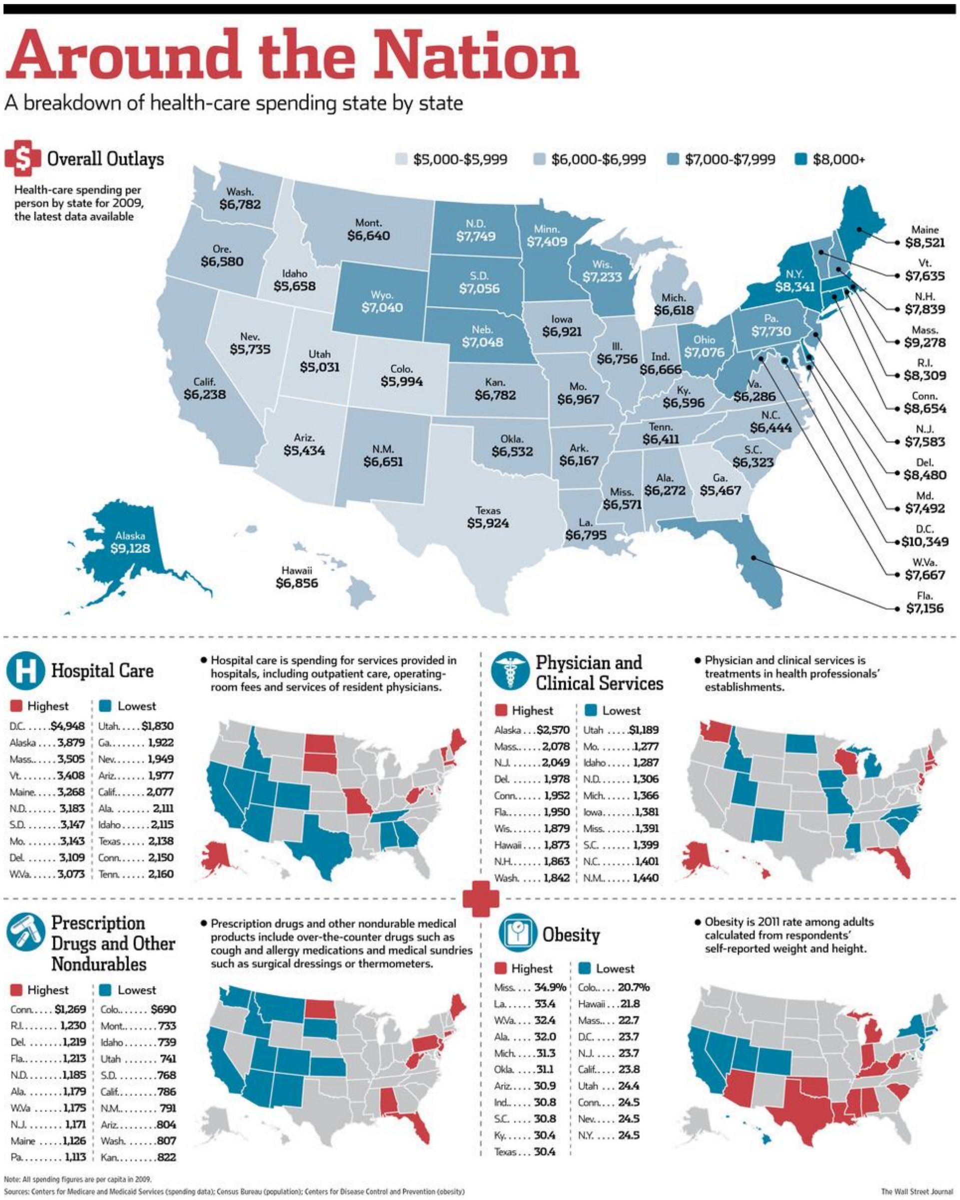


Figure 1.1: Comparison between health care spending per person and potential contributing factors (Radnofsky 2014)

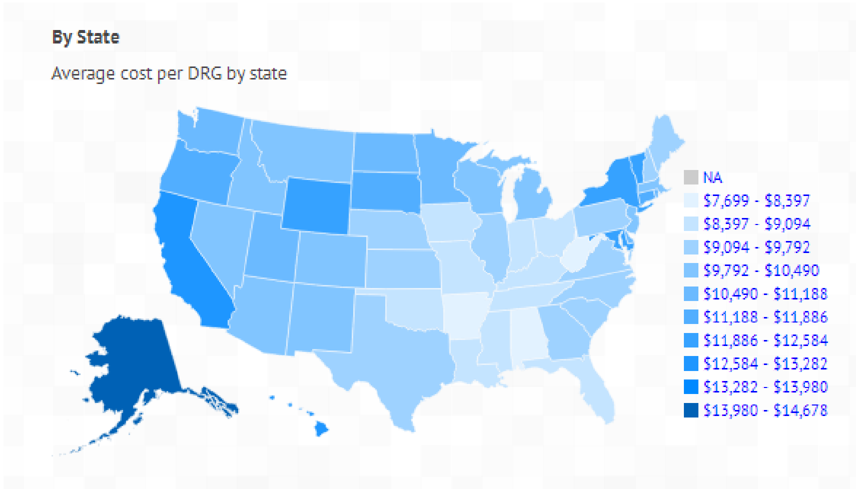


Figure 1.2: Variations of costs per Diagnosis Related Group (DRG) (Tera Insights 2013*)*

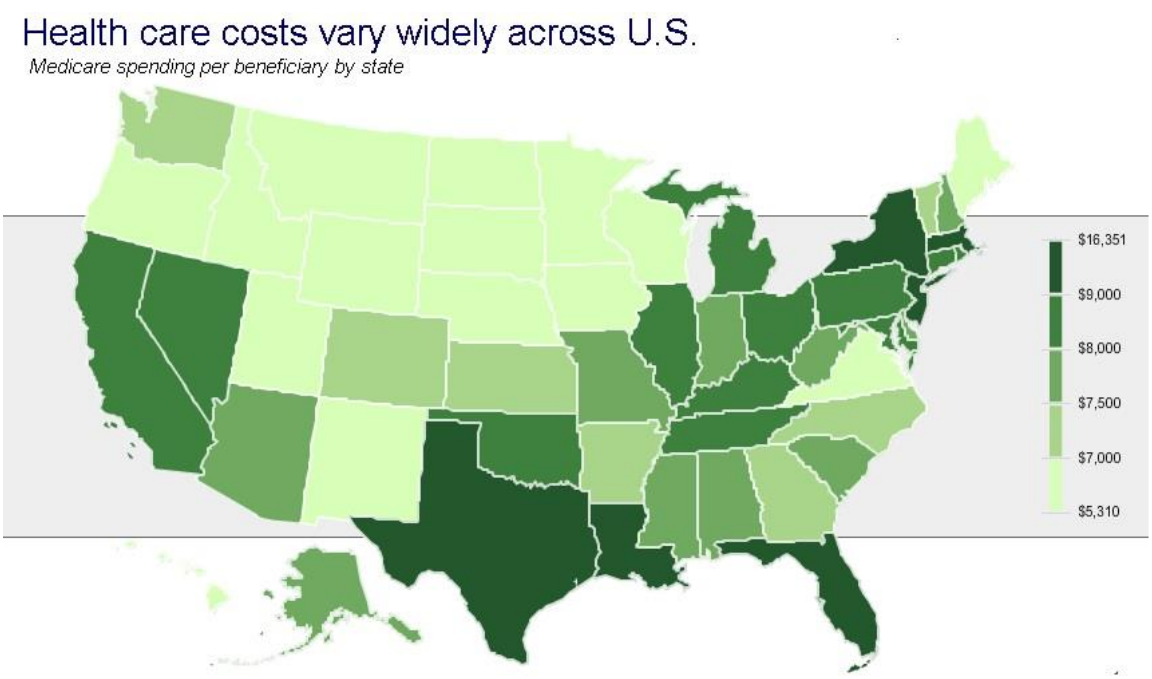


Figure 1. 3: Variations of costs by cost per beneficiary by state (NBC News 2014)

Some researchers have found geographic disparities in health care costs (calculated by Medicare on a cost per beneficiary basis) to be correlated with variables such as age, race, income, and/or education of patients in a particular area as well as patients' access to primary care providers (Fuchs et. al. 2001, Cutler and Sheiner 1999, Gage et. al. 1999, Welch et. al. 1993). Although such studies provide insight into geographic variations in health care costs, most researchers acknowledge that a full explanation of geographic disparities in health care costs requires additional research. Studies using finer data, especially those that examine geographic variation in costs associated with specific procedures (e.g., treatment of heart failure) and investigations of additional explanatory variables are needed. Such research will provide a more robust understanding of the factors that drive the geographic variation in health care costs.

It is noteworthy that research on geographic variation in health care costs has almost always been based upon the cost per Medicare beneficiary because such data are readily available. Medicare cost per beneficiary (often called "aggregate costs") data is an aggregate total of Medicare spending divided by the number of beneficiaries in an area. Additionally, Medicare cost per beneficiary data do reflect the amount hospitals charged Medicare for treatment and is simply a measure of Medicare payments. The amount the hospitals charged for treatment (hospital charges) would presumably remain constant regardless of the patient or insurance type and may be a better indicator of cost compared to Medicare payments for that same treatment. Therefore, studies that examine both Medicare payments and the amount hospitals charged Medicare for services rendered are needed since they represent different (but complementary) measures of "cost", and therefore may reveal similar or additional explanations for the geographic variations in health care costs.

A new dataset available from the Centers for Medicare and Medicaid Services (CMS) appears to be useful for addressing the limitations of data employed in previous studies, such as providing data by hospital rather than larger areal units (Centers for Medicare and Medicaid Services 2013). The CMS data detail both actual hospital charges (i.e., the amount each hospital charged Medicare for treatment and services) and Medicare payments for the top 100 Diagnosis Related Groups (DRGs) at hospitals throughout the United States. As a consequence, it is now possible to assess the amount charged for specific medical procedures. The CMS data also includes hospital discharge data, which represents the number of patients that were treated and released. These data make it feasible to better identify variables that contribute to observed geographic variation in health care costs.

Research Objectives

The principal objectives of this thesis are (1) to determine how health care costs (Medicare payments and the amount the hospital charged) for the treatment of heart failure vary geographically across the major metropolitan areas of the U.S, (2) to assess whether geographic variation in Medicare payments and hospital charges for heart failure treatment can be explained by variables found in relevant previous research based on Medicare cost per beneficiary, and (3) to develop a suitable Geographic Information System (GIS) methodology for assessing these variations. Socio-demographic factors such as patient race, age, income, and education found, in previous studies, to be correlated with health care costs at the cost per beneficiary level will be reassessed using the CMS hospital data. The number of discharges, the number of patients that were admitted and treated for heart failure that survived and were released from the hospital, will be used to determine if increased utilization of services results in higher charges or payments as previously indicated by the literature.

This thesis will focus on heart failure, a common and serious condition that has been a major concern for health officials for several decades (Figure 1.3)

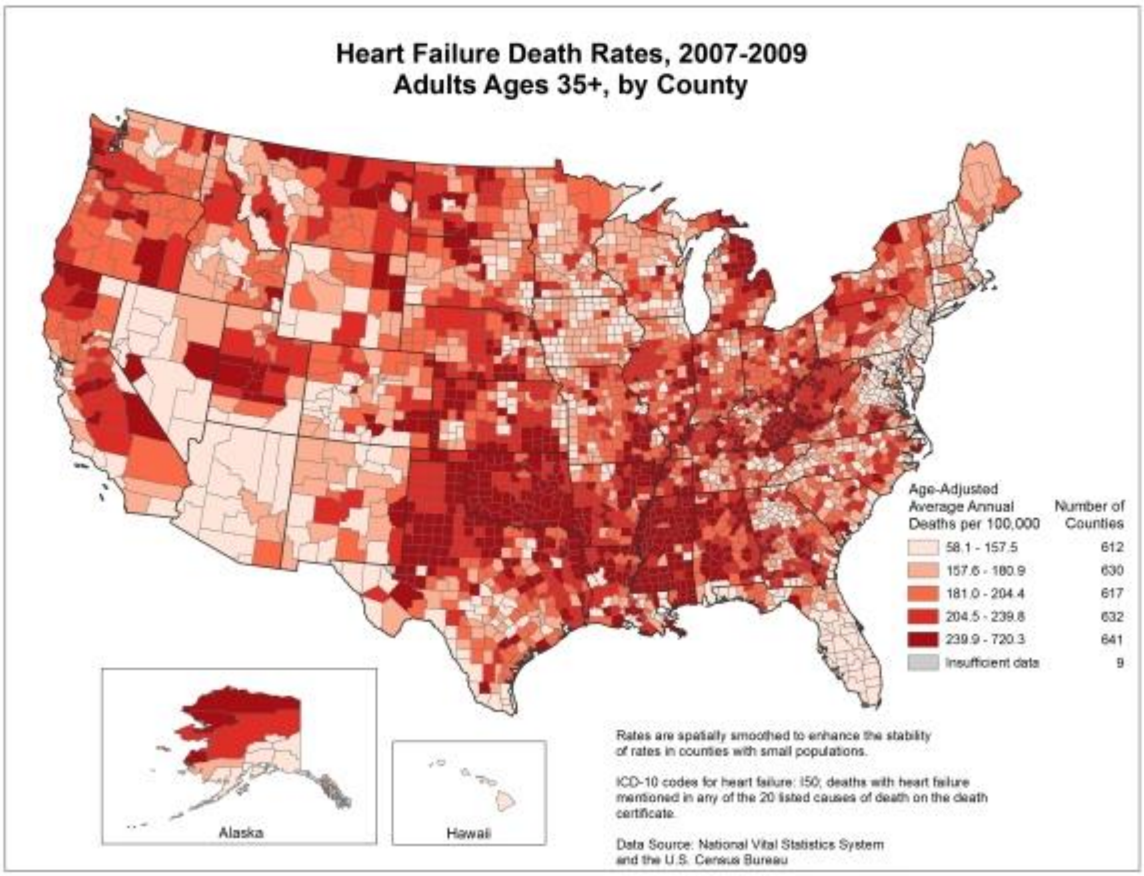


Figure 1.4: Heart failure death rates in the US (Centers for Disease Control 2010)

The Centers for Disease Control (2010) estimates that heart failure had a mortality rate, the rate at which deaths occur in a population, of 25.6 per hundred thousand in 2006. The number one cause of death, heart disease, is the primary cause of heart failure and 1 in 9 deaths are related to heart failure (Centers for Disease Control 2010). This research will be restricted to analyzing data on heart failure without major complication or comorbidity or comorbid conditions (DRG 293). Comorbidity indicates the presence of some other disease (a “comorbid” condition), thus restricting the study to heart failure without other diseases present will reduce opportunities for other diseases to influence the study results (i.e., results will only be related to heart failure). Extensive data on the costs of treatment for heart failure are available in the CMS dataset.

Heart failure is distinguished by being a malady that tends to be associated with socio-demographic factors rather than environmental factors that affect the incidence (the number of new cases) of many other diseases (Pneumonia, for example, exhibits clusters in colder regions). Thus, studying the costs of treating heart failure is expected to shed new light on the variables that influence geographic variation in costs. Finally, it is noteworthy that heart failure is a typical end result of many cardiovascular diseases, such as heart disease, and the results of an analysis of heart failure will likely be applicable to other cardiovascular diseases.

Study Area

This study will focus on the 356 Metropolitan Statistical Areas (MSAs) in the continental United States that have hospitals that received a Medicare payment for the treatment of heart failure in 2011. MSAs in Alaska, Hawaii, and those with no hospitals that received a Medicare payment for heart failure were excluded. These MSAs are defined by the Office of Management and Budget (OMB) to include a core city and the surrounding area determined to have significant interaction with each city (United States

Census Bureau 2013). MSAs were chosen for three reasons. First, over 80% of people in the United States live in urban areas, which will allow this study to encompass a large majority of the U.S. population (United States Census Bureau 2012). Second, nearly all of the MSAs in the continental U.S. have at least one hospital that treats heart failure, while many counties do not. Finally, extensive ancillary data describing socio-demographic characteristics are available for MSAs, which is a major focus of this study.

Research Methodology – An overview

The 2011 CMS dataset will be used to determine which hospitals that treat heart failure are within each MSA, the average number of discharges (patients treated for heart failure and released) within each MSA, the average amount hospitals charged for heart failure treatment within each MSA, and the average amount Medicare paid hospitals for the treatment of heart failure within each MSA. The difference between the average hospital charges and average Medicare payments will also be calculated. The difference is calculated as part of a supplementary analysis that will evaluate the potential for the difference to be used as a third complementary measure of cost. American Community Survey (ACS) data from 2011 will be used to estimate demographic factors for each MSA (e.g., age, sex, race, income, and education). Most data analysis will be conducted using ESRI ArcGIS version 10.1, a proprietary Geographic Information System (GIS). The CMS data will initially be imported into Microsoft Excel, and hospital charges, Medicare payments, and related information such as the number of discharges for heart failure and shock without CCs and MCCs will be isolated. The data will then be imported into ArcGIS for the remainder of the research. The CMS data will first be geocoded and aggregated into MSAs so that they are compatible with the ACS data, which are already reported by MSA. The ACS data will be spatially joined to the CMS data and Ordinary Least Squares (OLS) regression will be performed using explanatory variables that were identified as being correlated with costs in previous research (e.g., utilization, age, sex, race, income, and education) as independent variables and Medicare payments, and the difference between hospital charges and Medicare payments, as dependent variables. The difference between hospital charges and Medicare payments may also be a suitable measure of cost and help discover reasons behind the disparity between these measures of costs.

Implications of the Research

A better understanding of the geographic variation of health care costs is required in order to address issues associated with the increasing costs of medical care. Beginning in 2013, the Patient Protection and Affordable Care Act (PPACA) began to introduce many changes to the U.S health care system. The results of the research conducted for this thesis will facilitate future analyses designed to determine the effects of these changes. It seems likely that additional changes will need to be made as the PPACA begins to take effect and problems regarding costs are identified. To solve these problems, it is necessary to have a better understanding of factors that influence geographic variations in costs. This study helps form a basis in working with DRG data in analyses of geographic variations in health care costs and provides a working methodology from which other research can be based.

Chapter 2 – Background and Literature Review

Introduction

In this chapter, previous research on the geographic variation in health care costs will be summarized and reviewed. In addition, background information on heart failure, health care costs, and Medicare payments will be provided. The topics relevant to this research cover a broad spectrum; hence, this chapter cannot be exhaustive. The objective here is to provide a foundation so that the reader can understand the goals, terminology, analytical procedures, and implications of the research. As stated in Chapter 1, a key objective of the research is to assess variables (e.g. socio-demographic factors) that may explain the observed geographic variations in health care costs. In that regard, it should be noted that, in the literature on health care costs, such variables are often characterized as "supply variables", "demographic variables", and "lifestyle variables". Supply variables may include the number of hospital beds, number of primary care physicians, and costs of various services. Demographic variables include age, sex, race, income, and others. Lifestyle variables may include smoking rates, obesity rates, and stroke rates.

Heart Failure

Heart failure occurs when the heart can no longer pump enough blood and oxygen to support the body's organs (Centers for Disease Control 2010). The disease is common in the United States where, currently, approximately 5.7 million people are afflicted with heart failure. Heart failure is the primary cause of about 55,000 deaths per year, although heart failure does not always result in death (Centers for Disease Control 2010). Only about half of those with heart failure die within five years of diagnosis and can be treated using a variety of lifestyle changes, medical procedures, and medication (Centers for Disease Control 2010). Any disease or behavior that can damage the heart and restrict blood flow can increase the risk of heart failure.

Health Care Costs

Health care costs in the United States have been increasing dramatically in recent years. Since 1960, spending on health care as a percentage of the U.S. GDP has more than tripled from about 5.5% to 16% (Figure 2.1).

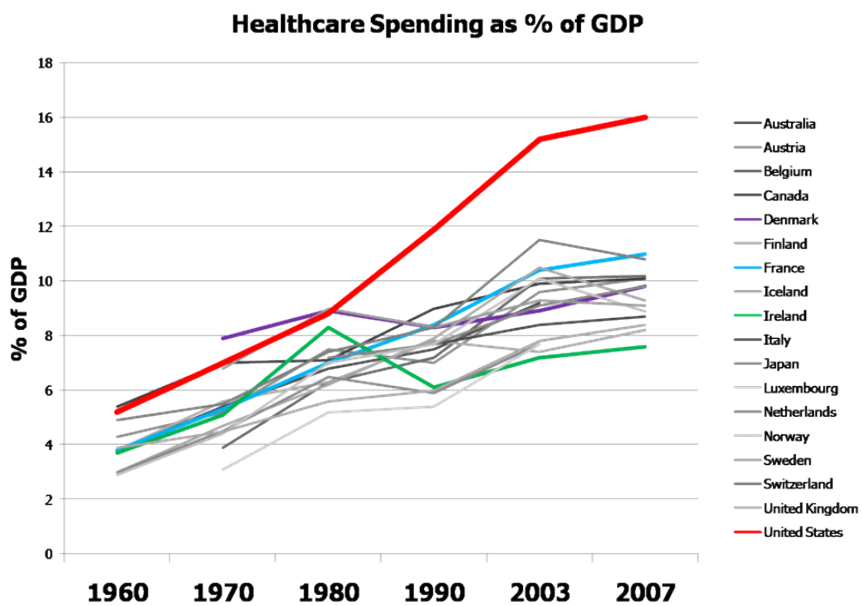


Figure 2.1: U.S. health care spending as a percentage of GDP compared to selected countries (Shook 2009)

On average, the nation now spends over $8,000 per person per year on medical care, nearly three times the average of the thirty-four nations comprising the Organization for Economic Co-operation and Development (OECD) average (Figure 2.2; Kane 2012).

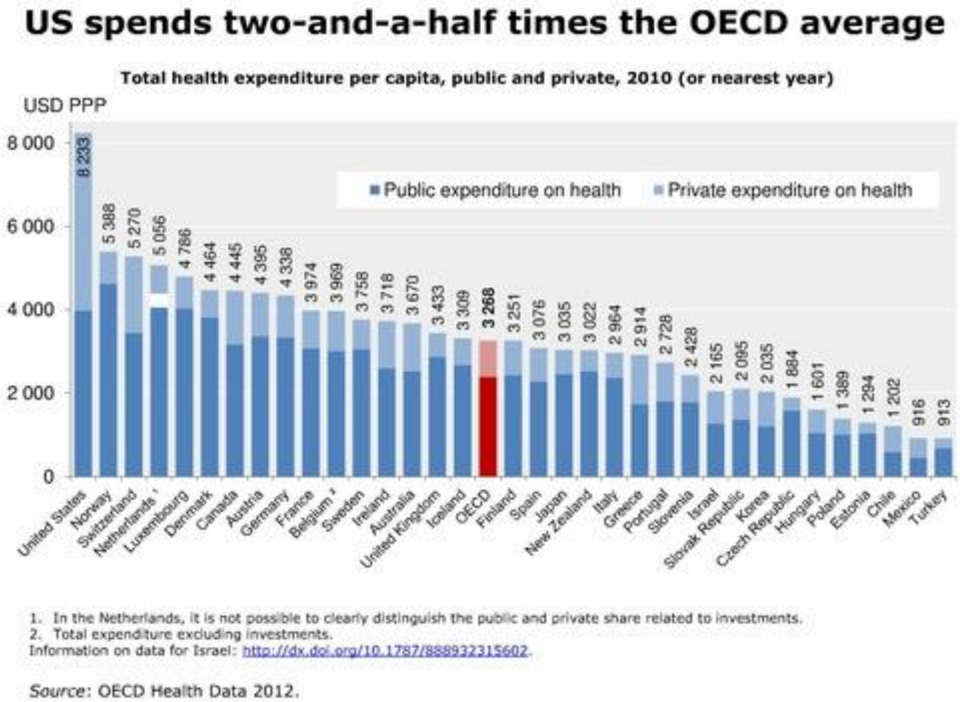


Figure 2.2: U.S. expenditure on health care compared to other OECD countries (Kane 2012)

While many in the US have insurance that assists with these costs, insurance premiums have gone up as well and have far outpaced salary increases. Between 1999 and 2009, the average U.S. salary increased 38%, but health care premiums in the same time period rose 131% (Institute of Medicine 2013). Health insurance also does not protect patents from the rising costs of health care. In 2007, at least 62.1% of all bankruptcies were related to medical debt, a rise of nearly 50% compared to 2001 (Himmelstein et. al. 2009). While conventional wisdom would suggest that those without insurance are at the greatest risk of financial hardship due to medical costs, Himmelstein (et. al. 2009) noted that about 75% of those that had bankruptcies related to medical costs actually had health insurance. Given the rise of premiums for already expensive health care and the lack of financial security a major health issue may bring, cost has become a major factor for many people in the United States.

Medicare Payments

The Medicare system provides insurance to individuals in the United States over 65 as well as some disabled persons under the age of 65. The system is comprised of four parts ranging from hospitalization to drug coverage. Part A covers hospitalizations, hospice and nursing, and a limited number of other health services. Part B covers necessary medical supplies as well as preventative care. Part C includes Medicare Advantage Plans, which are run by private insurers and cover everything that Medicare covers, although some plans in Part C may cover health services not covered by traditional Medicare. Part D is a collection of prescription medication coverage plans, each of which cover different medication. Data used in this study are derived from Medicare part A, and relates to patients that were admitted to a hospital with heart failure. These patients were treated on an impatient basis, which means that they were admitted based on a doctor’s direct order.

The Medicare system uses a formula, termed the Medicare Physician Fee Schedule (PFS), to determine payments to hospitals. The PFS equation is:

$$Payment = [(Work\ RVU * Work\ GPCI) + (PE\ RVU * PE\ GPCI) + (MP\ RVU * MP\ GPCI)] * CF$$

Where:

RVU = Relative Value Units

PE = Practice Expense

MP = Malpractice

GPCI = Geographic Practice Cost Indices

CF = Conversion Factor

RVUs consist of three factors: work, practice expenses, and malpractice. Work RVUs reflect the cost of the time and intensity (the amount of work involved in treatment), and account for about half of the total payment. Practice Expenses (PE) involve the costs of maintaining a particular practice, such as rent, equipment, and supplies. Malpractice (MP) RVUs reflect the malpractice costs (e.g. insurance) for medical procedures. Medicare also uses Geographic Practice Cost Indices (GPCIs) to calculate payments. These are applied to all three of the above payment variables (Work, PE, MP) based on the cost of practicing medicine in different areas of the country. The conversion factor (CF) is a constant dollar amount determined by the U.S. Congress based on inflation and the growth in demand for medical care.

Previous Research on Geographic Variability of Health Care Costs

Many previous studies have focused on the geographic variability of health care costs. Most have included attempts to identify variables that may help explain observed variations in Medicare cost per beneficiary, referred to as aggregate costs within this review. Welch et. al (1993), for example, studied aggregate health care costs in metropolitan areas of the U.S. using 1989 Medicare spending per beneficiary. Hospital admission rates, payments to physicians for inpatient services per admission and per beneficiary, payments to physicians for outpatient services per beneficiary, and overall payments to physicians per beneficiary were assessed as possible explanatory variables. They found that areas with high rates of admission had high costs per admission, and areas with high payments for inpatient services tended to have high payments for outpatient services (Welch et. al. 1993). While they found that costs were not related to the number of physicians per capita, they noted that costs were lower in areas with higher proportions of primary care physicians. These findings show that areas with high costs have high costs for a variety of services. Additionally, and perhaps more importantly, the fact that a higher number of primary care physicians results in lower costs shows that supply factors are also significant. Welch and his colleagues only examined the supply side of health care and failed to examine any socio-demographic factors. However, the study did examine admission rates. Such a variable should be similar to the number of discharges, since the number of discharges would be equal to the number of admissions minus any deaths. This study implies that discharges may be an explanatory variable for the costs of health care; however, an additional examination of discharges as a possible explanatory variable for the costs of health care is needed.

Cutler and Sheiner (1999) attempted to explain the geographic variation in costs of health care using variation of disease organized into four broad categories: those due to behavioral factors, environmental factors, genetic factors, and random chance (defined as disease not attributable to any of the other factors), variation in demand for medical services, demographic factors, and insurance coverage. They used data from the 1998 *Dartmouth Atlas* that showed 1995 Medicare spending per beneficiary data adjusted for differences in age, sex, and race. Using Medicare spending in MSAs, they found that higher illness rates were correlated with higher costs of health care and that the explanation of the spatial variation of costs could be improved by considering certain demographics such as age, race, and ethnicity. For example, higher proportions of Hispanic elderly were correlated with a rise in costs, while higher proportions of elderly females were correlated with a steep decline in costs (Cutler and Sheiner 1999). Cutler and Sheiner (1999) also determined that the availability of medical supplies (e.g. physicians and hospital beds) was correlated with health care costs. Nevertheless, they concluded that costs were best explained by the prevalence of illnesses and they suggested that many of the other variables that correlated well with costs acted as proxies for illness rather than being explanatory in and of themselves. However, the study revealed that supply factors are as significant as demographics. This is indicative of a complex system in which there are likely many factors behind cost variation. The specific factors in play and the ways in which such factors interact remains to be determined.

Gage et. al (1999) also used 1995 Medicare spending per beneficiary to study variations in health care at the state level. They investigated age, income, and differences in services (such as in-home care, ambulatory services, and inpatient services) to determine if they were related to variations in health care. People having only Medicare coverage were compared with those who were enrolled in both Medicare and Medicaid. The results showed that the primary factor influencing the state-to-state variability of aggregate health care costs was age. States having a higher proportion of those over 85 years of age in the population had a higher cost per beneficiary (Gage et. al 1999). Additionally, those who were enrolled in both Medicare and Medicaid had higher Medicare costs than those who were enrolled solely in Medicare, and areas with a higher proportion of dually enrolled individuals also had a higher proportion of those 85 years of age and older (Gage et. al 1999). Since Medicaid is offered only to those with lower incomes, income was also found to be a significant factor. It should be noted that income may simply be a surrogate factor as it is a determinant of Medicaid eligibility. Nevertheless, it is clear that age and income should be re-evaluated at the procedural level.

In 1999, the Center for the Evaluative Clinical Sciences conducted a study of variations in aggregate health care costs based on a new atlas of health care, a follow-up to the *Dartmouth Atlas of Healthcare 1998*. The study used 1995 Medicare cost per beneficiary data aggregated to the Hospital Referral Region (HRR) level (Center for the Evaluative Clinical Sciences 1999). Hospital Referral Regions denote areas from which the majority of a region's patients come. Variables found to be related to variation in health care costs included age, sex, race, illness rates, and supply factors such as hospital beds and number of physicians (Center for the Evaluative Clinical Sciences 1999). These results are consistent with most other research on geographic variation in health care costs. While the study is an important contribution to the body of research on the variability of health care costs, HRRs can sometimes be difficult to work with as demographic data are not reported by the U.S. census in these areal units. Errors may occur when demographic data are summarized and resampled from the MSA or state level. Thus, although this study found that age and race are major variables correlated with health care costs, they should be re-examined in further studies.

Fuchs et.al (2001) carried out research using Medicare cost per beneficiary from 1989 to 1991. Mortality was used as a proxy for health care costs because most costs occur within the last few years of life (Fuchs et. al 2001). The study focused on U.S. metropolitan areas. Those over 85 were excluded because of the difficulty in obtaining information on education and income (Fuchs et. al. 2001). The investigators found that the variables that best explained differences in mortality were education, income, cigarette smoking rates, obesity rates, air pollution, and the percent African American (Fuchs et. al 2001). However, they concluded that there was still a lot of variation in costs that were not explained by these variables, and that further study was needed. Although income, education, and race were shown to be related to variability in health care costs, supply factors such as the number of physicians, specialists, and hospital beds within a given area were not examined.

A 2013 study by the Medicare Payment Advisory Commission investigated the socio-demographic factors related to geographic variation in Medicare spending (Medicare Payment Advisory Commission 2013). The Medicare Payment Advisory Commission (MedPAC) was established to advise the U.S. Congress on issues related to Medicare. Using 2009 Medicare Current Beneficiary Survey (MCBS) data, MedPAC (2013) determined that a key factor in determining costs was the age of the beneficiary.

Those 65 and older were broken into three categories: 65-74, 74-84, and 85 and older. The results showed the percentage of the population 85 and older was a factor correlated with higher costs. Costs were observed to increase as age increased (Medicare Payment Advisory Commission 2013). This study again showed that age is a primary factor correlated with the costs of health care.

Summary and Conclusions

Studies of the geographic variation in aggregate U.S. health care costs have shown that some factors are consistently correlated with the costs of health care. These include income, age, sex, race, and others. It is not clear, however, if the results of research based on cost per beneficiary applies to costs calculated for specific procedures, such as treatment of heart failure. Research is needed to identify variables that help explain procedural costs, especially at fine scales (e.g., MSAs).

It is also noteworthy that a review of previous research has revealed some problems related to the definition and classification of explanatory variables. For example, many studies have found age to be important, but the way in which age groups have been classified varies across studies. In some studies, the “elderly” are identified as those aged 85 and older, while other studies define the elderly as those 65-85, and yet others include both or combine the two age groups. Such inconsistencies can make it difficult to compare research. In the research contained within this thesis, age will be defined as those aged 65 and older in order to best encompass all of the various age groups.

Virtually all previous studies use Medicare cost per beneficiary data to estimate health care costs. The costs for specific procedures, such as the treatment of heart failure, were not considered in such studies. Research directed towards the costs of specific treatments is desperately needed.

Finally, attention needs to be given to the reporting areal units used in previous studies. Many studies use MSAs, while others use states or HRRs as their units. This can make it difficult to compare studies. The modifiable area unit problem, which references potential statistical bias introduced due to aggregation between units, could very well introduce bias into these studies depending on the method of aggregation and the data aggregated. Although age, sex, race, income, education, and supply factors have all been shown to be significant in studies based on commonly-used areal units (states, MSAs, and HRRs), additional studies are required to confirm the results of previous research. Research that examines different levels of cost, such as those using data reported by DRG, are required to provide insight into the geographic disparities in costs across the complex U.S healthcare system.

Chapter 3 – Methodology

Introduction

This chapter provides details on the methodology used in this research and the steps taken to address the specific objectives identified in Chapter 1. In brief, American Community Survey data were used to estimate demographic variables for each MSA in the U.S. (e.g., age, sex, race, income, and education). Data from the Centers for Medicare and Medicaid Services (CMS) were then used to estimate the costs of treatment of heart failure and the number of discharges. The CMS data were resampled to MSAs so that they could be integrated with the ACS data. Ordinary Least Squares (OLS) regression was then employed to identify explanatory variables correlated with the costs of treatment for heart failure. The chapter is organized into three sections: study area, data characteristics and assemblage, and data analysis.

Study Area

This study will focus on the 356 Metropolitan Statistical Areas (MSAs) in the contiguous United States (United States Census Bureau 2013; Figure 3.1) that have hospitals that received a Medicare payment for the treatment of heart failure in 2011. MSAs in Alaska, Hawaii, and those with no hospitals that received a Medicare payment for heart failure were excluded.

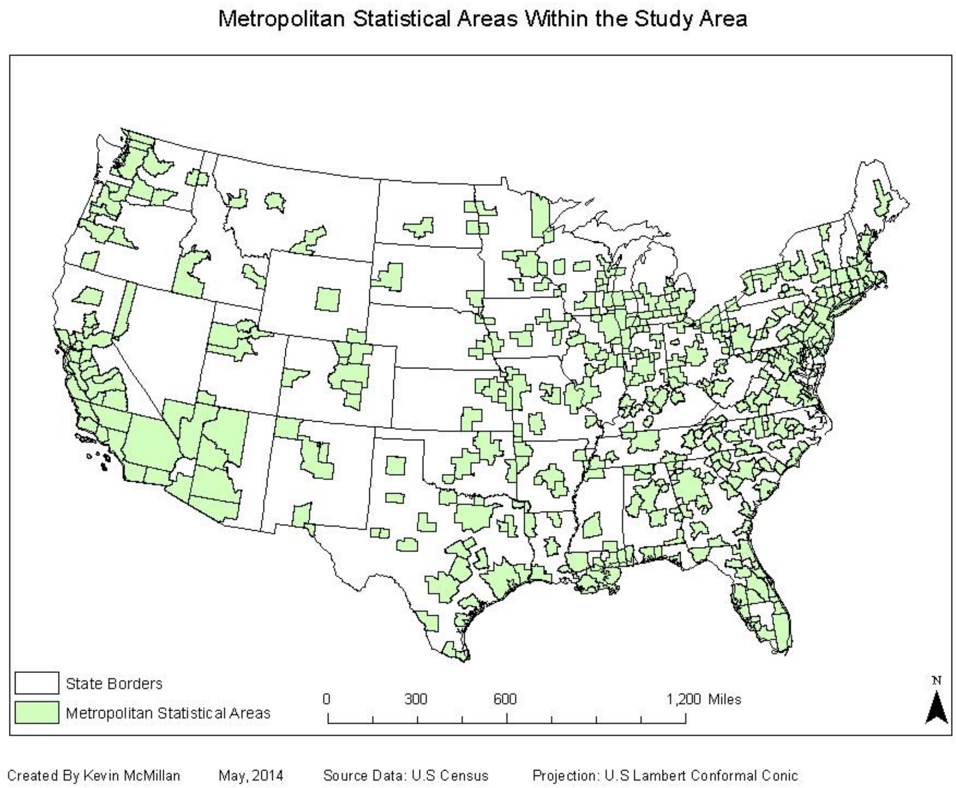


Figure 3.1: MSAs within the study area (Source- Author)

These Office of Management and Budget (OMB) defines MSAs to include a core city of 50,000 or more and the surrounding area determined to have significant interaction with each city, usually including several counties (United States Census Bureau 2013). Land area for MSAs also includes intermittent water and glaciers (note the Great Lakes in Figure 3.1). Their land areas, populations, demographic characteristics, and healthcare spending on heart failure vary greatly (Table 3.1; Table 3.2)

Table 3.1: *Area and Demographic Characteristics of the Study Area (Rounded to Nearest Tenth Decimal) (United States Census Bureau 2013)*

Characteristic	Maximum	Minimum	Mean	Standard Deviation
Land Area (Square Miles)	27263.4	144.6	2445.2	2671.0
Population	18796078	55378	713707.3	1586333.0
Percent African American	52.7	0.6	11.7	10.7
Percent with a Bachelor's Degree or Higher	57.8	12.2	25.5	7.8
Percent Female	52.9	43.3	50.8	1.0
Percent in Poverty	35.3	7.5	15.2	4.0
Percent Aged Sixty-Five and Older	34	7	13.4	3.3

Table 3.2: *Medical Spending and Discharges for Heart Failure Treatment in the Study Area (Rounded to Nearest Tenth Decimal) (United States Census Bureau 2013)*

Characteristic	Maximum	Minimum	Mean	Standard Deviation
Average Hospital Discharges for Heart Failure (Patients)	106	11	38.9	17.3
Average Hospital Charges for the Treatment of Heart Failure (Dollars)	71644	5769	16041.2	7725.2
Average Medicare Payments for the Treatment of Heart Failure (Dollars)	10107.5	3465	4538.5	722.5

The distributions of these characteristics vary somewhat (Figures 3.2-3.10 – note that these figures display the number of MSAs on the Y axis and the characteristic in question on the X axis). The land area and populations of MSAs have a majority of the data on the left side (left skewed) and are not normally distributed (Figure 3.2; Figure 3.3). Many MSAs have relatively small land areas, with some that are very large. Population size is virtually all relatively low, with a few outliers, such as MSAs in California. The percent African American residents in MSAs is also skewed left, suggesting that most MSAs have a relatively low percentage of African American residents, with a few MSAs having an unusually high percentage (Figure 3.4). It should be noted that the average percentage of African Americans in MSAs is slightly lower than the national average (United States Census Bureau 2013). All other characteristics (described in Table 3.1 and 3.2) are reasonably normal with some outliers (Figure 3.5-3.10). These outliers will be counted for the purposes of this thesis, as they are known to be actual data values rather than mistakes. The summary statistics (Table 3.1) are fairly close to the national averages.

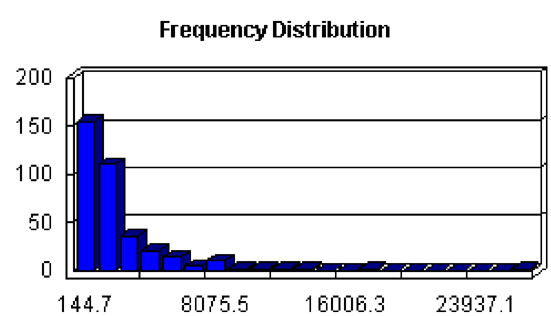


Figure 3.2: Histogram of MSA land areas in square miles (United States Census Bureau 2013)

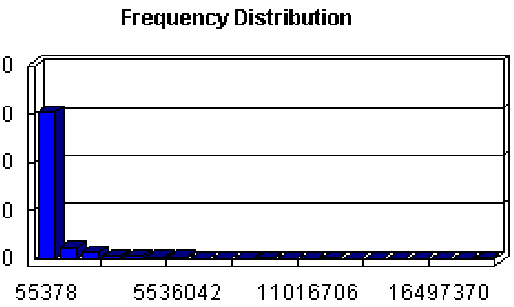


Figure 3.3: Histogram of MSA populations (United States Census Bureau 2013)

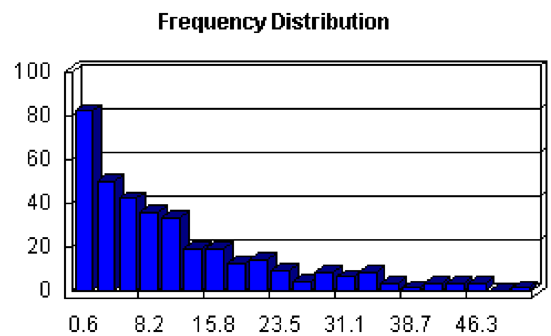


Figure 3.4: Histogram of the percent of African American MSA residents (United States Census Bureau 2013)

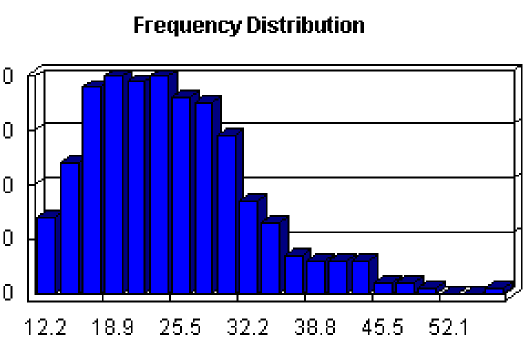


Figure 3.5: Histogram of the percent of MSA residents with a Bachelor's degree or higher (United States Census Bureau 2013)

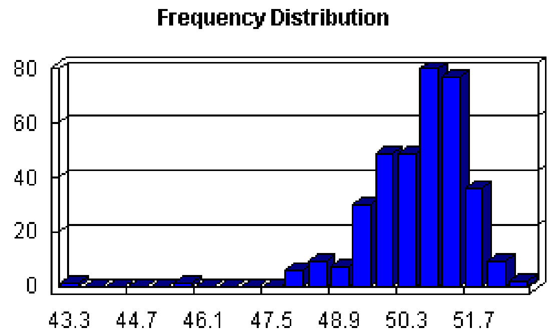


Figure 3.6: Histogram of the percent of female MSA residents (United States Census Bureau 2013)

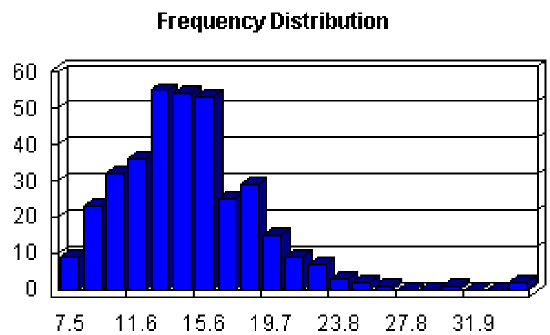


Figure 3.6: Histogram of the percent of MSA residents in poverty (United States Census Bureau 2013)

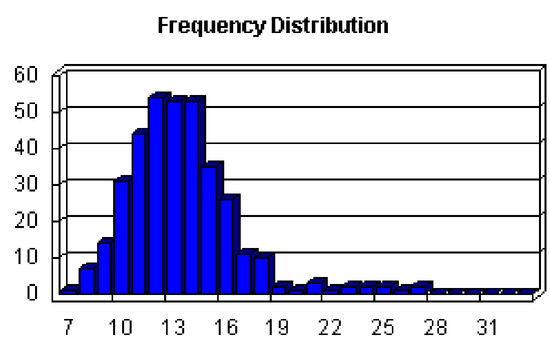


Figure 3.7: Histogram of MSA residents aged sixty-five and older (United States Census Bureau 2013)

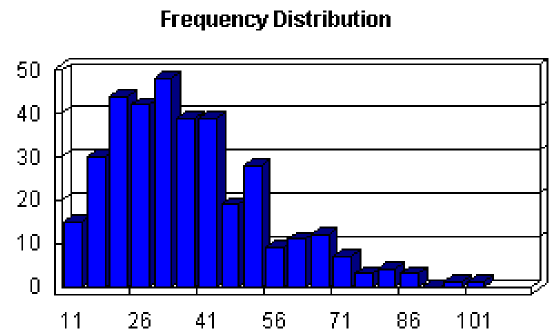


Figure 3.8: Histogram of average hospital discharges for the treatment of heart failure in MSAs (CMS 2013)

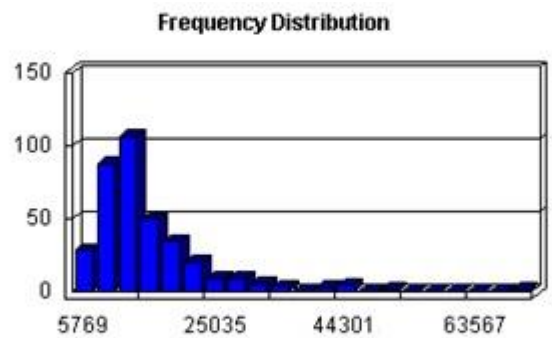


Figure 3.9: Histogram of average hospital charges ($) in MSAs for the treatment of heart failure (CMS 2013)

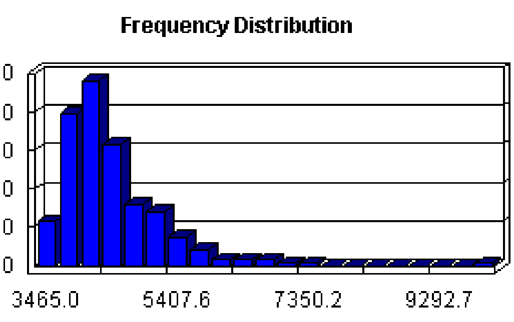


Figure 3.10: Histogram of average Medicare payments ($) to hospitals in MSAs for the treatment of heart failure

MSAs were chosen for three reasons. First, over 80% of people in the United States live in urban areas, which will allow this study to encompass a large majority of the U.S. population (United States Census Bureau 2012). Second, nearly all MSAs in the contiguous U.S. have at least one hospital that treats heart failure, while many counties do not. Finally, extensive ancillary data describing socio-demographic characteristics are available for MSAs.

Data

Data from the Centers for Medicare and Medicaid Services (CMS) dataset were used to determine the locations of hospitals, the number of discharges, and the Medicare payments for heart failure at each hospital. The CMS data cover all hospitals that received a Medicare payment in 2011 (the data are available from http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient.html). Demographic variables for each MSA (e.g., age, sex, race, income, and education) are derived from American Community Survey (ACS) data. The ACS data are organized as “sheets” (numbered 1-5 and organized based on topics). Data used for this study came from sheets 2, 3, and 5 and represent estimated data from 2007-2011 (the data can be downloaded from the U.S. Census from http://www.census.gov/population/metro/). Using the five year ACS estimates was necessary in order to obtain information for the MSAs with a population of less than 65,000. One year estimates do not include areas with populations below 65,000. Both datasets are freely available and can be downloaded from the CMS and the U.S. Census respectively. State boundaries were obtained from ESRI (available from http://www.arcgis.com/home/item.html?id=540003aa59b047d7a1f465f7b1df1950)

(ESRI 2013) All data were re-projected into the U.S Contiguous Lambert Conformal Conic projection.

Software

Most data analysis was conducted using geo-processing and statistical tools incorporated in ESRI ArcGIS version 10.1. Microsoft Excel was used for preliminary refinement of the CMS and ACS data. Refinement included isolating the entries related to heart failure and shock without MCC/CC, changing the formats of headings to ensure ArcGIS compatibility, and creating additional worksheets to facilitate geocoding. Excel was also used to produce a correlation matrix for the independent variables (see chapter 4). All other analysis was completed using ArcGIS.

Data Analysis Procedures

Data Preparation

Analysis began by importing the CMS data into Microsoft Excel. The amount hospitals charged, Medicare payments, and the number of discharges for the treatment of heart failure and shock without CCs and MCCs were isolated by selecting only DRG "293 – Heart Failure & Shock W/O CC/MCC". Special characters and spaces were removed from column headings. The resulting spreadsheet was then converted to .csv format. All analysis after this conversion was completed within ESRI ArcGIS 10.1.

Next, the data were geocoded based on the zip code for each of the 2514 hospitals that received Medicare payments for heart failure in 2011. A geocoder developed by the University of California – Los Angeles (available from http://gis.ats.ucla.edu/) was used to identify the zip code in which each hospital was located. This was accomplished by first adding the University's "Postal_US" address locator to ArcCatalog. Next, the hospitals were geocoded using the Geocode Addresses tool in ArcMap and matching the "Zip" field in the address locator to the "Provider_Zip_Code" in the CMS data. Automated geocoding yielded 188 unmatched hospitals. These were manually matched to their appropriate zip code, which resulted in a 100% match success rate. All hospitals were therefore matched with their proper zip codes and thus with their proper MSAs.

The MSA data were downloaded in geodatabase format. The data were previously joined to the 2007-2011 American Community Survey (ACS) demographic data. Micropolitan Statistical Areas (i.e. smaller towns) were included in the ACS geodatabase, but were not relevant to this research; consequently, these data were removed. The ArcGIS Select by Attribute tool was used with the query LSAD=M1, where LSAD is Legal/Statistical Area Description and M1 is Metropolitan Statistical Areas, to select 374 MSAs and create a new ArcGIS layer. Variables found by previous researchers to be correlated with costs per beneficiary were identified using Census codes (Table 3.3). Using the ArcGIS Delete Field tool, over 2,000 fields were deleted to reduce the size of the dataset.

Next, the ArcGIS Select by Location tool was used to identify the geocoded hospitals that were within MSAs. This resulted in 1855 hospitals remaining in the dataset. Hospitals were then spatially joined to the MSAs using the following input parameters:

- Target Layer: MSAs
- Join Layer: Geocoded Hospitals
- Join Method: one to one
- Field Join Methods:
    - Hospital Charges – Mean
    - Hospital Payments – Mean
    - Hospital Discharges - Mean

A join method of one to one was used to ensure that multiple hospitals within an MSA were aggregated together into that specific MSA. A field join method of “mean” denotes that these data were averaged when aggregated to the MSA level. A total of 356 MSAs and 1845 hospitals remained in the dataset after MSAs in Alaska, Hawaii, and MSAs without hospitals that received payments from Medicare for heart failure treatment in 2011, such as those in Puerto Rico, were removed from the dataset.

Six key independent variables were extracted from the ACS and CMS datasets (Table 3.3). These independent variables were selected because they have been found in previous research to be correlated with geographic variation in health care costs. The ACS data, aggregated by MSA, reflect the percentages of the MSA population in each category. The number of hospital discharges for the treatment of heart failure, extracted from the CMS dataset, denotes the average number of discharges for hospitals within a particular MSA.

Three dependent variables, entirely extracted from the CMS dataset, were selected to be measures of “cost” and are MSA averages (Table 3.4). Note that the hospital charges and the Medicare payments represent hospital averages for the treatment of heart failure (i.e., the average amount patients were charged and the average amount Medicare paid for the treatment).

Table 3.3: *Independent Variables Used in Analysis*

Variable Name	Data Source	Dates	Original Census Code or Name	Re-named
Percentage with a Bachelor’s Degree or Higher	ACS	2007-2011	DP2_HC03_VC94	BA_higher
Percentage of Families and People whose Income in the Past 12 Months was Below Poverty Level	ACS	2007-2011	DP3_HC03_VC166	Poverty
Percentage of Females	ACS	2007-2011	DP5_HC03_VC05	Female
Percentage aged 65 or Older	ACS	2007-2011	DP5_HC03_VC26	Sixfive_Ol
Percent African American	ACS	2007-2011	DP5_HC03_VC73	Af_Amer
Average Number of Hospital Discharges for the Treatment of Heart Failure	CMS	2011	Total Discharges	Discharges

Table 3.4: *Dependent Variables Used in Analysis*

Variable Name	Data Source	Dates	Original Census Code or Name	Re-named
Average Hospital Charges for the Treatment of Heart Failure	CMS	2011	Average Covered Charges	Charged
Average Medicare Payments for the Treatment of Heart Failure	CMS	2011	Average Total Payments	Paid
Average Difference Between Average Hospital Charges and Average	CMS (Calculated)	2011	-	Difference

The difference between the amount charged for treatment of heart failure and the amount Medicare paid was then calculated by creating a new field then using the ArcGIS Field Calculator. Additionally, a unique identification field was created for use in regression analysis using the Field Calculator to create values equal to the field identification number (FID) in the shapefile. The FID field is automatically created by ArcGIS, but it is not available when organizing data in a geodatabase. Columns were given descriptive names instead of the default Census code or name. For example, Census code DP3_HC03_CV166 was renamed Poverty (Table 3.3). It should be noted that ArcGIS limits field names to thirteen characters or less. Spaces and special characters often cause problems during data processing within ArcGIS, so these were avoided in favor of underscores.

Regression Analysis

Before regression analysis began, it was necessary to produce a correlation matrix. A correlation matrix indicates redundancy between variables. A variable is considered “redundant” if it is strongly correlated with another (i.e., a value of 0.50 or greater). Each value in the table represents the correlation coefficient between the two independent variables. A value of 1 would indicate that the variables represent exactly the same data, while a value of zero would indicate none of the same values. A negative symbol (-) before the value indicates a negative relationship between the two variables and the lack of a negative symbol indicates a positive relationship.

Regression analysis was used to determine which, if any, of the independent variables were correlated with each of the dependent variables (Research objective 2 in Chapter 1). This was accomplished using the ArcGIS Ordinary Least Squares (OLS) regression tool. Univariate (single variable) regression models were used by creating separate models for each independent variable combined with each dependent variable. The general equation of univariate OLS regression is $Y_i=\beta X_i+\alpha$, where:

- Y_i is the value of the dependent variable
- β is the slope of the regression line
- X_i is the value of the independent variable
- α is the value at which the regression line crosses the Y axis

Since there are six independent variables and three dependent variables, this resulted in a total of eighteen regression models. The Arc GIS OLS tool automatically develops and computes these models and produces all relevant statistics (See Chapter 4). Complete OLS reports for each model can be found in the appendix. Variables with statistically significant probability values (P values), defined as having a value less than 0.05 (5%), were considered to be correlated with the dependent variable.

Description of Regression Table Statistics

The tables (Table 4.2, 4.3, 4.4) described in Chapter 4 display a variety of important statistics that must be defined. The coefficient (β value in the regression equation) represents the amount of change between the independent variable and the dependent variable (i.e., Medicare payments). It is commonly referred to as the "slope" of the regression line. A positive value indicates a positive relationship between the two variables, while a negative value represents a negative relationship. For example, the coefficient of -54.404197 for the percent female indicates that Medicare payments decrease each time the percent female in a MSA increases by 1 (Table 4.2).

The standard error represents an estimate of the standard deviation of the coefficient. This value acts as a measure of the accuracy of the predictions and, ideally, should be fairly low.

The T-statistic is the number of standard deviations the regression coefficient is from zero. It is simply the coefficient divided by the coefficient's standard error. The T-statistic tests the null hypothesis that the coefficient is equal to zero. A very high or low T-statistic indicates that the coefficient is probably not zero. For example, a T-statistic of

-5.118356 for the percent aged sixty-five and older in a MSA indicates a very high probability that the coefficient is not zero (Table 4.2).

The probability (P-value) is determined based upon the T-statistic and represents the probability that the coefficient is actually zero. A statistically significant P-value represents a statistically significant correlation between the independent and dependent variable. A P-value of 0.05 (5%) or less is considered statistically significant for the purposes of this research.

The R-squared value is a measure of how well data points fit with the regression values. The higher the R-squared value, the better the regression model predicts the data points. This value can be expressed as a percentage. For example, the R-squared value of 0.068905 indicates that the model predicts about 6.8% of the variance of the dependent variable (Table 4.2). Ideally, this value would be very high (>0.50). However, in this study, the R-squared values will inevitably be very low because each variable was treated individually (univariate) rather than as a comprehensive model with multiple variables (multivariate) in the same regression equation. Additionally, there may be potential outliers (OLS reports in the appendix for scatterplots). Outliers were not removed because they were known to be actual data points. The objective of this study was to find correlated variables, not create a comprehensive model, and therefore the other statistics are far more important in this analysis – especially the P-value.

Standard Residual Maps

The standard residual (see maps in the appendix) is the difference between the observed value and the predicted value divided by the estimated standard deviation of the residuals. Essentially, the standard residual is the number of standard deviations the particular value is from the predicted value. In these maps, these standard deviations are mapped by MSA. The further the standard deviation is from zero, the worse the regression line is at predicting the specific value for that MSA.

Presentation of Results

Results of the analyses were presented as maps portraying the geographic variation in health care costs, a correlation matrix, and three summary tables. Note that maps of the variations in the average hospital charges and average Medicare payments use the natural breaks (Jenks) classification method. This method is based on natural groupings in the data. The breaks define classes which are internally similar while maximizing the differences between classes. Each summary table describes the relationship between the independent variables and the dependent variables as determined by OLS regression. The full reports for each OLS regression result and GIS-generated maps of the standard residuals can be found in the appendix. Standard residual maps indicate how far the actual MSA value was from the predicted (OLS regression line) result.

Summary

Data were first organized, narrowed, and prepared in Microsoft Excel. This included a variety of techniques, including isolating data entries for heart failure and removing special characters from column headings. All other preparation and analysis was completed using ArcGIS version 10.1. Preparation included isolating variables of interest, renaming column headings, geocoding hospitals, isolating hospitals within MSAs, and spatially joining the hospitals and MSAs.

Analysis began by creating a correlation matrix to evaluate redundancy between independent variables. Next, the variation in hospital charges and Medicare payments was assessed and two maps were created. Univariate models were then automatically created using the ArcGIS OLS regression tool, and a total of 18 models were created. These models compare each independent variable with each dependent variable. The results of this analysis is noted in Chapter 4.

Chapter 4 – Results

Introduction

This chapter presents the results of the analysis described in Chapter 3. The results are discussed, evaluated, and interpreted in the context of the research objectives outlined in Chapter 1. The discussion is based primarily upon two maps of the geographic variation in health care costs, a correlation matrix, and three summary tables.

Geographic Variations in Health Care Costs

The two maps presented below, created to meet the first goal of this research outlined in Chapter 1, facilitate the visualization of the geographic variations in health care costs by MSA (Figure 4.1; Figure 4.2). Each represents one of the two main measures of costs discussed in Chapter 1: average hospital charges for the treatment of heart failure and the average Medicare payment for the treatment of heart failure.

The geographic variation of Medicare payments for heart failure treatment (Figure 4.1) shows that there are clusters of MSAs where the Medicare payments are very high. These clusters exist in and near California and the North Atlantic region. These clusters may reflect the high cost of living associated in these areas, which would influence the Medicare PFS equation (see Chapter 2). However, there are also MSAs with high Medicare payments in certain parts of the Midwest, Great Plains, and southwest – especially in MSAs in and near American Indian reservations. These areas are often far from medical facilities, and contain populations with low insurance rates and high rates of disease (e.g., cardiovascular diseases that lead to heart failure) (Roubideaux 2004). In many cases, the Medicare payment for MSAs in and near American Indian reservations are actually greater than the amount the hospital charged for treatment.

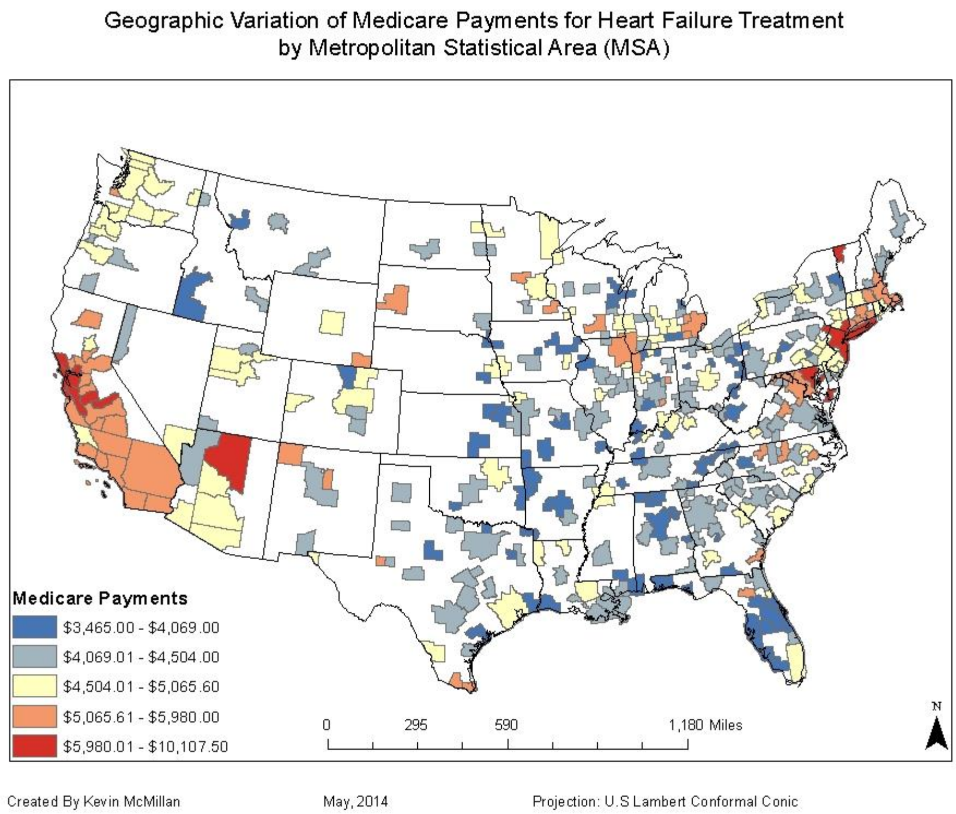


Figure 4.1: Geographic variation of Medicare payments for heart failure treatment

Anything included in the PFS schedule (rent, supplies, malpractice insurance, etc.) would increase the Medicare payment. Areas with higher payments, therefore, have higher costs of maintaining a practice. Areas with lower payments have lower costs of maintaining a practice. Given that the Medicare payment is entirely determined by the PFS equation, it is difficult to imagine extraneous factors directly influencing the Medicare payments. It should be noted that the Medicare payments for heart failure are approximately 28% of the hospital charge for the treatment.

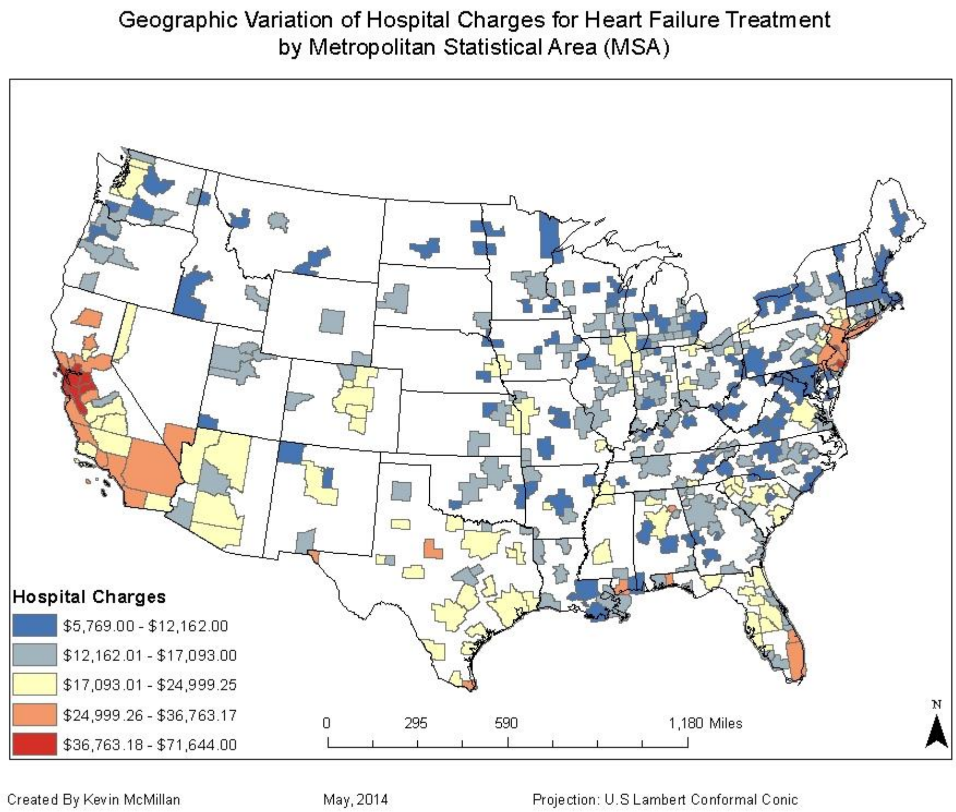


Figure 4.2: Geographic variation of hospital charges for heart failure treatment

A map of the geographic variation of hospital charges for heart failure treatment (Figure 4.2) shows similar patterns when compared to the map of the variation of Medicare payments (Figure 4.1). Many of the same major clusters (near California and the North Atlantic region) exist. However, a great many differences exist, and these can be explained in a variety of ways.

First, hospital size may be influential in determining the hospital charges for heart failure. Many MSAs have only a single hospital that is capable of treating heart failure. In very rural areas, the average travel time to a hospital that could treat a major ailment, such as heart failure, averages nearly an hour (Chan, Hart, and Goodman 2005). This may indicate that only larger hospitals treat heart failure, as there are usually far more clinics, physicians, and smaller hospitals in MSAs and rural areas. The Dartmouth Atlas (1998) notes an increase in costs in hospitals with more hospital beds, which would indicate higher costs in larger hospitals.

Second, it is possible that this map (Figure 4.2) shows areas where physicians are more likely to recommend additional services. A wide variety of studies have shown that the likelihood of additional services is strongly correlated with an increase in cost (Fisher and Skinner 2013; Fisher, Bynum and Skinner 2009; Newhouse and Garber 2013). This is often associated with increased amounts of fraud, but it is also possible that lower cost areas are charging less. However, studies have also suggested additional treatment is not correlated with better treatment (Fisher, Bynum, and Skinner 2009). Regardless, additional treatment would result in an increased charge and a higher Medicare payment. Any treatment for heart failure would be covered in the same DRG, and thus recommendations of additional services would be covered within the data.

Correlation Matrix

Because the values used to create this correlation matrix (Table 4.1) are percentages, some redundancy is to be expected due to the limited number of possible values. For the purposes of this research, these variables are not considered redundant (i.e., have very high correlation with another variable), but extreme care should be taken in future studies that contain variables such as the percentage of African Americans (Af_Amer) and the percentage of females (Female) due to the low-to-moderate correlation of 0.38. Given that none of the variables are redundant, regression analysis will produce worthwhile and will not be biased due to redundancy.

Table 4.1: *Correlation Matrix for Independent Variables*

Correlation Matrix for Independent Variables

Independent Variable	BA_higher	Poverty	Female	Af_Amer	Sixfive_ol	Discharges
BA or higher	1					
Poverty	-0.333419838	1				
Female	0.03340788	0.051830646	1			
Af_Amer	-0.0725536	0.227742659	0.386789282	1		
Sixfive_ol	-0.251121637	-0.19085253	0.284205447	-0.157080855	1	
Discharges	-0.077501151	0.038795251	0.171958183	0.284386269	0.117015395	1

Average Medicare Payments

Medicare spending per beneficiary has been the focus of most of the major research in the variation in health care costs. To meet the second goal of this research outlined in Chapter 1, Medicare payments were again revisited and explored. The results of each of the OLS regression analyses were summarized based on the OLS analysis (see full reports in the appendix).

Table 4.2: *Summary Statistics for the Average Medicare Payment Independent Variables*

Independent Variable	Coefficient	Standard Error	T-Statistic	Probability	R-Squared
Percent African American	-4.361013	3.549652	-1.228575	0.22005	0.004246
% with Bachelor's Degree or Higher	17.575348	4.811201	3.653006	0.000310*	0.036327
Average # of Hospital Discharges	1.366764	2.214329	0.617236	0.537478	0.001075
% Female	-54.404197	38.477205	-1.413933	0.158272	0.005616
% in Poverty	-0.912256	9.617648	-0.094852	0.924472	0.000025
% Aged Sixty-Five and Older	-57.551863	11.24421	-5.118356	0.000001*	0.068905

An Asterisk (*) Denotes Statistical Significance

The results suggest that the percentage of the MSA population with a Bachelor's degree or higher and the percent aged sixty-five and older within a MSA are statistically significant variables (Table 4.2). The P-values of these two variables (0.000310 and 0.000001 respectively) are very low, indicating that their coefficients are most likely not zero. This indicates a statistically significant relationship between these variables and the average Medicare payments for heart failure, and that future studies may wish to include them as part of a comprehensive regression model with many independent variables. The significance is consistent with previous research (Fuchs et. al. 2001; Gage et. al. 1999; Centers for the Evaluative Clinical Sciences 1999; MedPAC 2013)

The fact that the percentage of those with a Bachelor's degree or higher is correlated with the average Medicare payments for the treatment of heart failure seems to fit the previous research reviewed in Chapter 2 (Fuchs et. al. 2001). According to the results of this thesis, a one percent rise in the percent in a MSA with a Bachelor's degree results in approximately a $17 rise in average Medicare payments for heart failure. This may be due to a variety of reasons, such as the increased resources associated with collegiate degrees or the fact that higher education leads to better health behaviors, such as receiving regular check-ups from a physician (Cutler 2007).

The results of this analysis showed that the percentage in the MSA population aged sixty-five and older was negatively correlated with Medicare payments for the treatment of heart failure. These results are at odds with research conducted by previous investigators. Virtually all such research that included age as a variable determined that an increase in age should increase Medicare payments (Gage et. al. 1999; Centers for the Evaluative Clinical Sciences 1999; MedPAC 2013). According to the results of this analysis, a one percent rise in the percent aged sixty-five and older in a MSA results in approximately a 57 dollar decrease in average Medicare payments for the treatment of heart failure. This phenomenon may be unique to heart failure, however it should be noted that approximately one-third of heart failure patients are under the age of sixty-five. Additionally, the risk of heart failure decreases dramatically for females over the age of sixty-give, but increases for males over the age of sixty-five (Centers for Disease Control). It is likely that many of those over the age of sixty-five are women due to longer life expectancy. Therefore, it may be necessary to obtain data for those over sixty-five, rather than for the entire population of a MSA. With such data, a secondary analysis may confirm this negative relationship.

No other variables (percent in poverty, percent female, percent African American, and the average number of hospital discharges) were statistically significantly correlated with Medicare payment for heart failure at the 0.05 (5%) significance level. Statistically, these variables had very low coefficients – suggesting there is very little related change between the independent and dependent variables. Additionally, these variable influence the Medicare PFS equation very indirectly or not at all. However, it should also be noted that these statistics were for the MSA as a whole, not only for those over the age of sixty-five (the primary user of Medicare). Data for only those over the age of sixty-five would be very beneficial for this analysis, and a secondary analysis could re-evaluate the significance of these variables.

Average Hospital Charges

The average amount hospitals charged for treatment of heart failure has not been thoroughly explored. As indicated in Chapter 1, an objective of this thesis was to determine if variables identified as being correlated with Medicare spending per beneficiary could also help explain hospital charges. Three variables were found to be significant: the average number of hospital discharges for heart failure in a MSA, the percent female, and the percent aged sixty-five and older (Table 4.3)

Table 4.3: *Summary Statistics for the Average Hospital Charge Independent Variables*

Independent Variable	Coefficient	Standard Error	T-Statistic	Probability	R-Squared
Percent African American	-28.459909	38.006564	-0.748816	0.454457	0.001581
% with Bachelor's Degree or Higher	39.012215	52.364993	0.745006	0.456755	0.001565
Average # of Hospital Discharges	-48.919487	23.54707	-2.077519	0.038464*	0.012045
% Female	-1004.875204	409.118904	-2.456194	0.014510*	0.016756
% in Poverty	65.773354	102.78174	0.639932	0.52263	0.001155
% Aged Sixty-Five and Older	-297.77611	123.592659	-2.409335	0.016479*	0.016133

An Asterisk (*) Denotes Statistical Significance

The P-values of each of these variables (0.038464, 0.014510, and 0.016479 respectively) are very low, and indicate that these variables would likely be useful in a comprehensive model with many variables.

It is noteworthy, however, that these variables are inconsistent with previous literature regarding Medicare payments per beneficiary. They suggest that these variables (the average number of hospital discharges for heart failure in a MSA, the percent female, and the percent aged sixty-five and older) result in a decrease in average hospital charges within a MSA.

The results of this analysis show that a one percent rise in the percent female in a MSA results in over a $1000 decrease in average hospital charges for the treatment of heart failure in a MSA. The most likely reason that the results found in this thesis are different may be that men are far more at risk of heart failure after age sixty-five than females (Centers for Disease Control 2013). However, it should be noted that it is possible that this is a statistical phenomenon: the percent female within a MSA has very little variation (close to 50%) compared to other variables. A lack of variation may introduce statistical bias, and cause the coefficient to be difficult to interpret.

It was also noted in the results of this analysis that a one percent rise in the percent aged sixty-five and older results in nearly a $300 decrease in hospital charges for the treatment of heart failure. Interestingly, a statistically significant negative relationship between the percent aged sixty-five and older exists in both the average Medicare payments for the treatment of heart failure and the average hospital charges for the treatment of heart failure. Research into this relationship deserves attention, as it does not seem to fit with the studies based on Medicare payments per beneficiary. However, given the significance and large reduction in hospital charges for the treatment of heart failure in MSAs with higher percentages of females, it may be that many of those over the age of sixty-five are women. Therefore, it may be necessary to obtain data for those over sixty-five, rather than for the entire population of a MSA. With such data, a secondary analysis may confirm or refute this negative relationship.

A less dramatic, but still unusual, relationship exists between the average number of hospital discharges for heart failure in a MSA and the average hospital charge for that MSA. The analysis in this thesis shows that a one percent rise in the average number of hospital discharges for the treatment of heart failure results in just under a $50 decrease in average hospital charges. This seems counter-intuitive, and is at odds with previous research on Medicare payments per beneficiary that suggests an increase in the number of hospital discharges should result in increase in costs. Conventional wisdom suggests that increased demand on a service should result in an additional cost for that service. Given this, it is difficult to speculate about reasons why such a negative relationship exists.

All of the remaining variables (percent with a bachelor's degree or higher, and percent in poverty, percent African American) were not statistically significantly correlated with the average hospital charge for the treatment of heart failure at the 0.05 (5%) significance level. However, it should also be noted that these statistics were for the MSA as a whole, not just for those over the age of sixty-five (the primary user of Medicare). Data for only those over the age of sixty-five would be very beneficial for this analysis, and a secondary analysis could re-evaluate the significance of these variables. However, it is very important to note that this inconsistency with the literature may be due to the simple fact that the research tends to focus on Medicare payments per beneficiary, rather than hospital charges. Variables may behave differently when examining hospital charges.

The Difference Between Average Hospital Charges and Average Medicare Payments

As noted in Chapter 3, the difference between average hospital charges and average Medicare payments (charge – payment) was calculated in an effort to discover why such a disparity between these values exists. Table 4.4 shows that three variables show a significant relationship with the difference between average hospital charges for heart failure treatment and the average Medicare payments for heart failure treatment: the average number of hospital discharges for heart failure in a MSA, the percent female, and the percent aged sixty-five and older.

Table 4.4: *Summary Statistics for Difference Independent Variables*

Independent Variable	Coefficient	Standard Error	T-Statistic	Probability	R-Squared
Percent African American	-24.098913	36.712738	-0.656418	0.51198	0.001216
% with Bachelor's Degree or Higher	21.43698	50.599917	0.423656	0.672087	0.000507
Average # of Hospital Discharges	-50.28628	22.722884	-2.213024	0.027520*	0.013646
% Female	-950.472074	395.256669	-2.404696	0.016686*	0.016072
% in Poverty	66.685749	99.258785	0.671837	0.502123	0.001273
% Aged Sixty-Five and Older	-240.225031	119.658869	-2.007582	0.045440*	0.011257

An Asterisk (*) Denotes Statistical Significance

It was found that a one percent increase in any of the three variables discussed above decreased the difference between the average hospital charge for heart failure and the average Medicare payment for heart failure. This means that as the magnitude of these variables increases, the average hospital charge for the treatment of heart failure and the average Medicare payment for the treatment of heart failure becomes more similar. It is unclear if the difference is a suitable measure of cost, but it is certain that many of the independent variables identified in research based on Medicare payments per beneficiary are statistically significant. Nevertheless, research into the difference between hospital charges and Medicare payments as a suitable measure of cost may provide insight about the disparity in health care and help shape future policy for both hospitals and Medicare. A higher difference, for example, may indicate that a particular MSA charges disproportionately high for the treatment of heart failure compared to the Medicare payment. Given that the PFS equation suggests a “fair” payment for a service, research into disproportionately high-charging MSAs may reveal additional demographic or systemic variables that influence the cost of health care.

Standard Residual Maps

Maps portraying standard residuals were prepared to facilitate visualization of potential outliers for each of the three dependent variables discussed above and their relationships with each of the six variables (see Chapter 3). A total of 18 maps of standard residuals were generated (available in the appendix). For example, the regression results show a standard residual of over 2.5 for the average amount hospitals charged in a MSA and the percent aged sixty-five and older in a MSA, an indication that the average amount charged was much higher than the predicted value.

The maps show that the regression models are generally good in the Midwest and Great Plains, but are generally not as good near the coasts. Specifically, all maps show that the models predicted poorly for MSAs in and around California and most show poor prediction in the Northern Atlantic region. This may be due to the cost of living in these areas, as the patterns are similar (Missouri Department of Economic Development 2013)

Summary

The results of this research clearly demonstrate that the costs of treating heart failure vary geographically. It was found that any of the same variables used in previous studies based on Medicare payments per beneficiary can be used for studies based on Medicare payments or hospital charges reported by DRG. However, many of the variables have opposite relationships when compared to studies using cost per beneficiary data. This may be due to a variety of secondary factors such as the nature of heart failure (i.e., men are far more likely to be diagnosed with heart failure compared to women) or may be statistical. It may be necessary to obtain data for those aged sixty-five and older, Medicare's primary demographic, within a MSA for re-evaluation rather than using statistics for the entire population of a MSA.

A GIS-based approach to evaluating the variation in the cost of health care is a reasonably unique approach compared to the studies reviewed in Chapter 2. A GIS methodology was successfully employed to evaluate various socio-demographic variables. While many of the results presented were unexpected, the methodology used in this research seems sound and will remain similar for any additional analysis.

Chapter 5: Conclusion

Thesis Summary

Health care costs in the U.S. have been rising steadily for decades, yet vary geographically. Explanations of costs have been studied extensively, however, there are many uncertainties regarding the drivers of these inequities. Previous research on the geographic variation of health care costs has been primarily focused on Medicare spending per beneficiary. This research has two major shortcomings: (1) it has generally not focused on individual diseases or treatments, such as Diagnosis Related Groups (DRGs), and thus cannot give insight into the explanation of the variation in Medicare payments at the DRG level and (2) it has not considered hospital charges, a particular concern in understanding costs associated with uninsured patients.

This thesis attempted to address both of these issues by using a new dataset. The study examined treatment of heart failure in 365 Metropolitan Statistical Areas (MSAs) distributed across the U.S. Heart failure is a major and costly disease that affects thousands of people across the nation.

The principal objectives of this theses were: (1) to determine how health care costs (Medicare payments and the amount the hospital charged) for the treatment of heart failure vary geographically across the major metropolitan areas of the U.S, (2) to assess whether geographic variation in Medicare payments and hospital charges for heart failure treatment can be explained by variables found in relevant previous research based on

Medicare cost per beneficiary, and (3) to develop a suitable Geographic Information System (GIS) methodology for assessing these variations.

Ordinary Least Squares regression was applied to the average Medicare payments for the treatment of heart failure in the MSAs. Results of the analysis were compared to previous literature that used Medicare payments per beneficiary data to see if similar variables (percent with a Bachelor's degree or higher, percent in poverty, percent female, percent aged sixty-five and older, percent African American, and the average number of hospital discharges for the treatment of heart failure) were applicable when using DRG data for heart failure treatment. Results suggest that an increase in the percent with a Bachelor's degree or higher increases Medicare payments, but an increase in the percent aged sixty-five or older decreases Medicare payments. While the increase in payments correlated with the percent with a Bachelor's degree or higher is consistent with the literature, the decrease in payments correlated with the percent aged sixty-five and older is not. This may be attributable to the fact that a majority of older individuals are women, as women over sixty-five are less at risk of heart failure.

In a second phase of analysis, average hospital charges for the treatment of heart failure were examined with the same variables. Those results also indicated that many of the same variables are correlated with average hospital charges, but again with opposite effects. The percent female, the average number of discharges for heart failure, and the percent aged sixty-five and older were correlated with a decrease in hospital charges for the treatment of heart failure. This inconsistency with the literature may be due to the simple fact that the research tends to focus on Medicare payments per beneficiary, rather than hospital charges. Nevertheless, it is also possible that these are due to the fact that women are at a lower risk of heart failure than men after the age of sixty-five.

Further research will be necessary to determine why such opposite effects exist, but the results of this thesis are integral to providing a base-line understanding of explanations of variations in the cost of health care at the DRG level. This thesis provides evidence that similar variables can be used when examining DRG payments, and begins to explain the variations in hospital charges for heart failure.

Conclusions

Objective 1

Two maps that portray the geographic variation in average Medicare payments and the average amount hospitals charged for the treatment of heart failure were prepared to address the first objective (Figures 4.1 and 4.2) . These maps demonstrate that the costs for the treatment of heart failure vary geographically. They show that there are clusters as well as isolated MSAs with high costs. These variations are consistent with the literature, which suggests that health care costs do vary geographically. However, these maps suggest that the geographic variation of health care costs remains true when using DRG data.

Objective 2

Results of OLS regression showed that most variables found relevant in previous research (based on Medicare payments per beneficiary data) were different (Table 5.1; Table 5.2).

Table 5.1: *Independent Variables for Medicare Payments*

Variable Name	Consistent with Literature?	Reason for Inconsistency
Percent with a Bachelor's Degree or Higher	Yes	-
Percentage in Poverty	No	Not significant
Percent Female	No	Not significant
Percent aged 65 or Older	No	Decreases Medicare Payments
Percent African American	No	Not significant
Average Number of Hospital Discharges for the Treatment of Heart Failure	No	Not significant

Table 5.1: *Independent Variables for Hospital Charges*

Variable Name	Consistent with Literature ?	Reason for Inconsistency
Percent with a Bachelor's Degree or Higher	No	Not significant
Percentage in Poverty	No	Not significant
Percent Female	No	Decreases hospital charges
Percent aged 65 or Older	No	Decreases hospital charges
Percent African American	No	Not significant
Average Number of Hospital Discharges for the Treatment of Heart Failure	No	Decreases hospital charges

Reasons for inconsistencies may vary. In many cases (the percent aged sixty-five and older for both Medicare payments and hospital charges and the percent female for hospital charges), there may be a data problem. Data for an entire MSA, rather than for only those aged sixty-five and older, was used for this research. It is possible that, when using data for those aged sixty-five and older, that these results may change. However, it is also possible that the variables behave very differently when using DRG data compared to studies that use Medicare payments per beneficiary. Indeed, this may especially be the case when evaluating hospital charges for the treatment of heart failure. Additional research is necessary regardless; however, the methodology seems sound.

Objective 3

In spite of the uncertainties in the results of this analysis noted above, this thesis did demonstrate the virtues and shortcomings of utilizing the CMS dataset in a Geographic Information System (GIS) based methodology for assessing geographic variation in the treatment of heart failure. Many researchers have used statistical software to assess variation in health care costs; however, such software does not allow for the exploration of a spatial context. Using a GIS provides the distinct advantage of utilizing and visualizing the spatial context, a valuable tool for communicating results. However, a GIS is also far more limited in statistical tools compared to statistical software. A statistical program will, by default, often have a much wider variety of analysis that can be used. This study utilized a tool, OLS regression, already part of a standard extension pack in ArcGIS. However, this program and extension can be cost prohibitive. Regardless of these potential issues, this study has shown that utilizing a GIS-based methodology for analyzing the costs of health care using DRG data is viable. Notably, other GIS programs, such as GeoDa, are far more statistically based and are often open source and free to download.

Limitations

It is noteworthy that this thesis was limited in several regards. These limitations may have affected the results, but in ways still unclear. The most important limitation was in the selection of variables, which were limited to a handful of socio-demographic variables found to be related to health care costs in previous research based on analyses of Medicare payments per beneficiary. The importance of other variables that may be important drivers of geographic variation in costs of health care could, thus, not be determined.

Second, although the new CMS dataset appears to have substantial utility, some limitations were observed in ancillary data used. For example, Medicare only covers those aged sixty-five and older and some disabled persons under the age of sixty-five. Therefore, any research based upon the Medicare population may not be applicable to the population as a whole. Additionally, only a single year (2011) of the CMS data was available. This makes it impossible to do any sort of year-to-year analysis. It may be that 2011 was not a representative year for the health care industry or Medicare, but this cannot be determined until additional data are released.

Third, this research was based on a simplistic univariate approach to determining correlations between selected socio-demographic variables and measures of costs. It is clear that a multivariate regression model is needed for an improved analysis of the geographic variation in health care costs using DRG data.

Implications of the Research

This thesis provides a methodology on which future research regarding the geographic variation in health care cost can be based. The framework is flexible, allowing individual researchers to select any individual disease (DRG) in the CMS dataset and choose other variables to employ.

This thesis should be viewed as an exploratory study of the potential for using the CMS dataset, OLS regression, and GIS to improve the understanding of the geographic variation in health care costs (specifically, costs for the treatment of heart failure). Since the results were inconclusive, it is recommended that others interested in costs of health care should use these results with caution. The relationships between health care costs using DRG data and demographics remains uncertain, and additional research is needed to confirm the results of this thesis.

With the implementation of the Patient Protection and Affordable Care Act (PPACA), the methodology will remain very similar, however; the results may change drastically. While the results of the PPACA remain uncertain, it is clear that it will radically change the health care system as more individuals become insured and Medicare expands. These may radically change the variables needed to explain health care costs in unforeseen ways. Hospital charges will remain a topic of concern as Medicare attempts to reduce expenditures and it is likely that there will always be a certain number of individuals who remain uninsured.

Suggestions for Future Research

Studies that examine geographic variation in health care costs will continue to be necessary. A better understanding of the factors that drive such disparities in cost is critical in order to address the issue. Of course, the impacts of the PPACA are currently unknown; and thus an emphasis of future research must be to determine how the PPACA influences the geographic variation in health care costs in upcoming years.

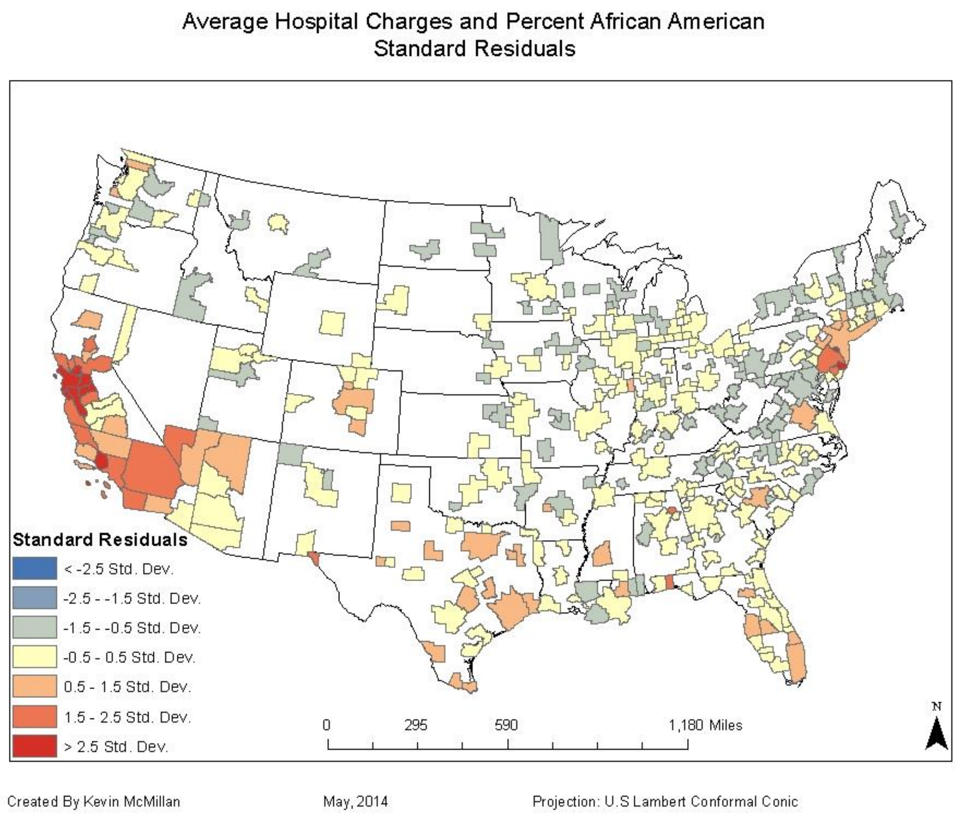
Meanwhile, it is recommended that researchers test (and, perhaps, modify) the methodology used in this thesis. The costs of other diseases (e.g., cancer) should be assessed, and the influence of additional variables (socio-economic and supply-side) on the variations in health care costs should be studied. There are currently 100 DRGs in the CMS dataset.

Future research should also be directed towards the development of a multivariate approach to explaining geographic variation in costs. The statistically significant variables (especially the percent with a Bachelor's degree or higher for the Medicare payments) identified in this research could be used in conjunction with others in an attempt to better understand the variations in the cost of heart failure treatment and other maladies. Both analyses of average Medicare payments for heart failure and the average hospital charges for the treatment of heart failure should be explored.

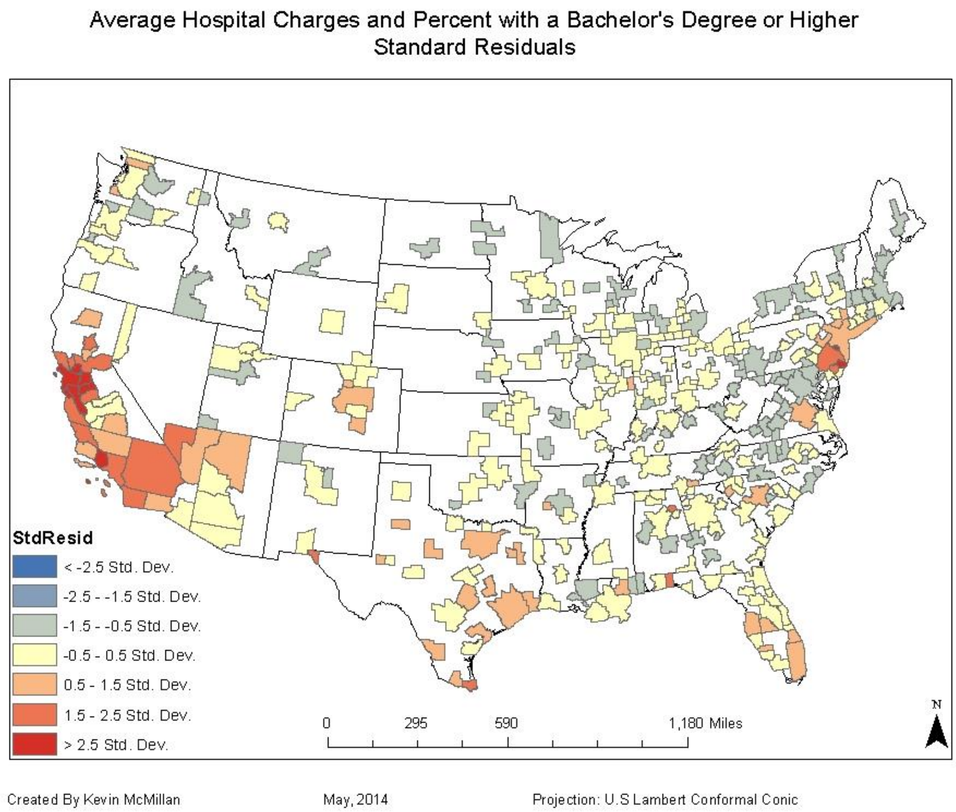
Additional research also needs to address the fact that this thesis found significant variables (e.g., percent aged sixty-five and older for both Medicare payments and hospital charges, and the percent female and the average number of hospital discharges for heart failure for hospital charges) to have the opposite effect that was expected from research based on Medicare payments per beneficiary. It must be determined if these relationships truly exist, or are artifacts of data used in the analysis. Reasons why variables identified from previous research on variations in health care cost using Medicare payments per beneficiary data were not found to be statistically significant must be found. Finally, additional investigation is needed to determine the possible merits of using the difference between the average hospital charge and the average Medicare payment in a MSA as another potential form of cost or disparity.

Appendix

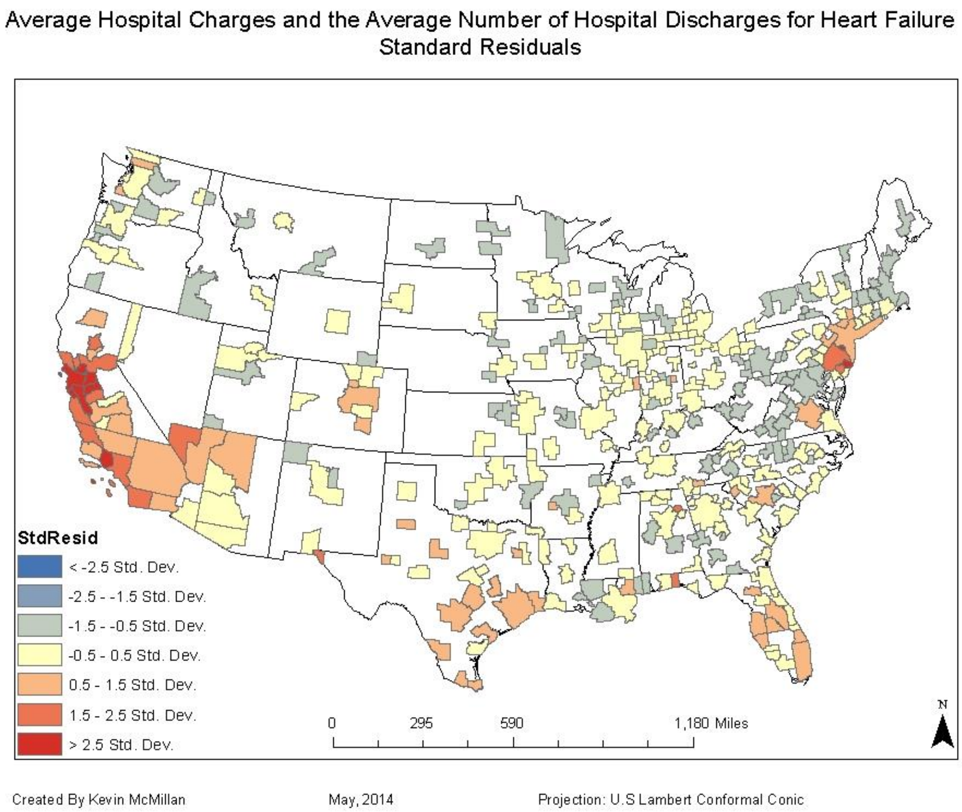
Standard Residual Maps



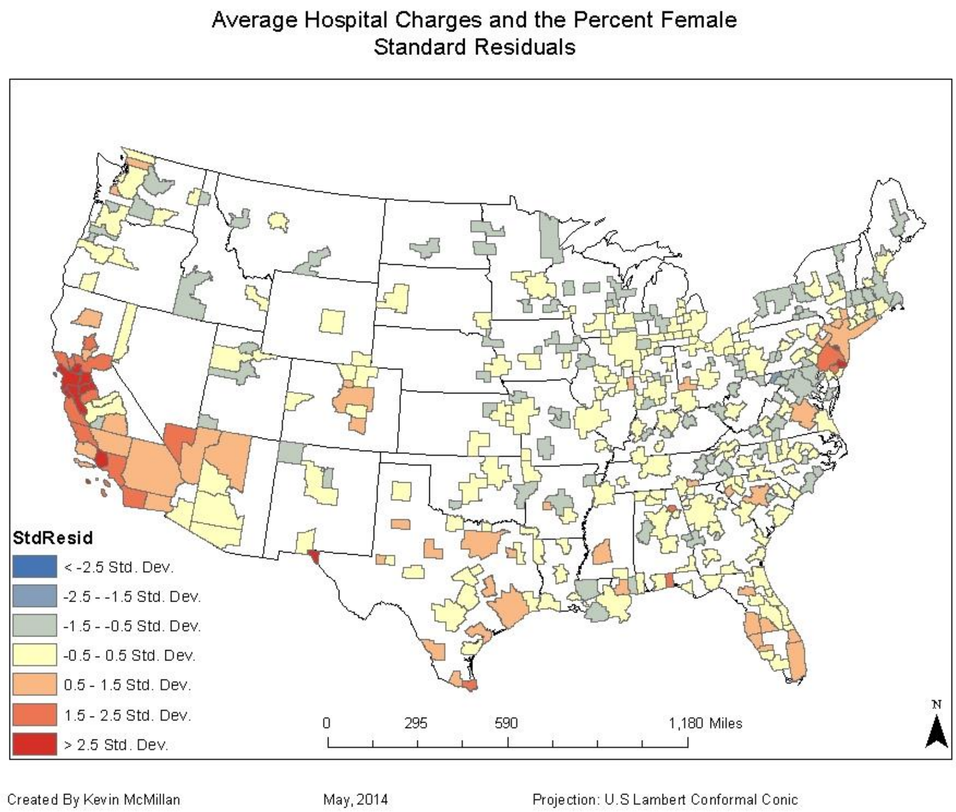
Map 1: A Map of Standard Residuals from OLS Regression using Average Hospital Charges and the Percent African American (Source- Author)



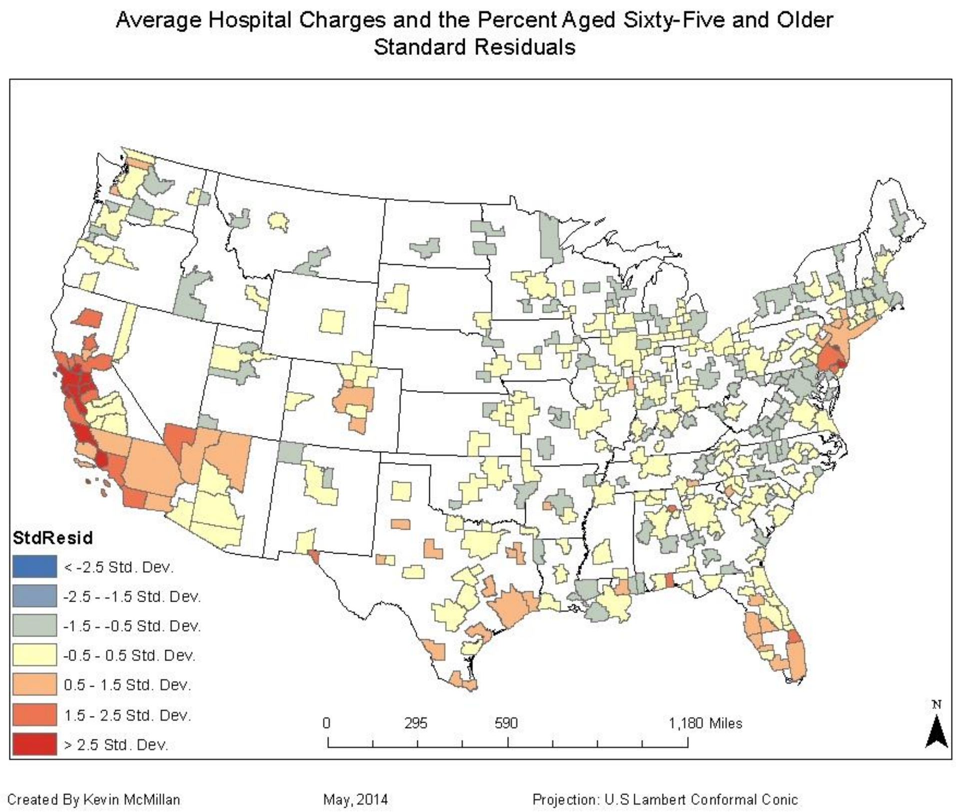
Map 2: A Map of Standard Residuals from OLS Regression using Average Hospital Charges and the Percent with a Bachelor’s Degree or Higher (Source- Author)



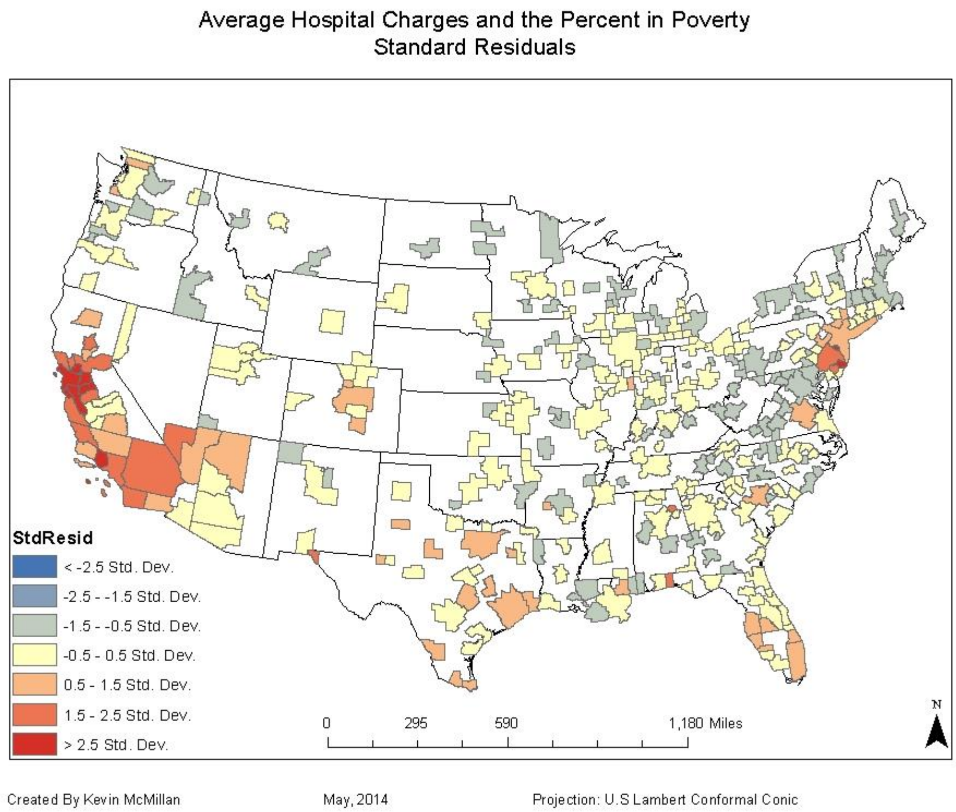
Map 3: A Map of Standard Residuals from OLS Regression using Average Hospital Charges and the Average Number of Hospital Discharges for Heart Failure (Source-Author)



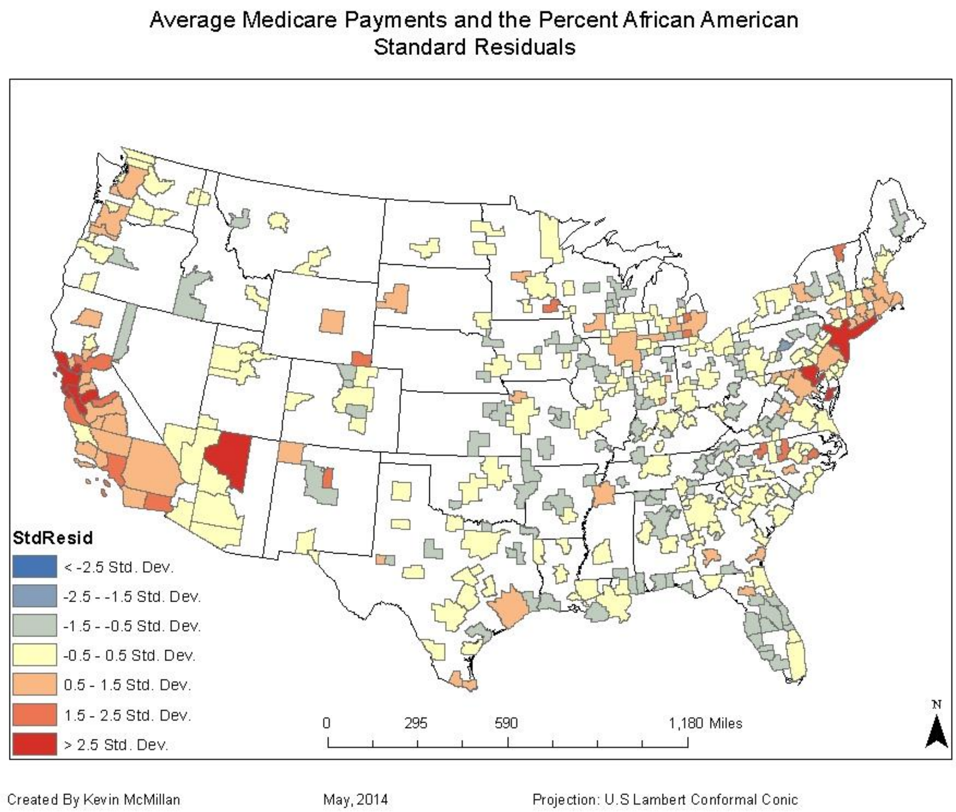
Map 4: A Map of Standard Residuals from OLS Regression using Average Hospital Charges and the Percent Female (Source- Author)



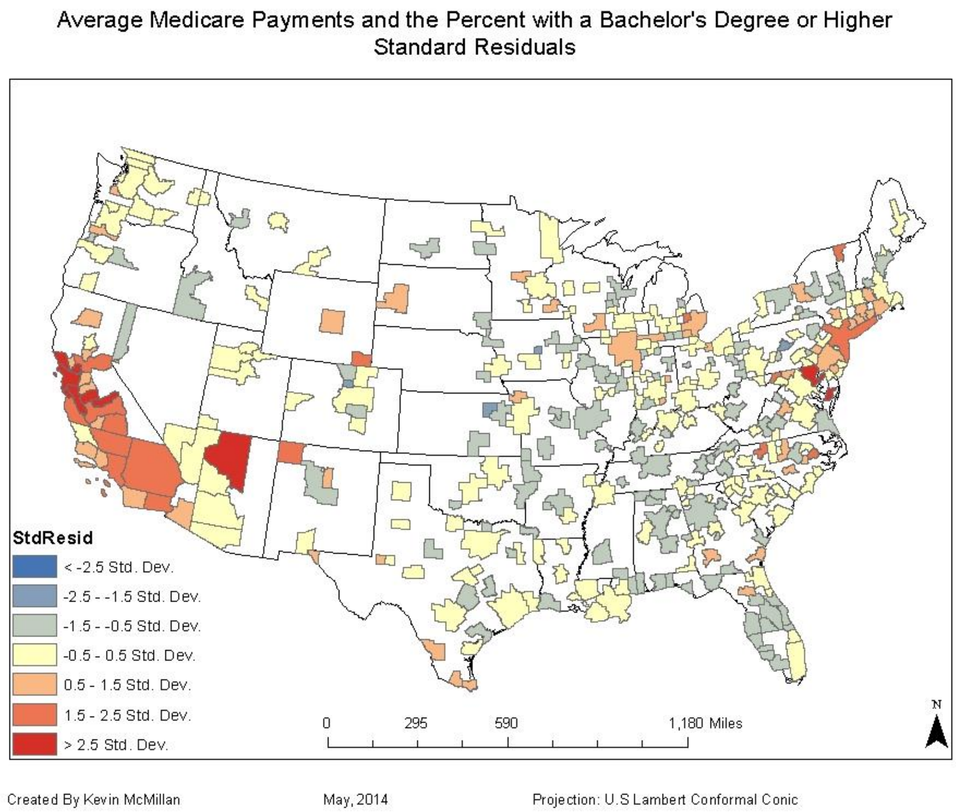
Map 5: A Map of Standard Residuals from OLS Regression using Average Hospital Charges and the Percent Aged Sixty-Five and Older (Source- Author)



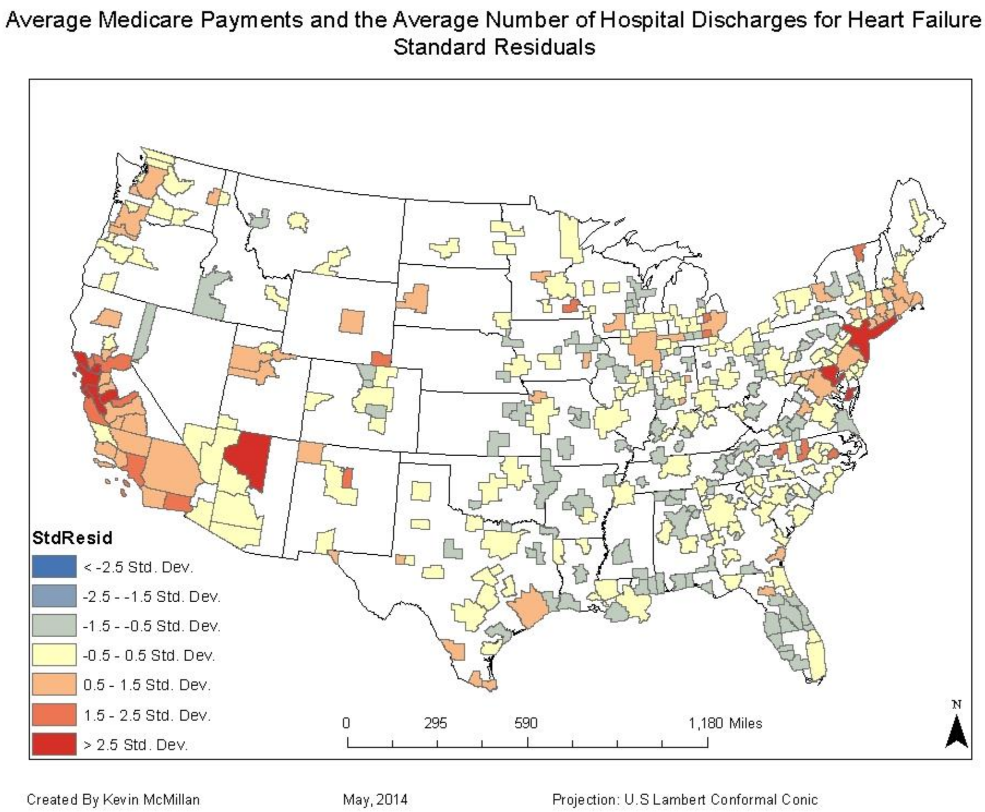
Map 6: A Map of Standard Residuals from OLS Regression using Average Hospital Charges and the Percent in Poverty (Source- Author)



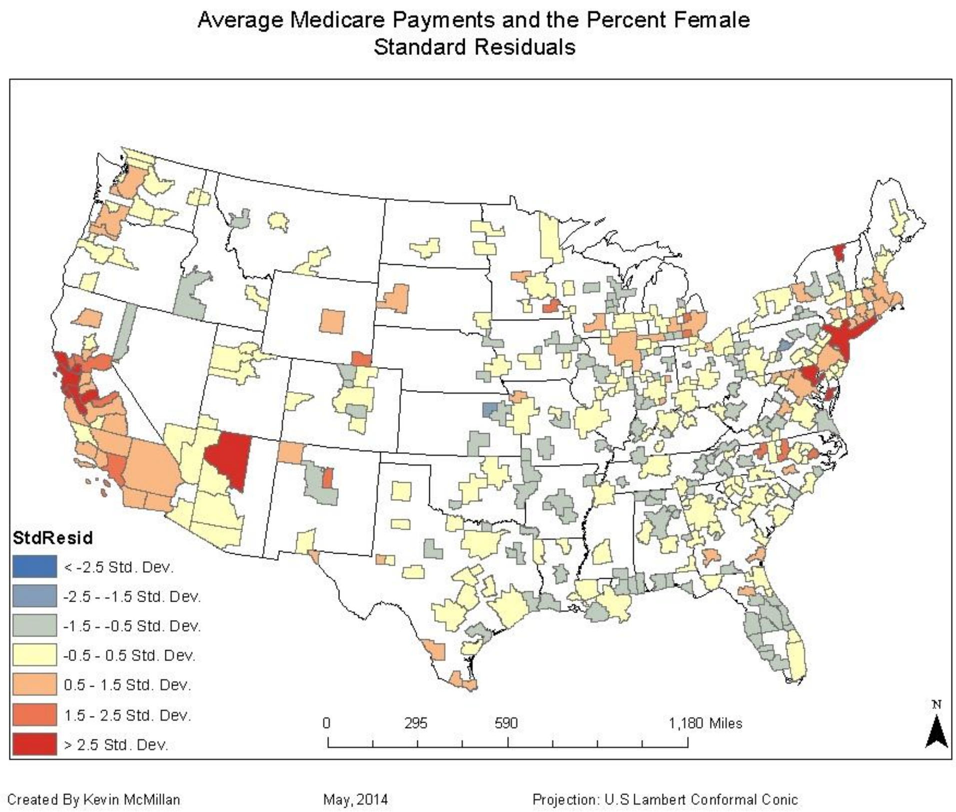
Map 7: A Map of Standard Residuals from OLS Regression using Average Medicare Payments and the Percent in African American (Source- Author)



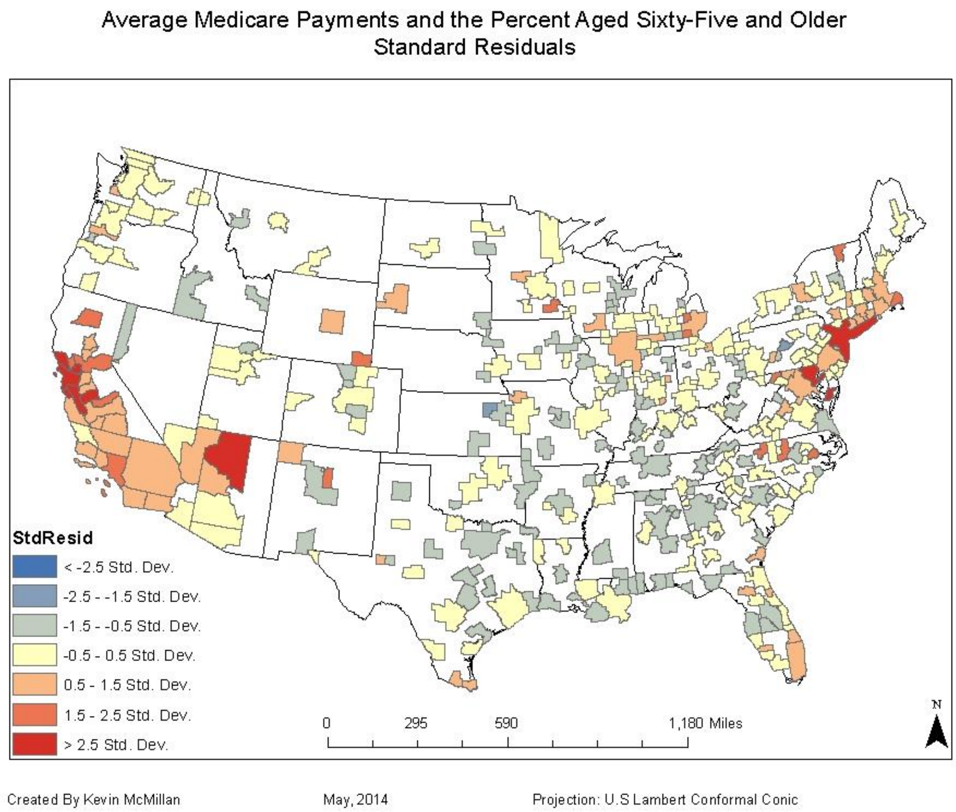
Map 8: A Map of Standard Residuals from OLS Regression using Average Medicare Payments and the Percent with a Bachelor's Degree or Higher (Source- Author)



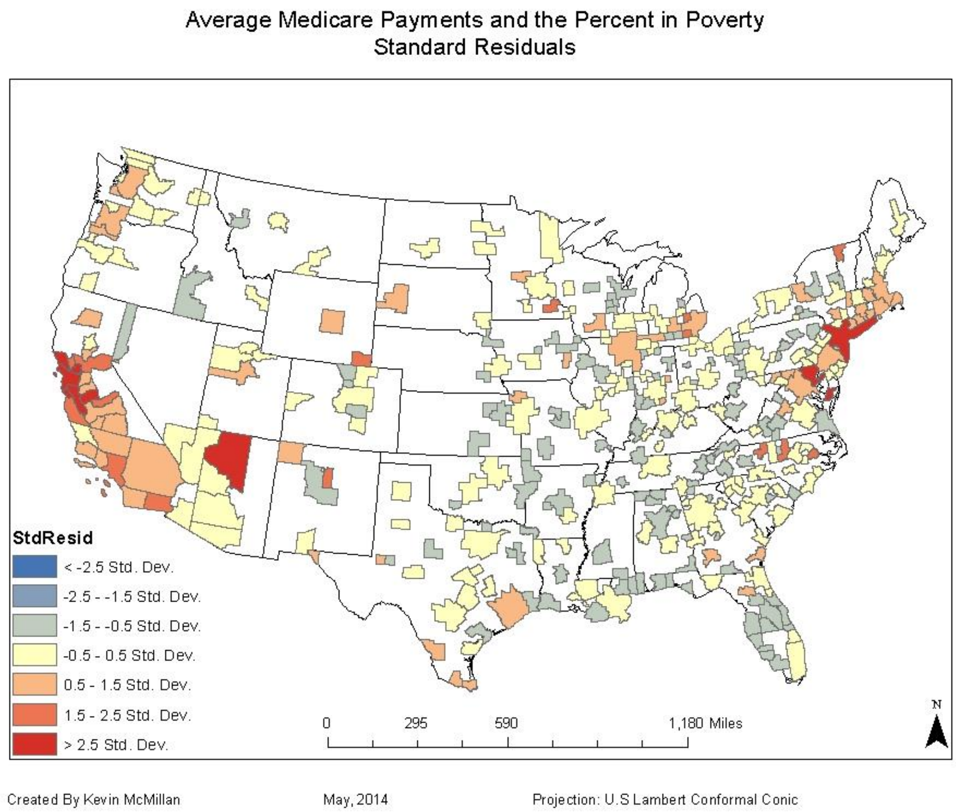
Map 9: A Map of Standard Residuals from OLS Regression using Average Medicare Payments and the Average Number of Hospital Discharges for Heart Failure (Source- Author)



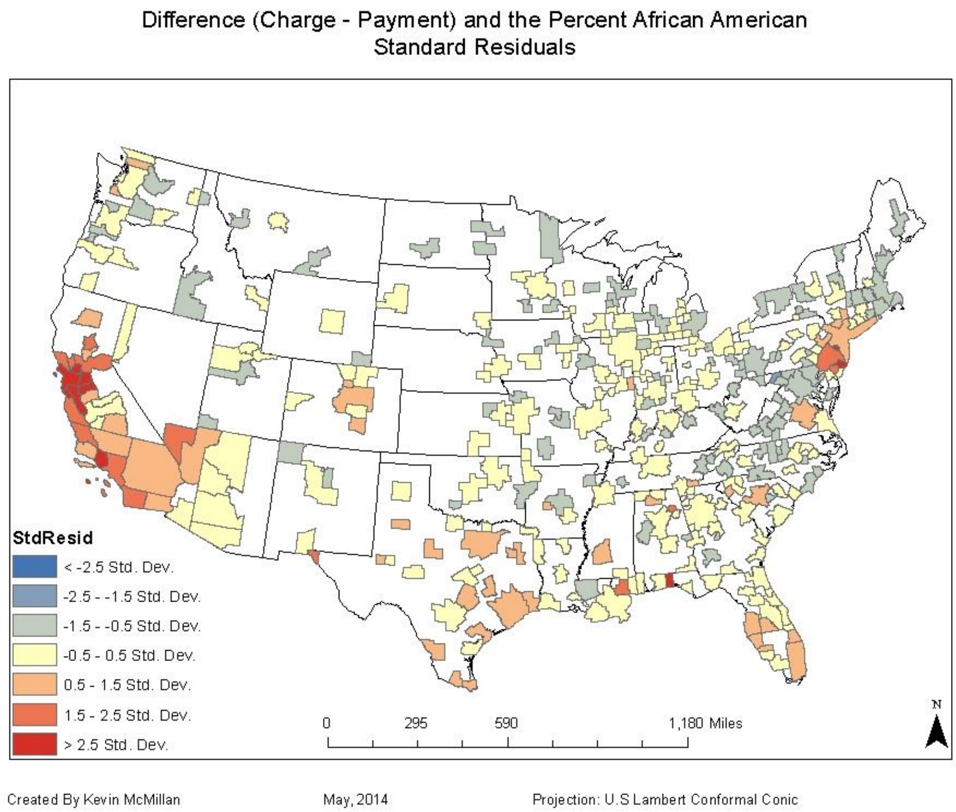
Map 10: A Map of Standard Residuals from OLS Regression using Average Medicare Payments and the Percent Female (Source- Author)



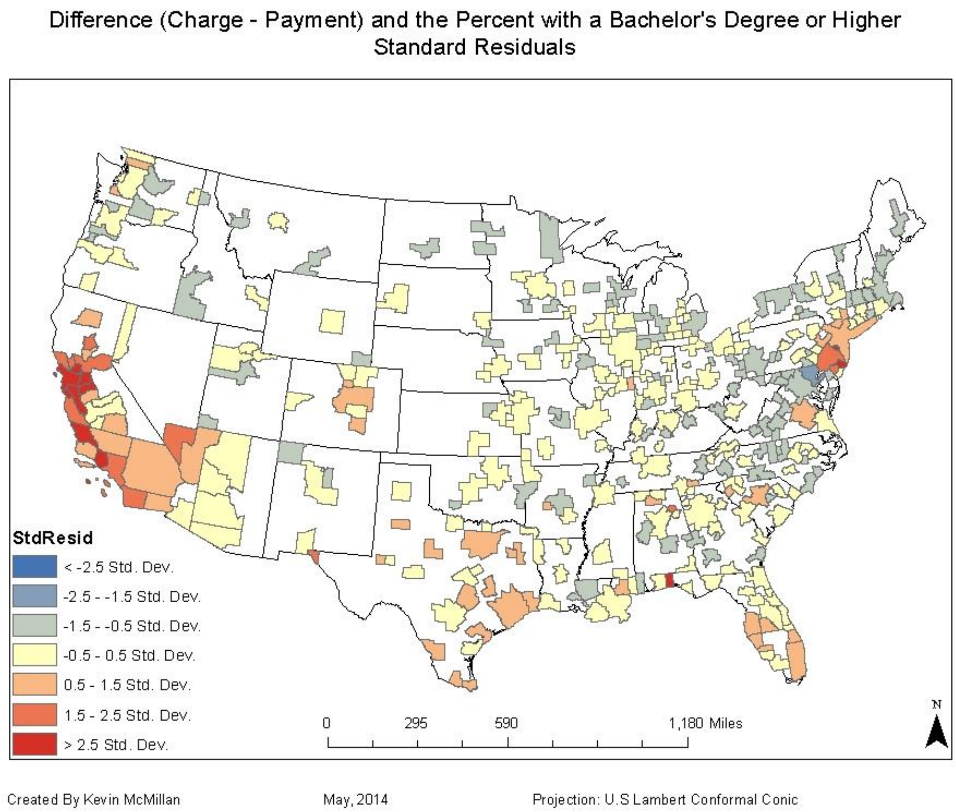
Map 11: A Map of Standard Residuals from OLS Regression using Average Medicare Payments and the Percent Aged Sixty-Five and Older (Source- Author)



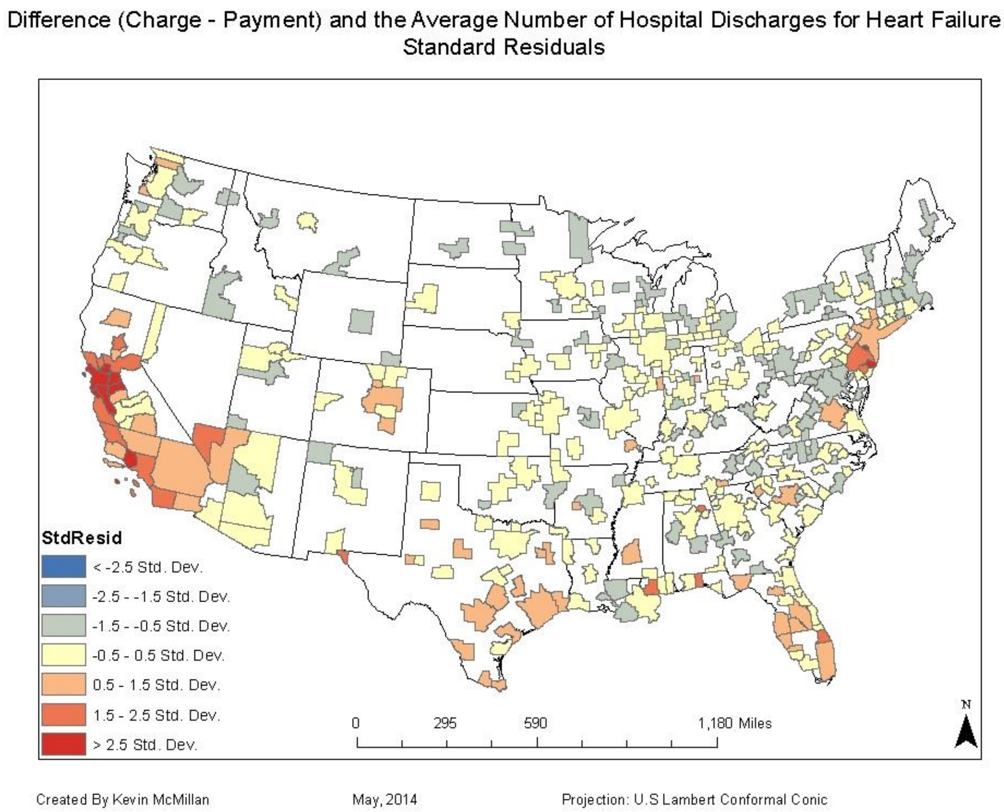
Map 12: A Map of Standard Residuals from OLS Regression using Average Medicare Payments and the Percent in Poverty (Source- Author)



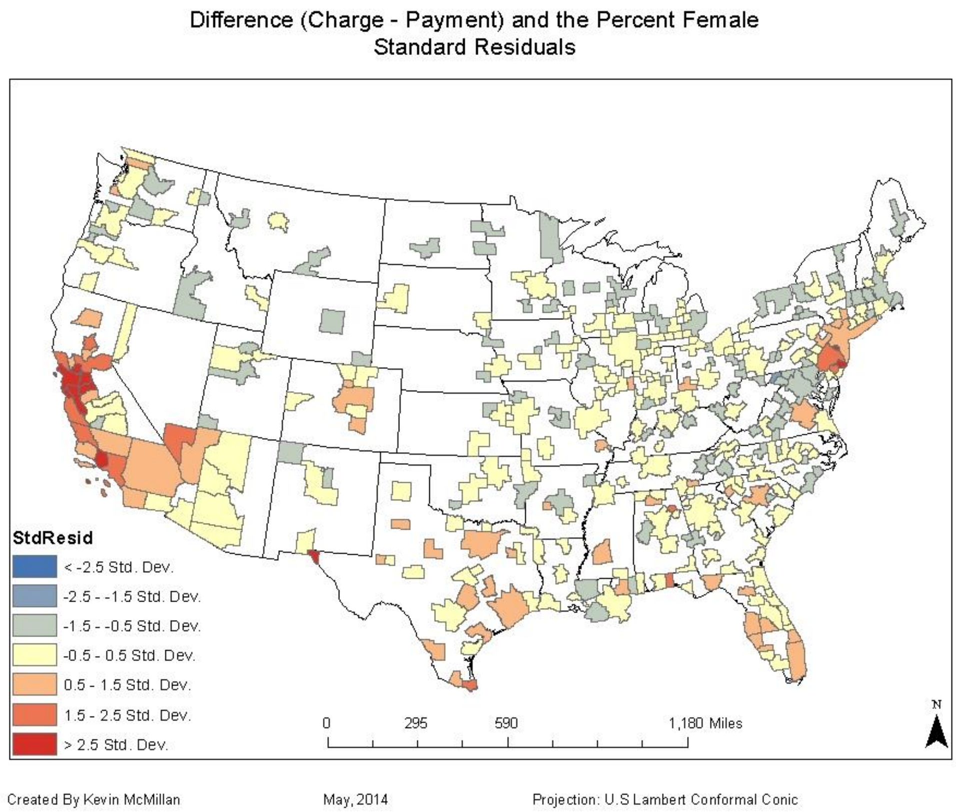
Map 13: A Map of Standard Residuals from OLS Regression using the Difference between Average Hospital Charges and Average Medicare Payments (Charge – Payment) and the Percent African American (Source- Author)



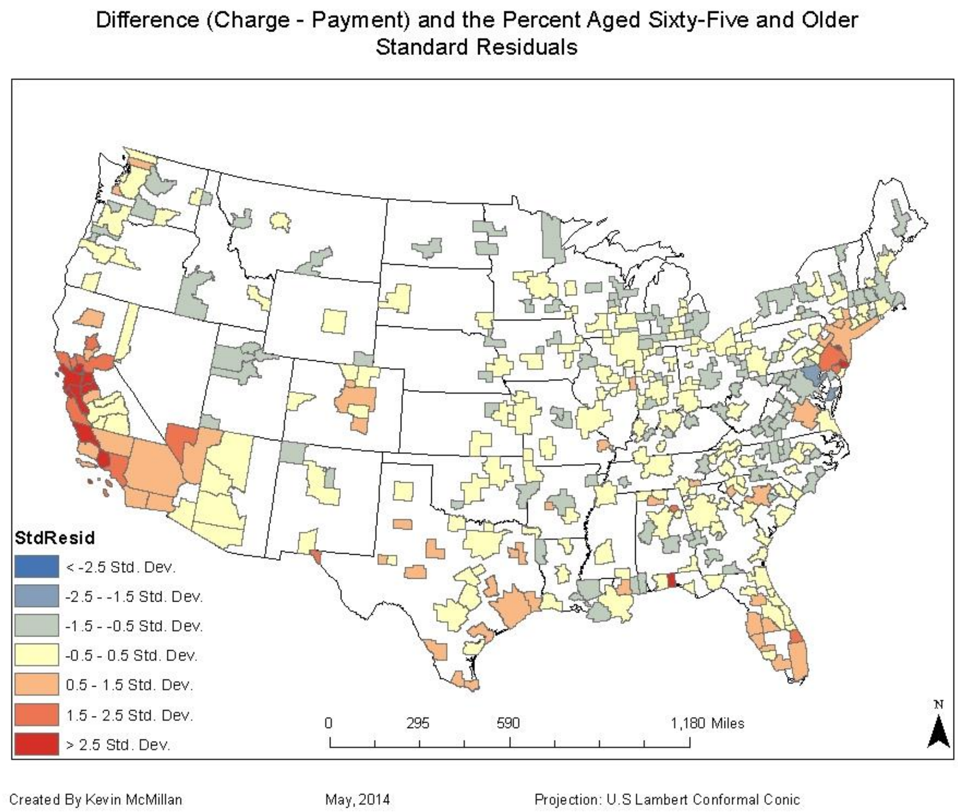
Map 14: A Map of Standard Residuals from OLS Regression using the Difference between Average Hospital Charges and Average Medicare Payments (Charge – Payment) and the Percent with a Bachelor's Degree or Higher (Source- Author)



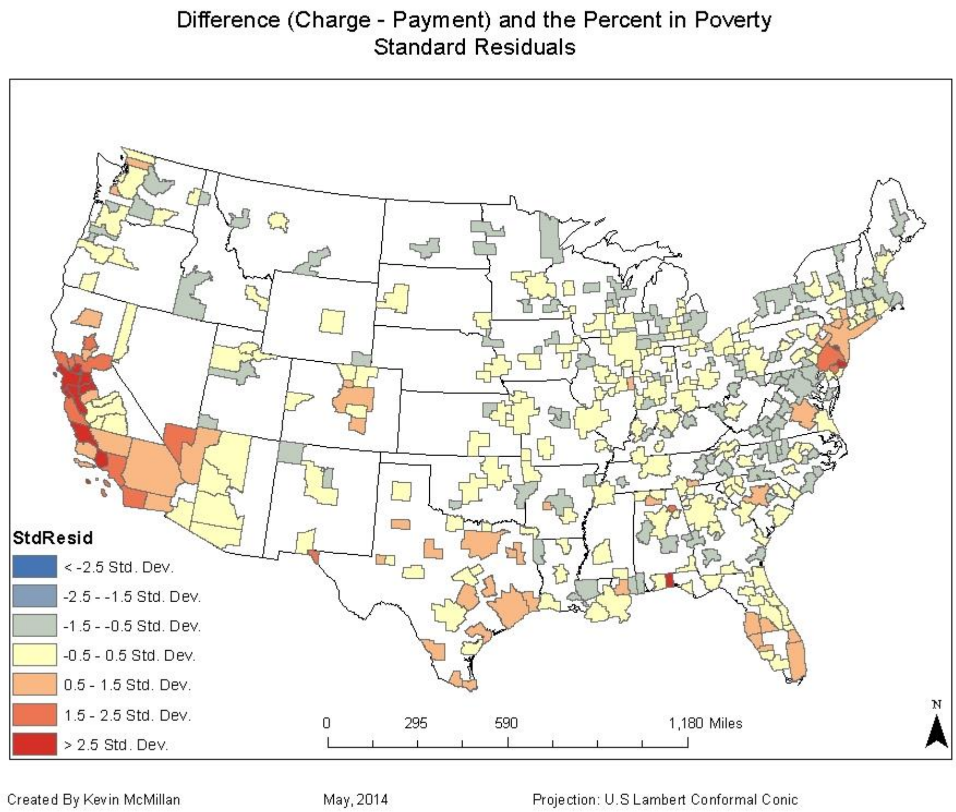
Map 15: A Map of Standard Residuals from OLS Regression using the Difference between Average Hospital Charges and Average Medicare Payments (Charge – Payment) and the Average Number of Hospital Discharges for Heart Failure (Source- Author)



Map 16: A Map of Standard Residuals from OLS Regression using the Difference between Average Hospital Charges and Average Medicare Payments (Charge – Payment) and the Percent Female (Source- Author)



Map 17: A Map of Standard Residuals from OLS Regression using the Difference between Average Hospital Charges and Average Medicare Payments (Charge – Payment) and the Percent Aged Sixty-Five and Older (Source- Author)



Map 18: A Map of Standard Residuals from OLS Regression using the Difference between Average Hospital Charges and Average Medicare Payments (Charge – Payment) and the Percent in Poverty (Source- Author)

Ordinary Least Squares (OLS) Regression Reports

Average Hospital Charge Regression Reports

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	16375.400528	606.216782	27.012450	0.000000*	577.879421	28.337054	0.000000*
AF_AMER	-28.459909	38.006564	-0.748816	0.454457	28.770251	-0.989213	0.323223

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	CHARGED
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7389.782095
Multiple R-Squared [d]:	0.001581	Adjusted R-Squared [d]:	-0.001239
Joint F-Statistic [e]:	0.560725	Prob(>F), (1,354) degrees of freedom:	0.454466
Joint Wald Statistic [e]:	0.978543	Prob(>chi-squared), (1) degrees of freedom:	0.322559
Koenker (BP) Statistic [f]:	0.573507	Prob(>chi-squared), (1) degrees of freedom:	0.448869
Jarque-Bera Statistic [g]:	2164.450174	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

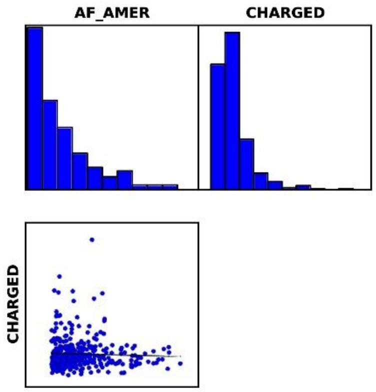
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

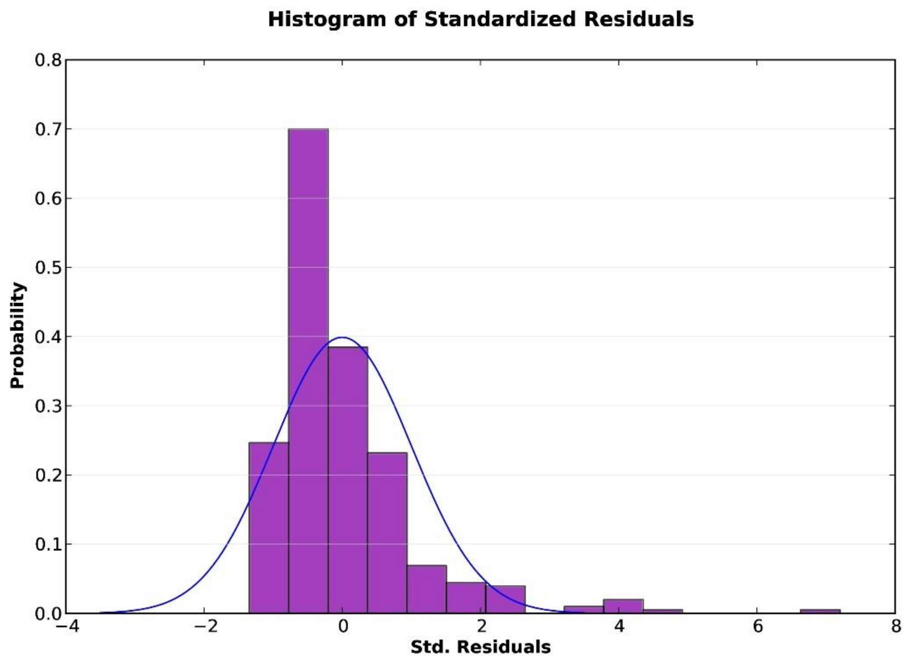
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships



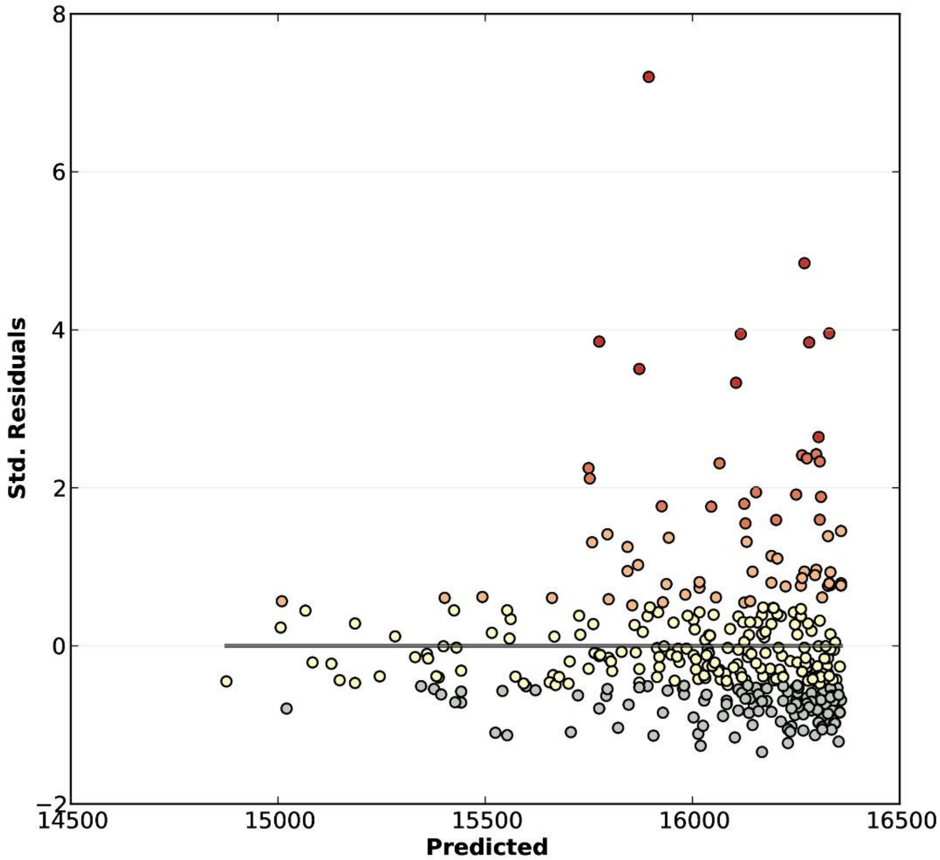
The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.

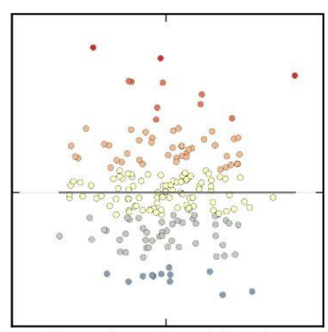


Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).

Residual vs. Predicted Plot



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	CHARGED
Explanatory Variables	AF_AMER
Selection Set	False

Report 1: OLS Report for Average Hospital Charges for Heart Failure and Percent African American

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	15045.023150	1398.669930	10.756664	0.000000*	1545.323870	9.735838	0.000000*
BA_HIGHER	39.012215	52.364993	0.745006	0.456755	60.053200	0.649628	0.516353

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	CHARGED
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7389.787809
Multiple R-Squared [d]:	0.001565	Adjusted R-Squared [d]:	-0.001255
Joint F-Statistic [e]:	0.555033	Prob(>F), (1,354) degrees of freedom:	0.456763
Joint Wald Statistic [e]:	0.422016	Prob(>chi-squared), (1) degrees of freedom:	0.515933
Koenker (BP) Statistic [f]:	0.399136	Prob(>chi-squared), (1) degrees of freedom:	0.527536
Jarque-Bera Statistic [g]:	2151.995827	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

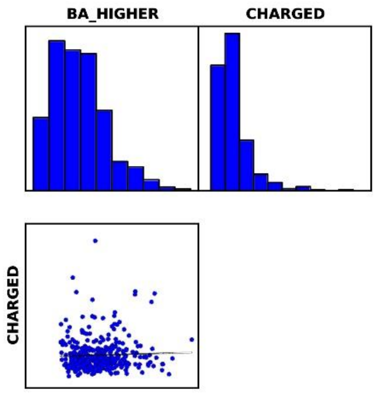
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

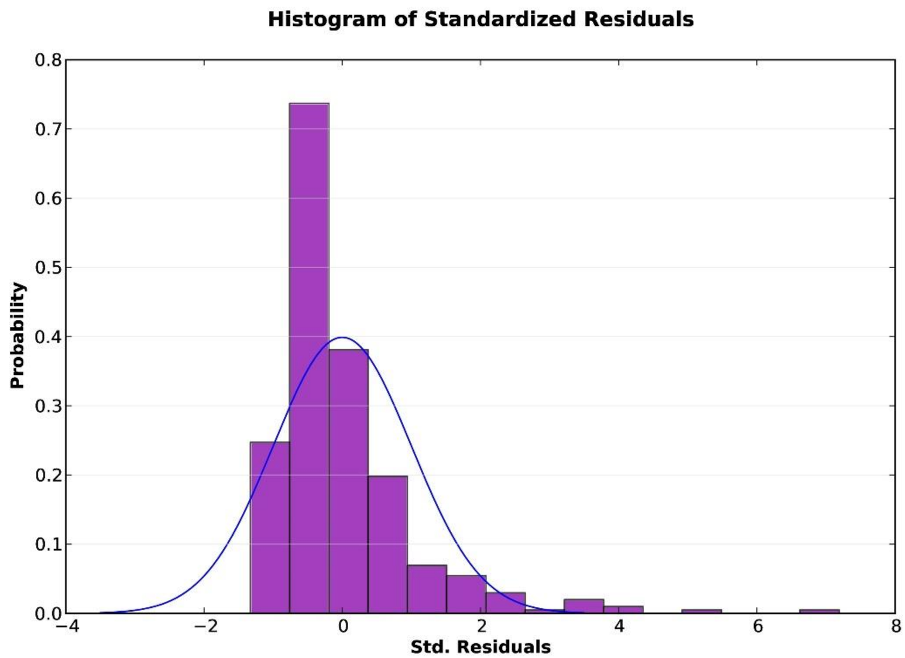
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

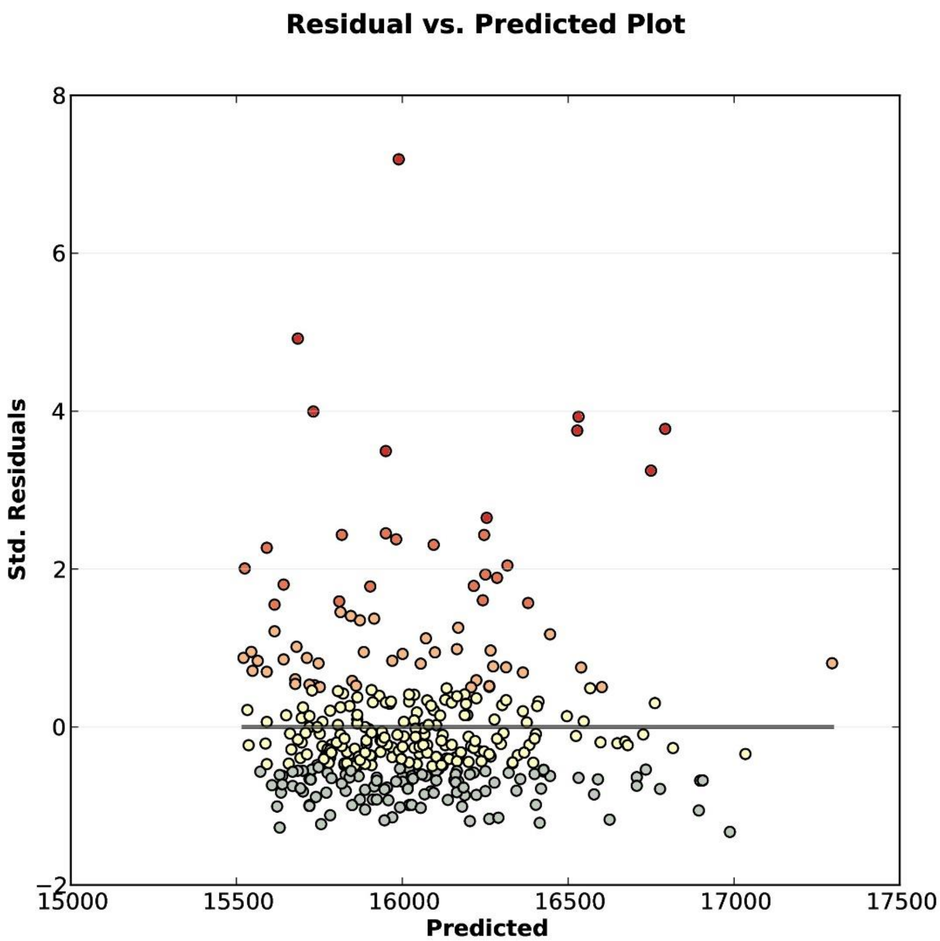


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

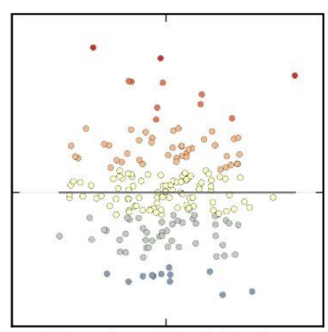
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	CHARGED
Explanatory Variables	BA_HIGHER
Selection Set	False

Report 2: OLS Report for Average Hospital Charges for Heart Failure and Percent with a Bachelor's Degree or Higher

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	17944.053006	1002.729944	17.895200	0.000000*	1175.935675	15.259383	0.000000*
DISCHARGES	-48.919487	23.547070	-2.077519	0.038464*	24.158274	-2.024958	0.043614*

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	CHARGED
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7386.031319
Multiple R-Squared [d]:	0.012045	Adjusted R-Squared [d]:	0.009255
Joint F-Statistic [e]:	4.316085	Prob(>F), (1,354) degrees of freedom:	0.038474*
Joint Wald Statistic [e]:	4.100454	Prob(>chi-squared), (1) degrees of freedom:	0.042872*
Koenker (BP) Statistic [f]:	5.894873	Prob(>chi-squared), (1) degrees of freedom:	0.015185*
Jarque-Bera Statistic [g]:	1927.881246	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

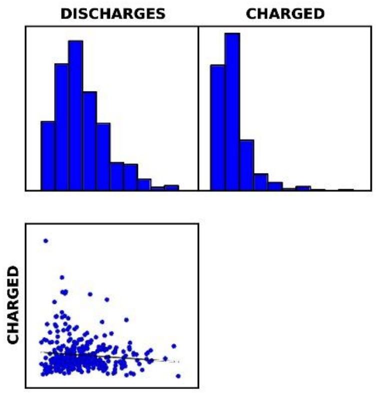
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

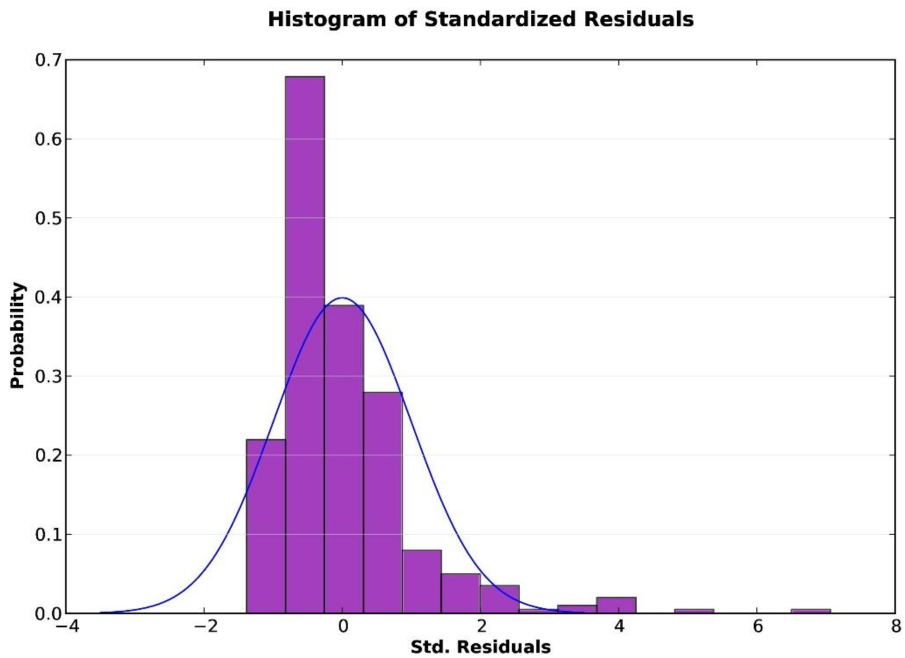
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

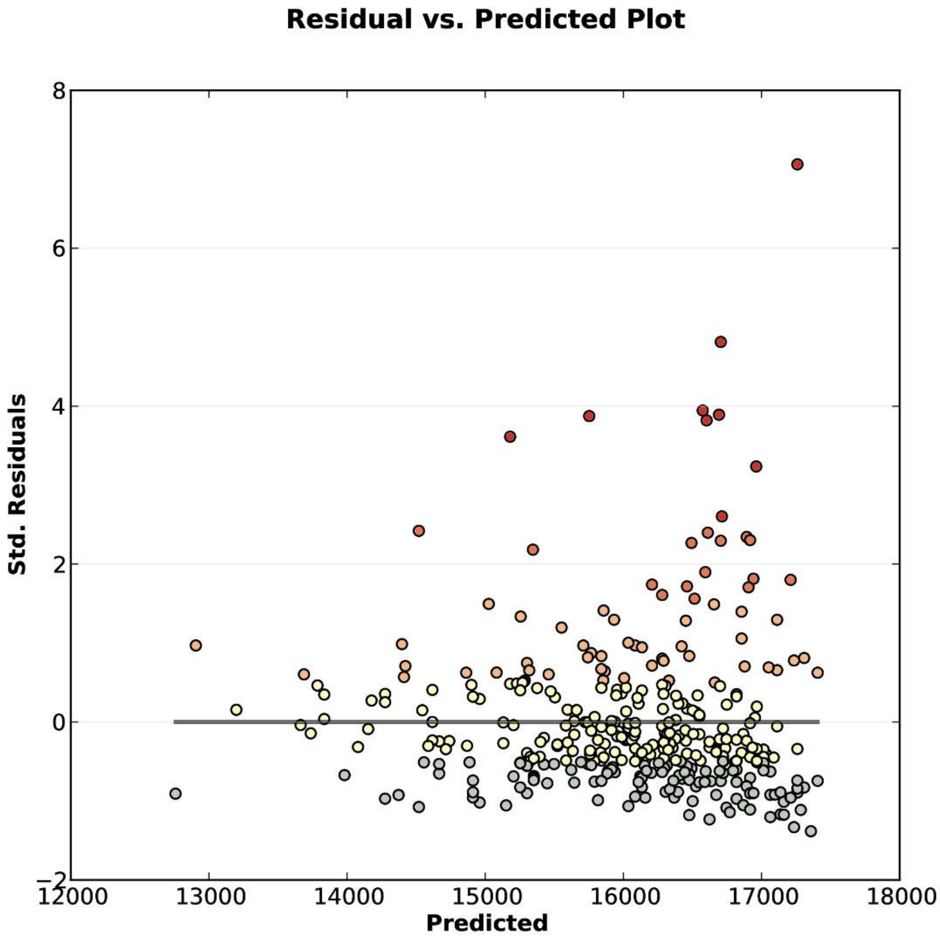


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

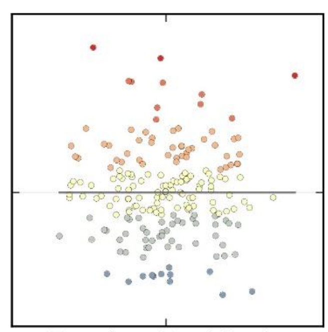
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	CHARGED
Explanatory Variables	DISCHARGES
Selection Set	False

Report 3: OLS Report for Average Hospital Charges for Heart Failure and the Average Number of Hospital Discharges for Heart Failure

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	67055.274863	20773.554686	3.227915	0.001376*	20416.357917	3.284390	0.001137*
FEMALE	-1004.875204	409.118904	-2.456194	0.014510*	399.171261	-2.517404	0.012253*

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	CHARGED
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7384.329688
Multiple R-Squared [d]:	0.016756	Adjusted R-Squared [d]:	0.013979
Joint F-Statistic [e]:	6.032887	Prob(>F), (1,354) degrees of freedom:	0.014521*
Joint Wald Statistic [e]:	6.337321	Prob(>chi-squared), (1) degrees of freedom:	0.011822*
Koenker (BP) Statistic [f]:	3.603650	Prob(>chi-squared), (1) degrees of freedom:	0.057653
Jarque-Bera Statistic [g]:	2067.098006	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

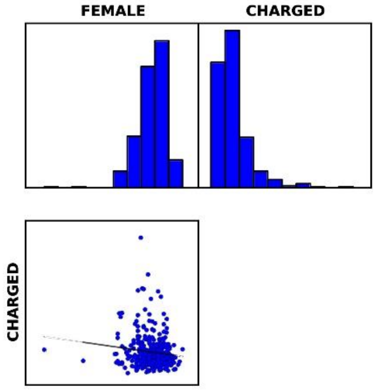
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

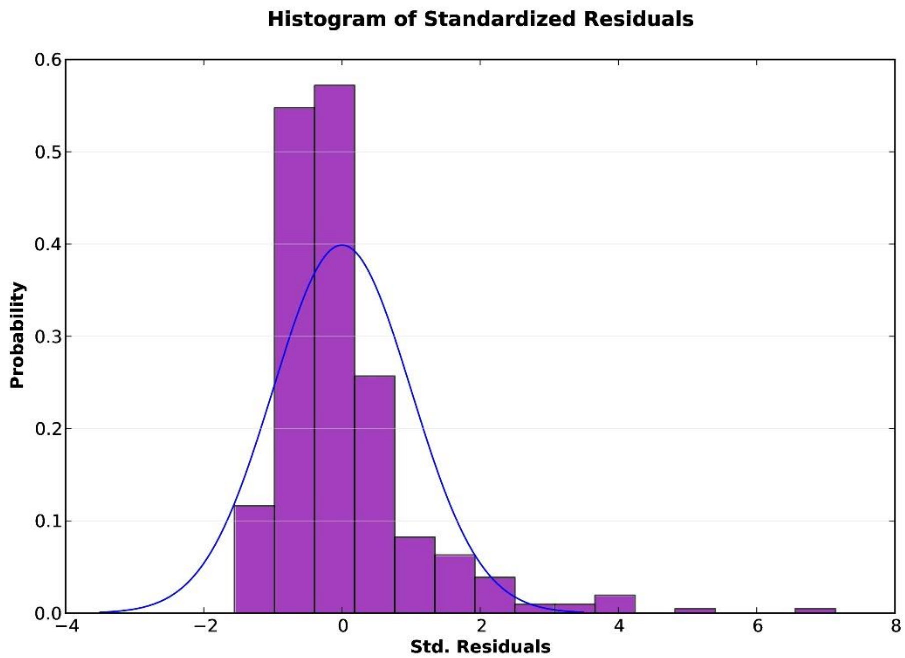
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

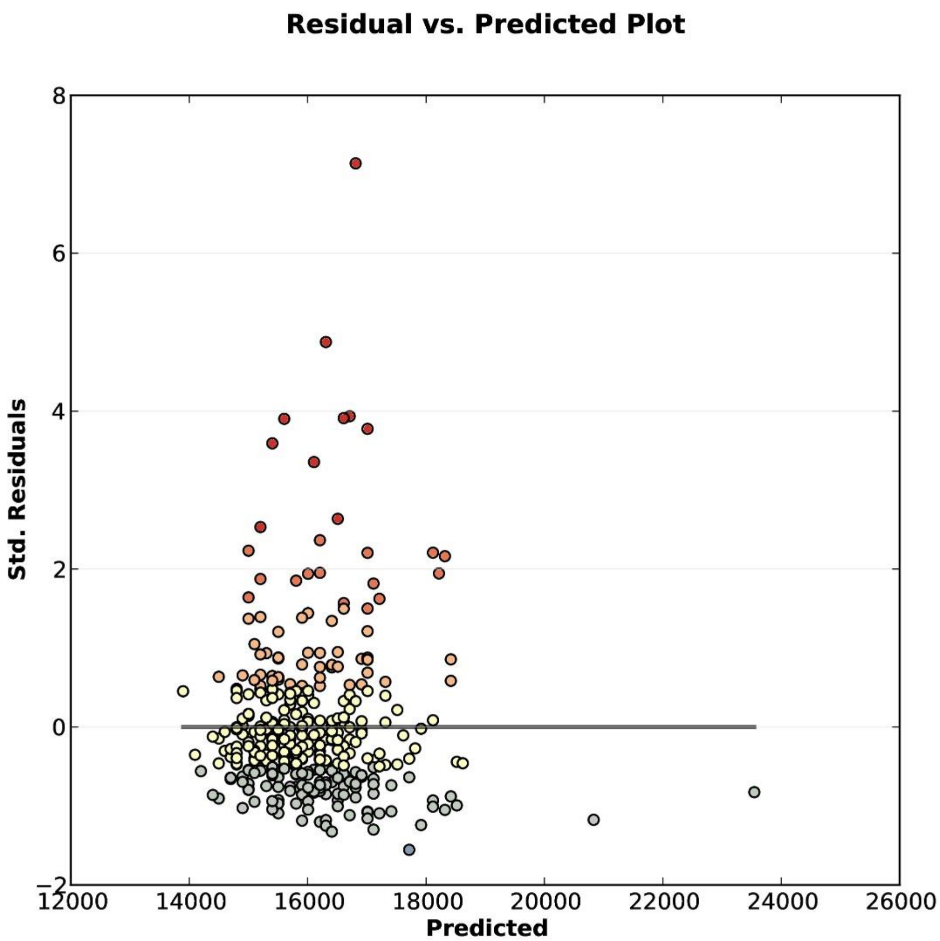


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

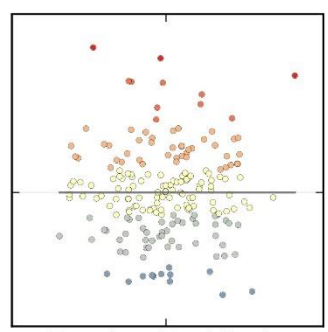
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	CHARGED
Explanatory Variables	FEMALE
Selection Set	False

Report 4: OLS Report for Average Hospital Charges for Heart Failure and Percent Female

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	15041.837469	1614.688732	9.315627	0.000000*	1788.222383	8.411615	0.000000*
POVERTY	65.773354	102.781740	0.639932	0.522630	107.931516	0.609399	0.542653

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	CHARGED
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7389.933953
Multiple R-Squared [d]:	0.001155	Adjusted R-Squared [d]:	-0.001666
Joint F-Statistic [e]:	0.409513	Prob(>F), (1,354) degrees of freedom:	0.522631
Joint Wald Statistic [e]:	0.371367	Prob(>chi-squared), (1) degrees of freedom:	0.542260
Koenker (BP) Statistic [f]:	1.948466	Prob(>chi-squared), (1) degrees of freedom:	0.162752
Jarque-Bera Statistic [g]:	2243.791148	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

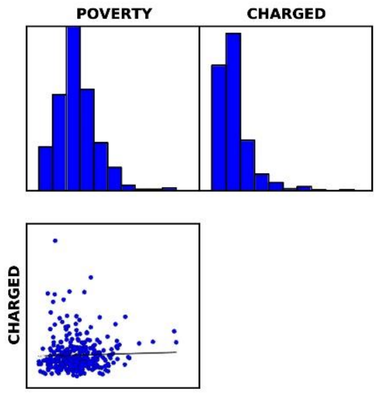
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

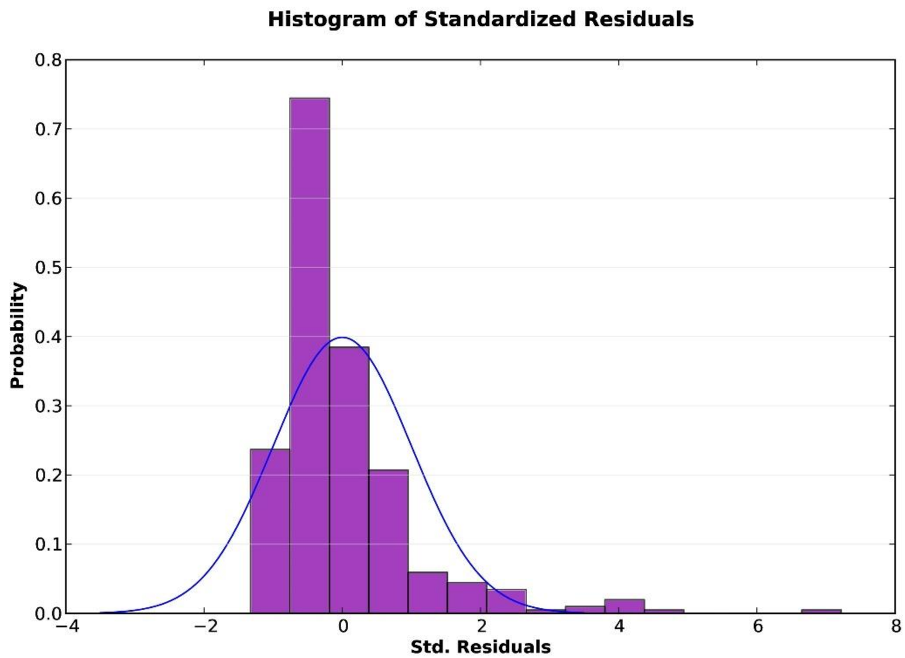
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

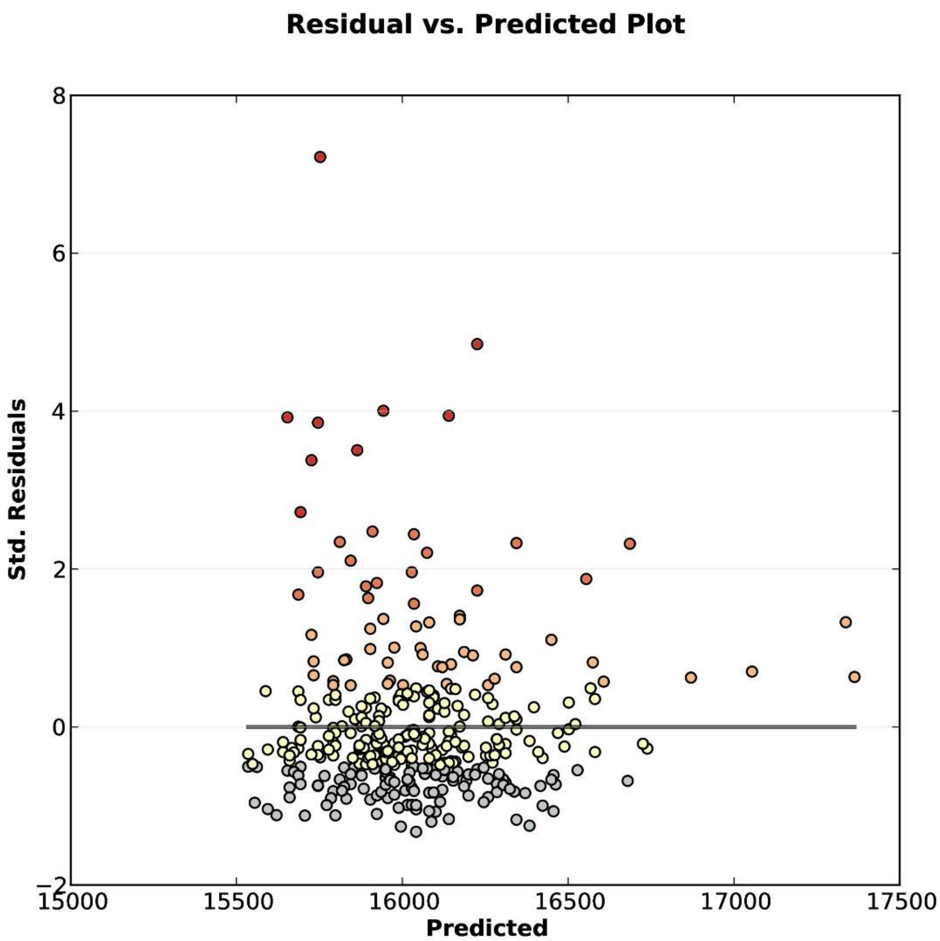


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

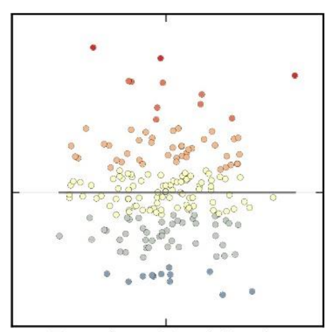
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	CHARGED
Explanatory Variables	POVERTY
Selection Set	False

Report 5: OLS Report for Average Hospital Charges for Heart Failure and Percent in Poverty

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	20031.906199	1705.683843	11.744208	0.000000*	1732.349491	11.563432	0.000000*
SIXFIVE_OL	-297.776110	123.592659	-2.409335	0.016479*	117.240730	-2.539869	0.011507*

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	CHARGED
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7384.555197
Multiple R-Squared [d]:	0.016133	Adjusted R-Squared [d]:	0.013354
Joint F-Statistic [e]:	5.804895	Prob(>F), (1,354) degrees of freedom:	0.016491*
Joint Wald Statistic [e]:	6.450935	Prob(>chi-squared), (1) degrees of freedom:	0.011089*
Koenker (BP) Statistic [f]:	2.142301	Prob(>chi-squared), (1) degrees of freedom:	0.143287
Jarque-Bera Statistic [g]:	2021.261100	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

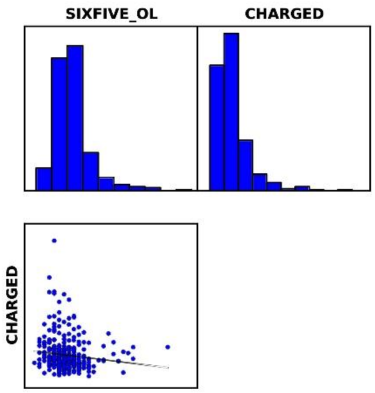
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

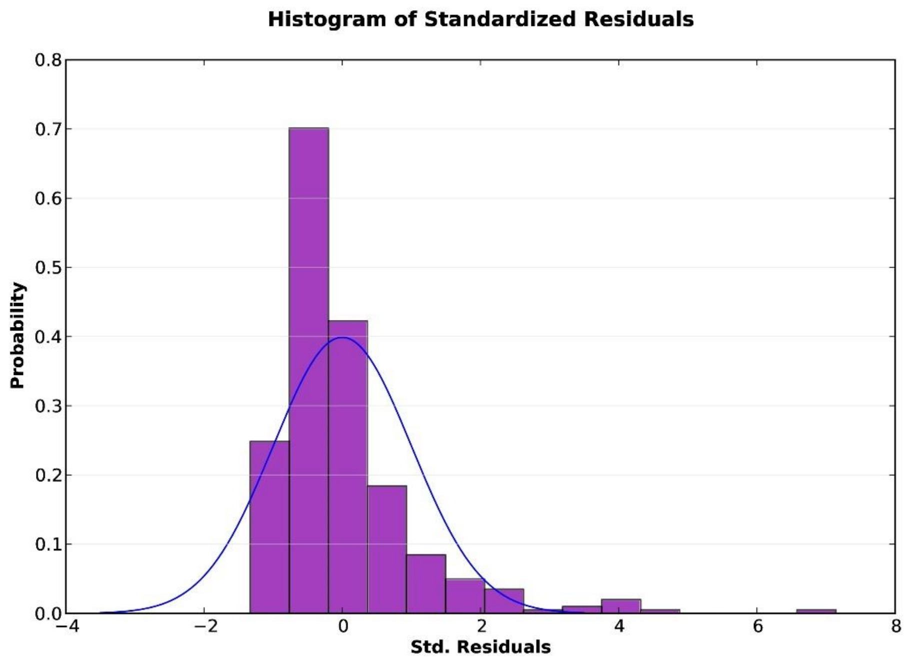
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

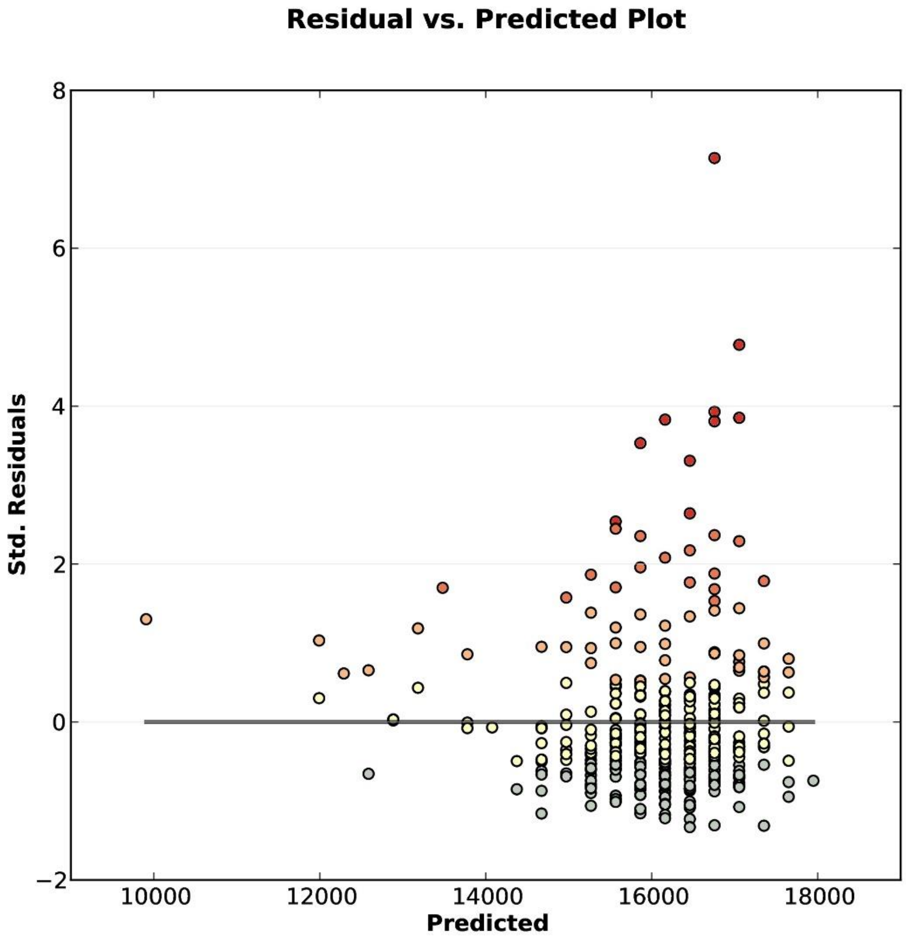


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

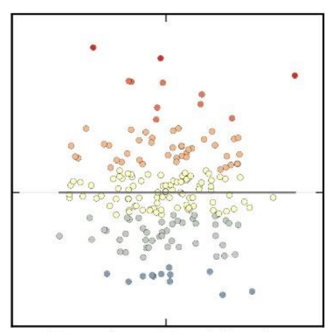
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	CHARGED
Explanatory Variables	SIXFIVE_OL
Selection Set	False

Report 6: OLS Report for Average Hospital Charges for Heart Failure and Percent Aged Sixty-Five and Older

Average Medicare Payment Regression Reports

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	4589.664856	56.618078	81.063594	0.000000*	58.803301	78.051143	0.000000*
AF_AMER	-4.361013	3.549652	-1.228575	0.220050	3.082597	-1.414720	0.158041

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	PAID
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	5701.694659
Multiple R-Squared [d]:	0.004246	Adjusted R-Squared [d]:	0.001433
Joint F-Statistic [e]:	1.509396	Prob(>F), (1,354) degrees of freedom:	0.220047
Joint Wald Statistic [e]:	2.001434	Prob(>chi-squared), (1) degrees of freedom:	0.157150
Koenker (BP) Statistic [f]:	1.329674	Prob(>chi-squared), (1) degrees of freedom:	0.248863
Jarque-Bera Statistic [g]:	1959.087340	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

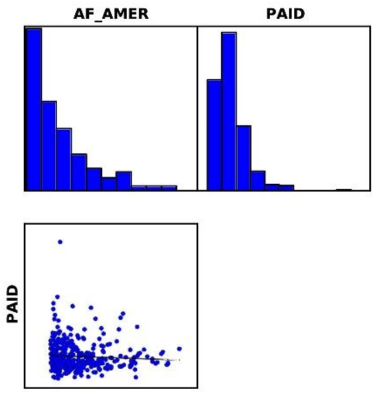
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

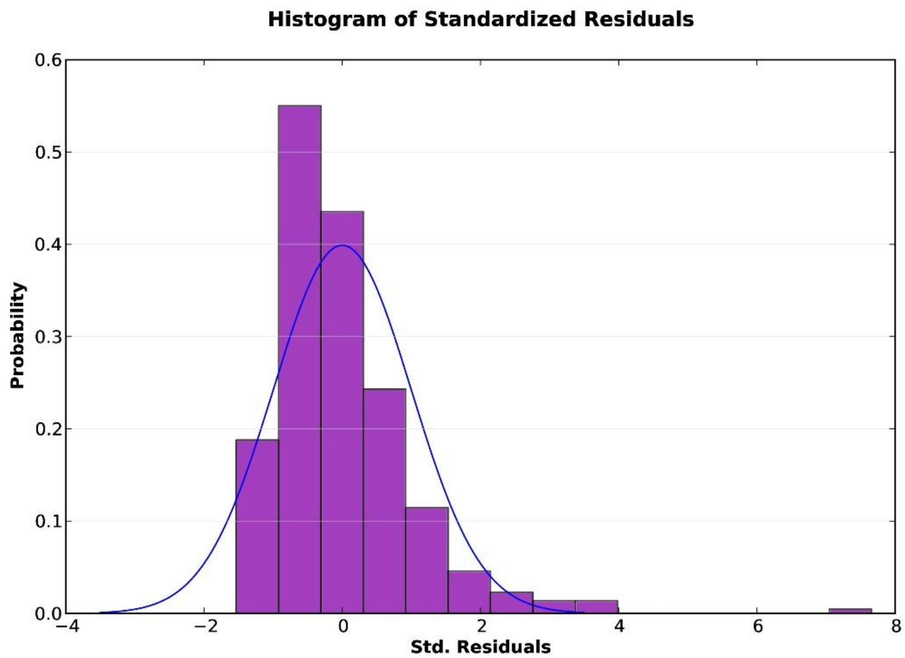
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

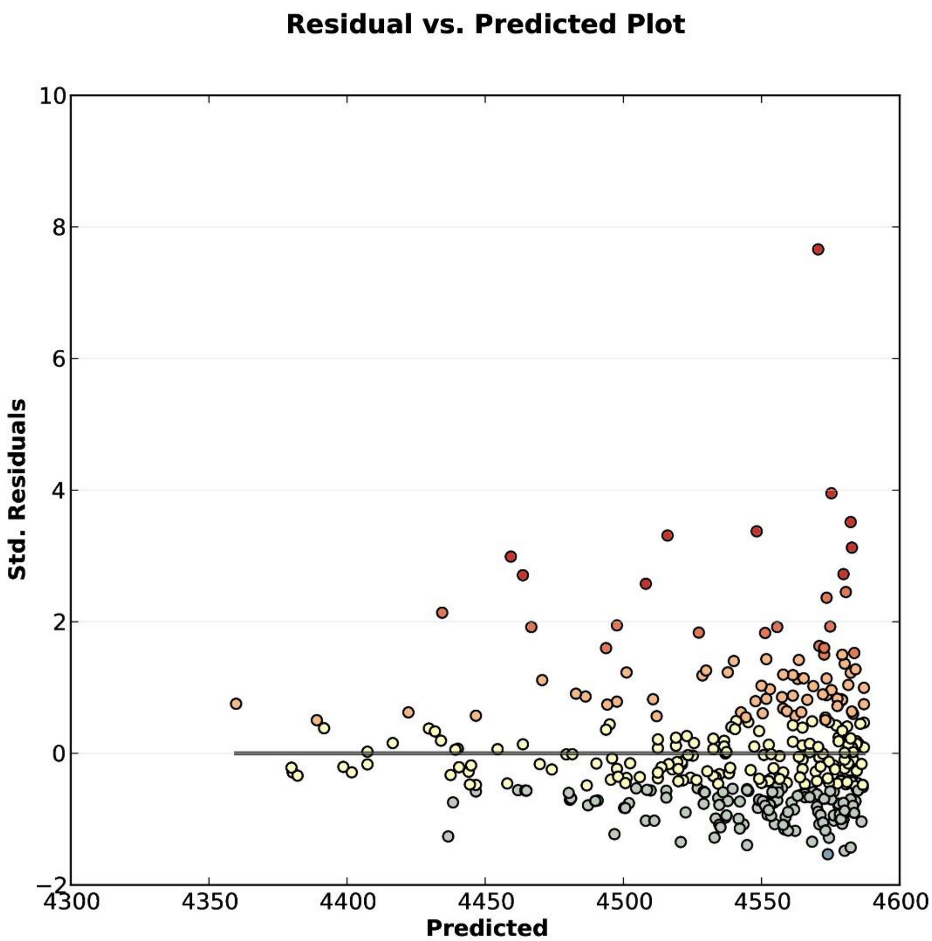


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

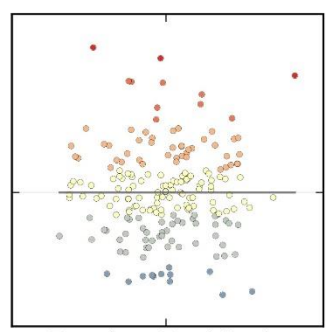
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	PAID
Explanatory Variables	AF_AMER
Selection Set	False

Report 7: OLS Report for Average Medicare Payments for Heart Failure and Percent African American

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	4089.666303	128.507275	31.824395	0.000000*	177.439617	23.048214	0.000000*
BA_HIGHER	17.575348	4.811201	3.653006	0.000310*	6.761853	2.599191	0.009728*

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	PAID
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	5690.036266
Multiple R-Squared [d]:	0.036327	Adjusted R-Squared [d]:	0.033605
Joint F-Statistic [e]:	13.344455	Prob(>F), (1,354) degrees of freedom:	0.000298*
Joint Wald Statistic [e]:	6.755795	Prob(>chi-squared), (1) degrees of freedom:	0.009344*
Koenker (BP) Statistic [f]:	0.018884	Prob(>chi-squared), (1) degrees of freedom:	0.890700
Jarque-Bera Statistic [g]:	2816.938904	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

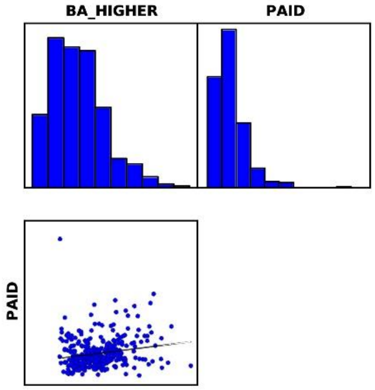
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

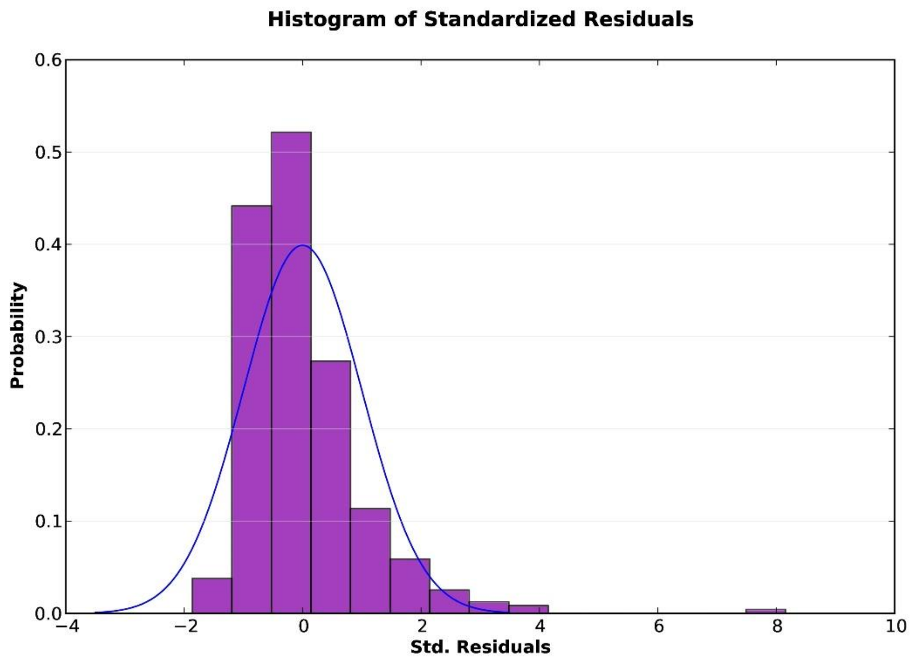
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

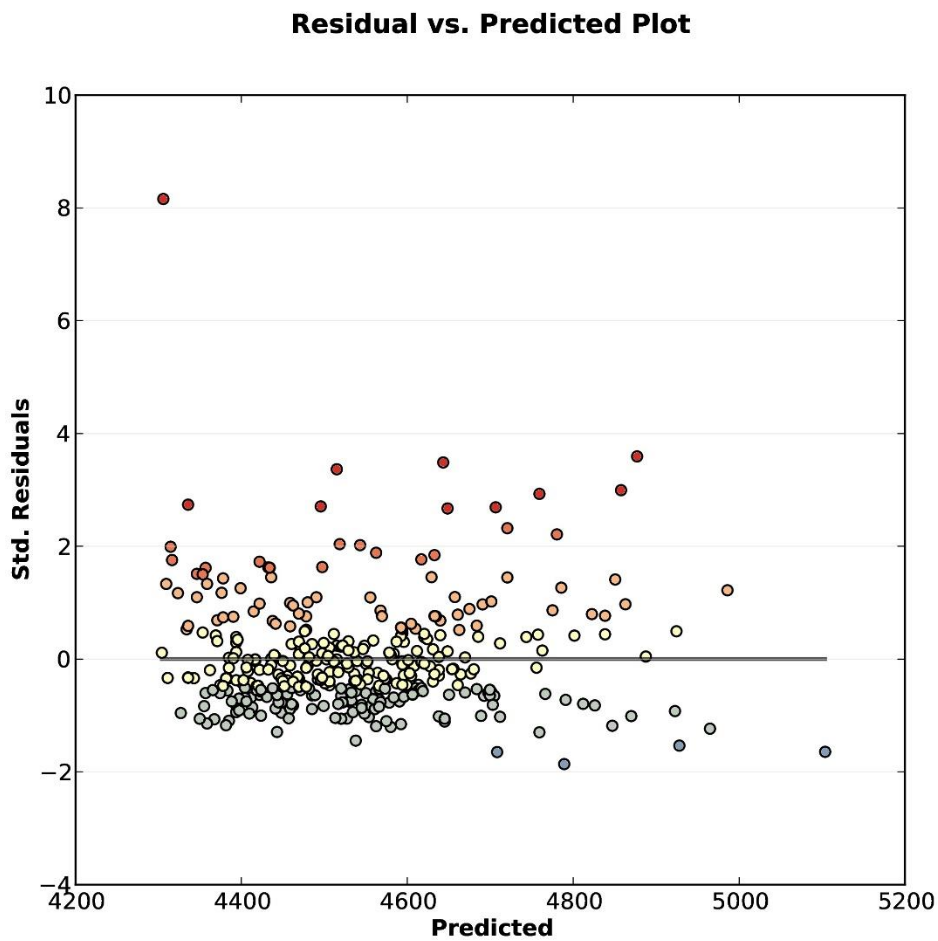


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

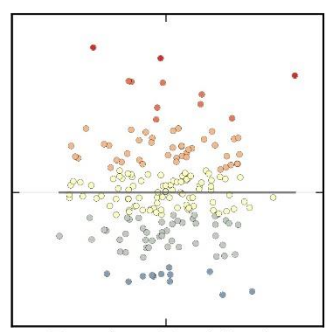
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	PAID
Explanatory Variables	BA_HIGHER
Selection Set	False

Report 8: OLS Report for Average Medicare Payments for Heart Failure and Percent with a Bachelor's Degree or Higher

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	4485.291020	94.295144	47.566511	0.000000*	118.859357	37.736121	0.000000*
DISCHARGES	1.366764	2.214329	0.617236	0.537478	2.651295	0.515508	0.606530

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	PAID
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	5702.826429
Multiple R-Squared [d]:	0.001075	Adjusted R-Squared [d]:	-0.001747
Joint F-Statistic [e]:	0.380980	Prob(>F), (1,354) degrees of freedom:	0.537476
Joint Wald Statistic [e]:	0.265748	Prob(>chi-squared), (1) degrees of freedom:	0.606198
Koenker (BP) Statistic [f]:	2.601254	Prob(>chi-squared), (1) degrees of freedom:	0.106779
Jarque-Bera Statistic [g]:	2126.182406	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value (p < 0.05).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant (p < 0.05); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

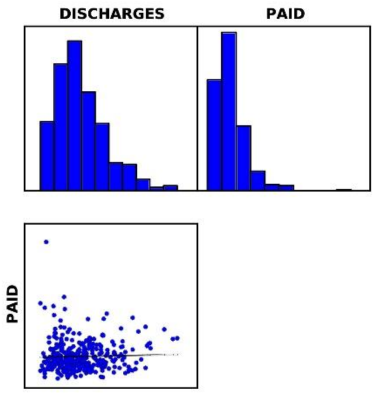
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance (p < 0.05); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant (p < 0.05), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

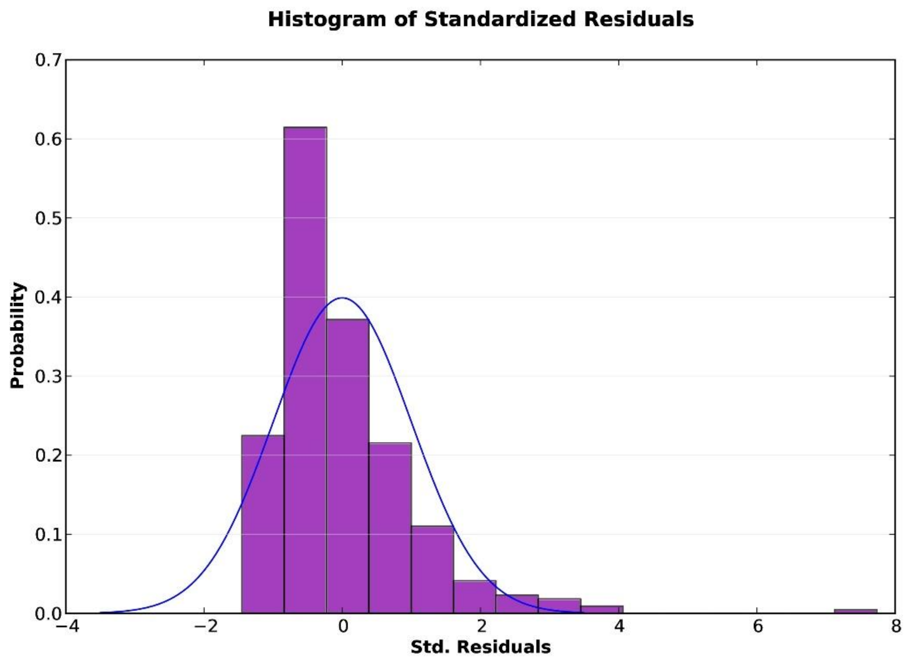
[g] Jarque-Bera Statistic: When this test is statistically significant (p < 0.05) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

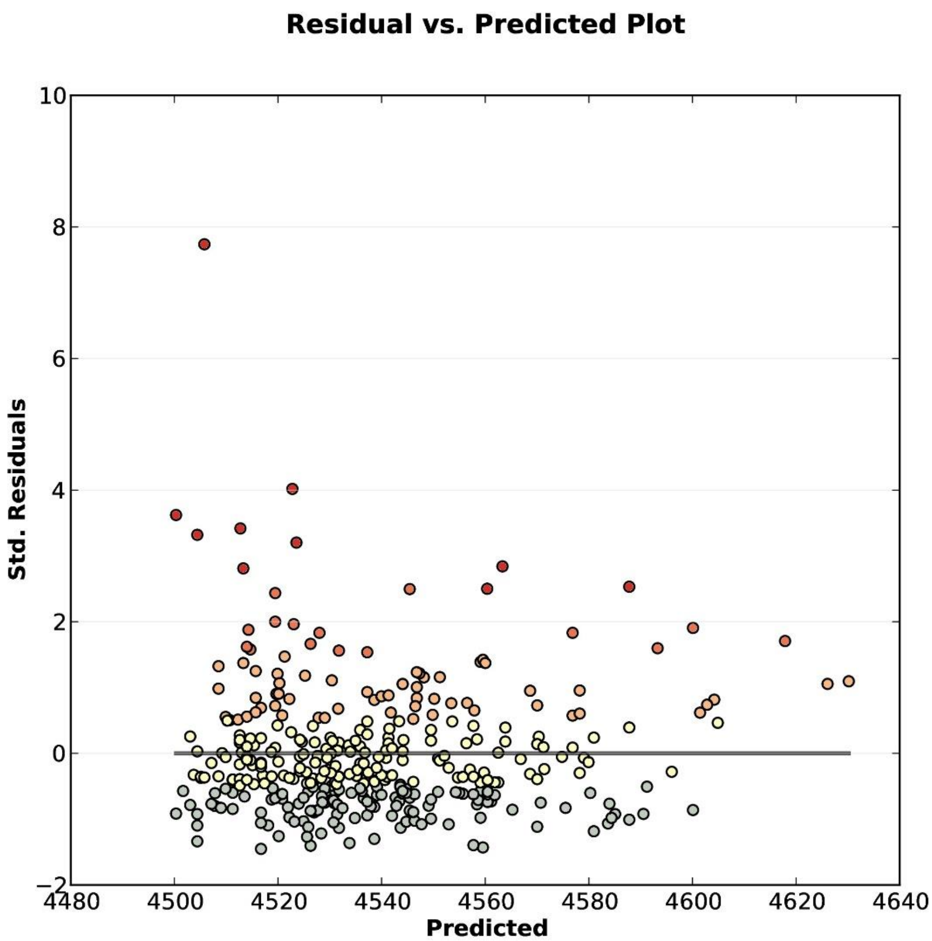


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

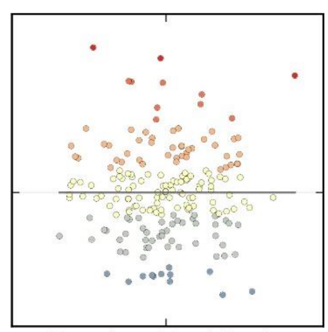
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	PAID
Explanatory Variables	DISCHARGES
Selection Set	False

Report 9: OLS Report for Average Medicare Payments for Heart Failure and the Average Number of Hospital Discharges for Heart Failure

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	7300.369446	1953.731067	3.736630	0.000228*	1928.470012	3.785576	0.000190*
FEMALE	-54.404197	38.477205	-1.413933	0.158272	37.797851	-1.439346	0.150948

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	PAID
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	5701.204510
Multiple R-Squared [d]:	0.005616	Adjusted R-Squared [d]:	0.002807
Joint F-Statistic [e]:	1.999207	Prob(>F), (1,354) degrees of freedom:	0.158260
Joint Wald Statistic [e]:	2.071718	Prob(>chi-squared), (1) degrees of freedom:	0.150052
Koenker (BP) Statistic [f]:	1.632928	Prob(>chi-squared), (1) degrees of freedom:	0.201299
Jarque-Bera Statistic [g]:	1914.186433	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value (p < 0.05).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant (p < 0.05); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

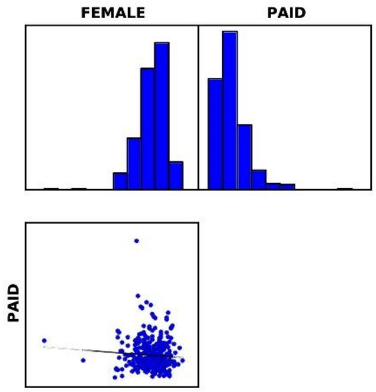
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance (p < 0.05); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant (p < 0.05), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

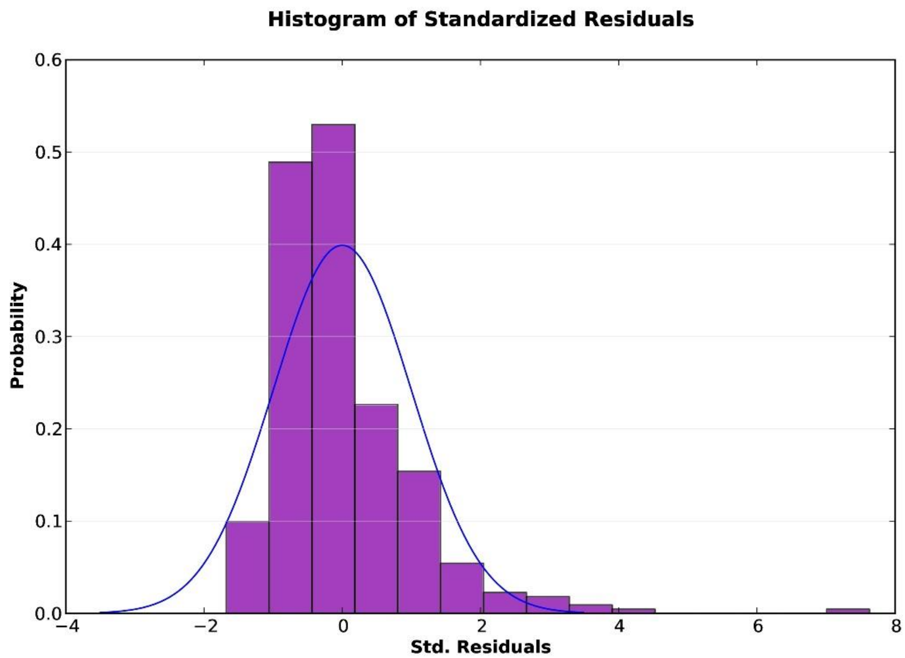
[g] Jarque-Bera Statistic: When this test is statistically significant (p < 0.05) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

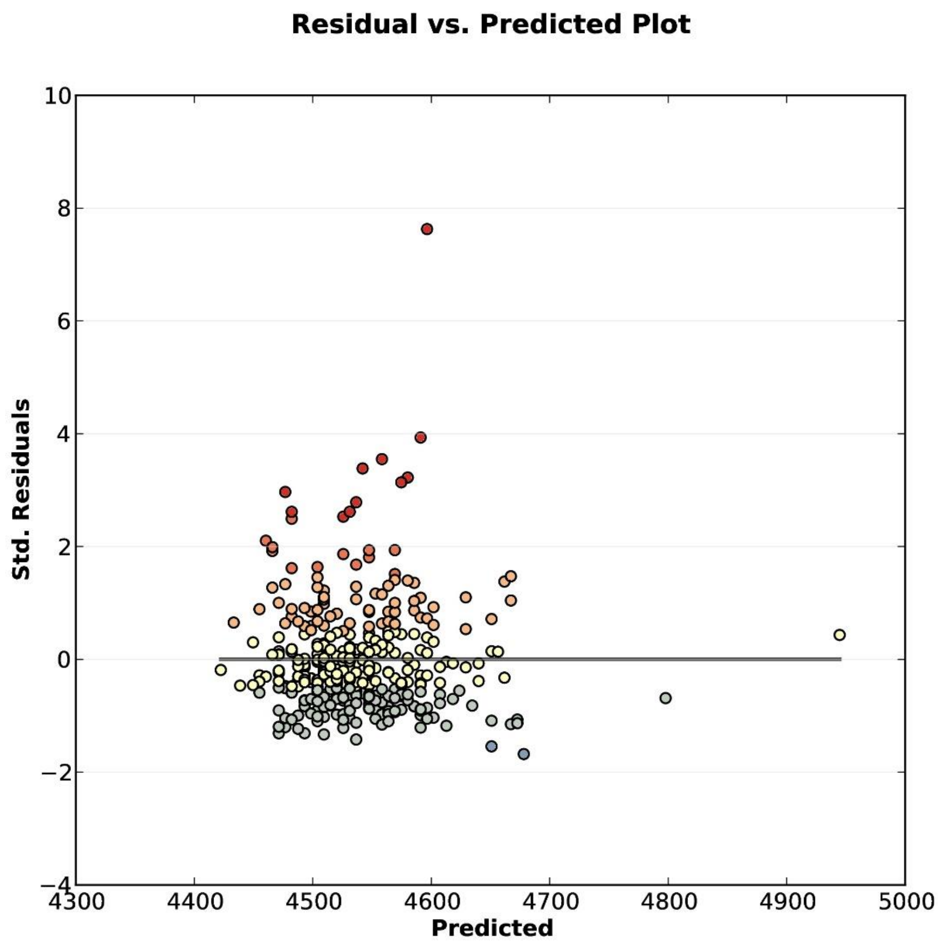


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

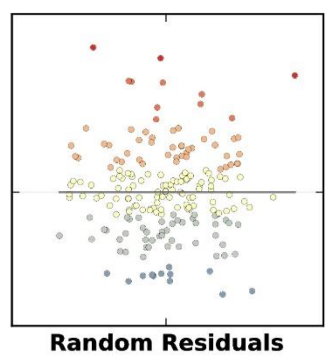
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	PAID
Explanatory Variables	FEMALE
Selection Set	False

Report 10: OLS Report for Average Medicare Payments for Heart Failure and Percent Female

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	4552.315710	151.092095	30.129410	0.000000*	185.987063	24.476518	0.000000*
POVERTY	-0.912256	9.617648	-0.094852	0.924472	12.350329	-0.073865	0.941146

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	PAID
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	5703.200309
Multiple R-Squared [d]:	0.000025	Adjusted R-Squared [d]:	-0.002799
Joint F-Statistic [e]:	0.008997	Prob(>F), (1,354) degrees of freedom:	0.924486
Joint Wald Statistic [e]:	0.005456	Prob(>chi-squared), (1) degrees of freedom:	0.941118
Koenker (BP) Statistic [f]:	0.978570	Prob(>chi-squared), (1) degrees of freedom:	0.322552
Jarque-Bera Statistic [g]:	2028.500262	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value (p < 0.05).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant (p < 0.05); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

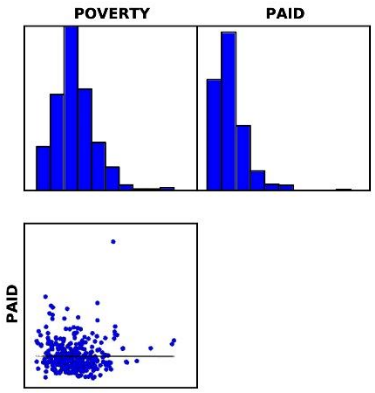
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance (p < 0.05); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant (p < 0.05), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

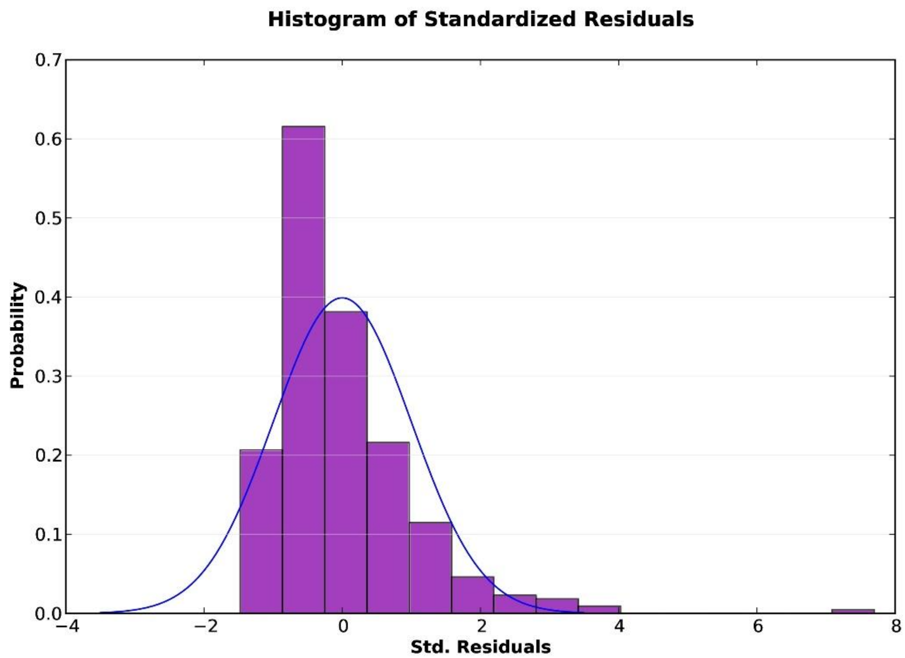
[g] Jarque-Bera Statistic: When this test is statistically significant (p < 0.05) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

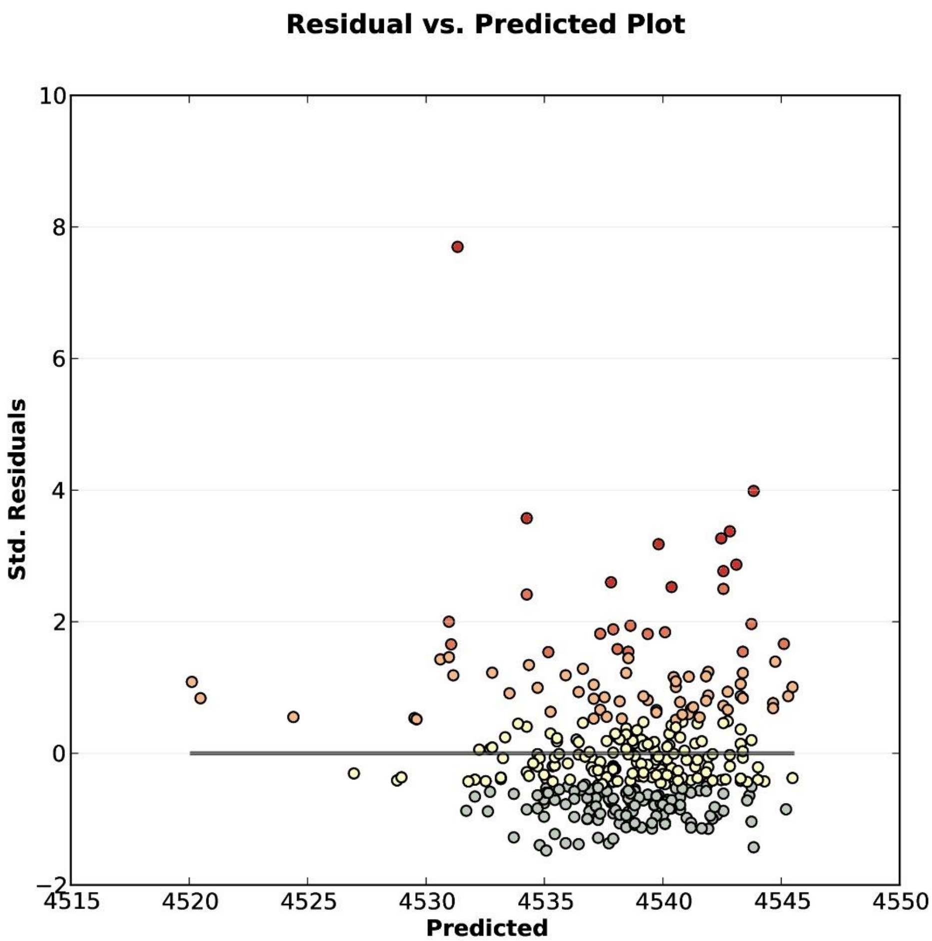


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

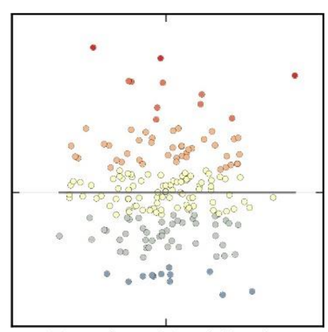
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	PAID
Explanatory Variables	POVERTY
Selection Set	False

Report 11: OLS Report for Average Medicare Payments for Heart Failure and Percent in Poverty

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	5309.746758	155.179661	34.216770	0.000000*	166.248234	31.938666	0.000000*
SIXFIVE_OL	-57.551863	11.244210	-5.118356	0.000001*	10.942064	-5.259690	0.000000*

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	PAID
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	5677.793054
Multiple R-Squared [d]:	0.068905	Adjusted R-Squared [d]:	0.066275
Joint F-Statistic [e]:	26.197563	Prob(>F), (1,354) degrees of freedom:	0.000001*
Joint Wald Statistic [e]:	27.664336	Prob(>chi-squared), (1) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	5.407729	Prob(>chi-squared), (1) degrees of freedom:	0.020048*
Jarque-Bera Statistic [g]:	1817.446034	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

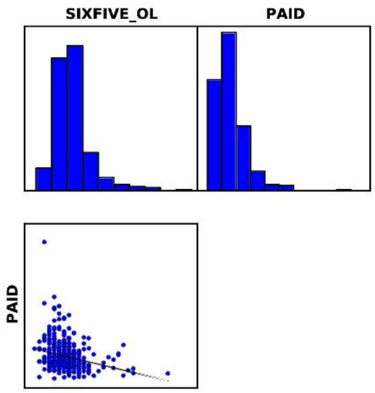
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

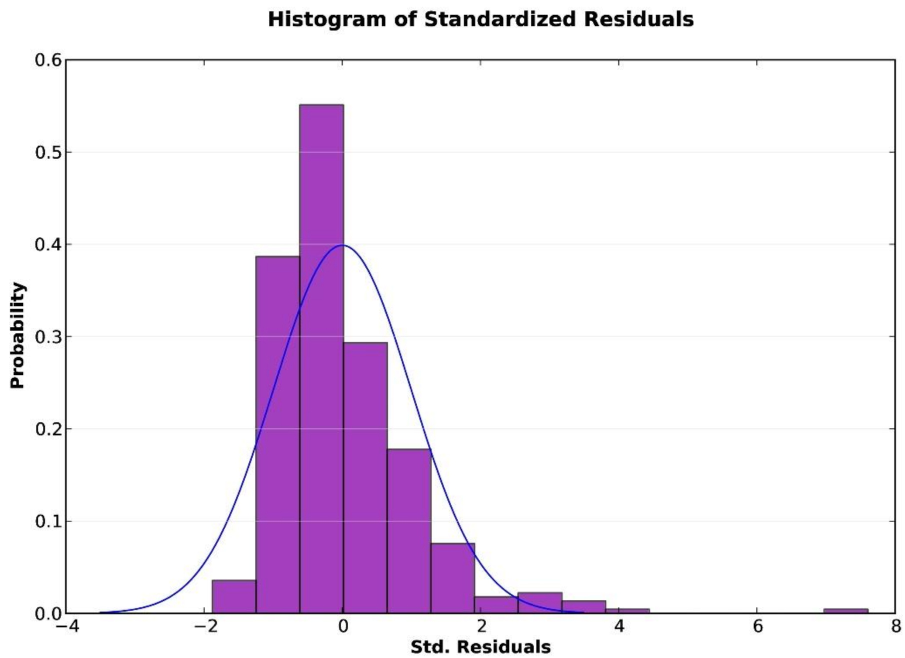
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships



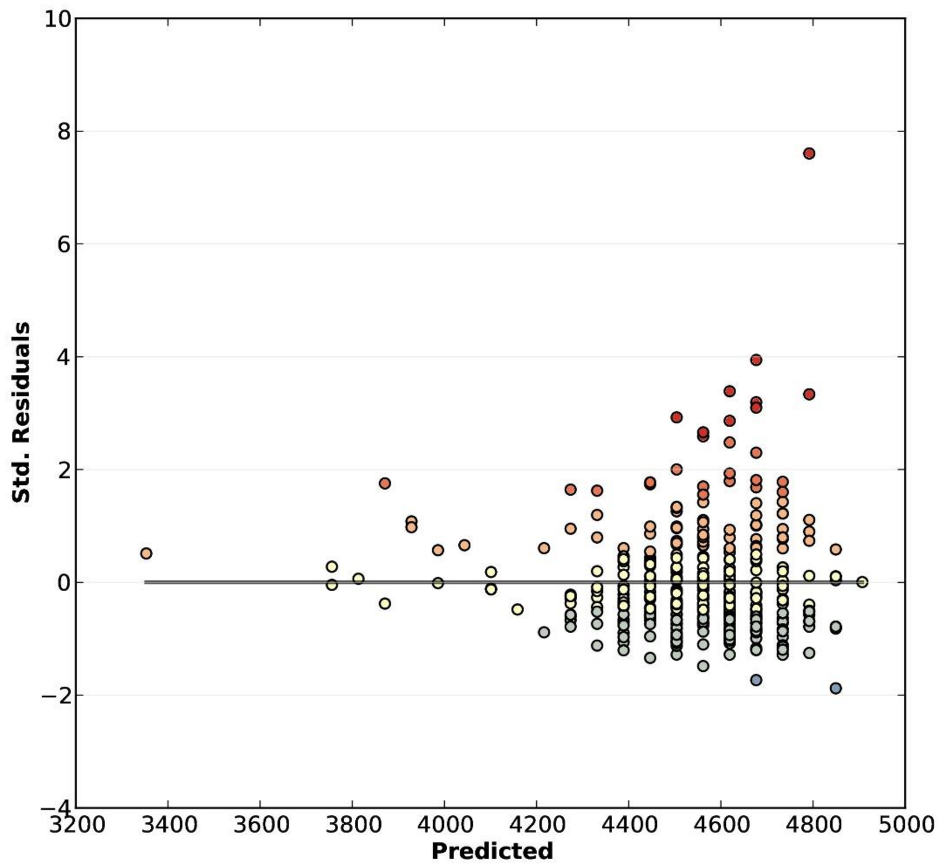
The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.

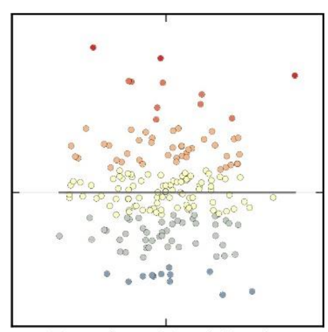


Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).

Residual vs. Predicted Plot



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	PAID
Explanatory Variables	SIXFIVE_OL
Selection Set	False

Report 12: OLS Report for Average Medicare Payments for Heart Failure and Percent Aged Sixty-Five and Older

Difference (Average Hospital Charges – Average Medicare Payment) Regression Reports

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	11785.734806	585.579841	20.126606	0.000000*	552.249401	21.341327	0.000000*
AF_AMER	-24.098913	36.712738	-0.656418	0.511980	28.210415	-0.854256	0.393528

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	DIFFERENCE
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7365.121895
Multiple R-Squared [d]:	0.001216	Adjusted R-Squared [d]:	-0.001606
Joint F-Statistic [e]:	0.430885	Prob(>F), (1,354) degrees of freedom:	0.511981
Joint Wald Statistic [e]:	0.729753	Prob(>chi-squared), (1) degrees of freedom:	0.392963
Koenker (BP) Statistic [f]:	0.376475	Prob(>chi-squared), (1) degrees of freedom:	0.539496
Jarque-Bera Statistic [g]:	2024.144240	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

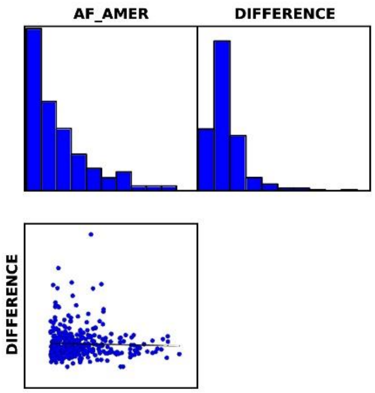
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

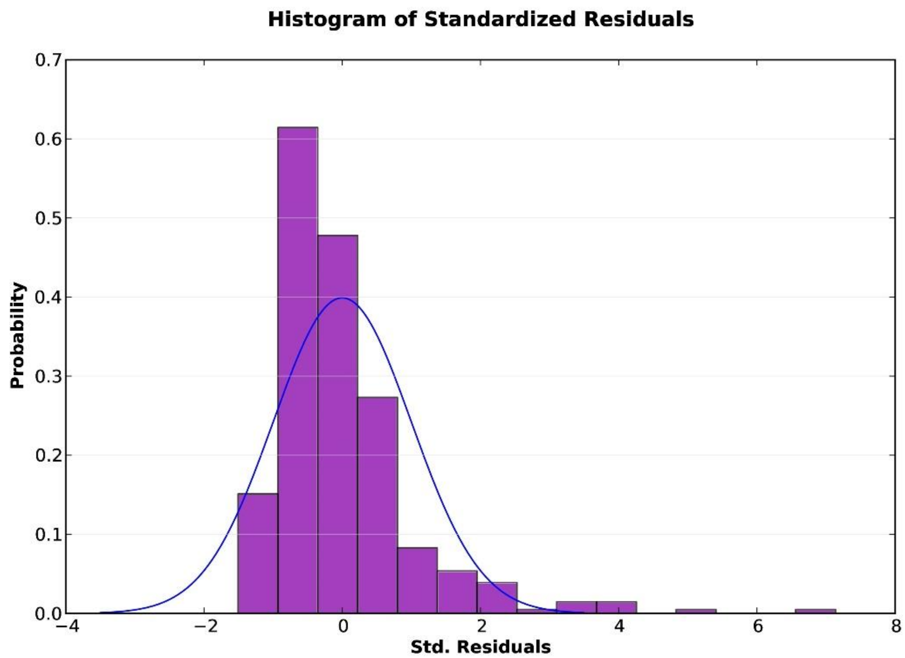
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

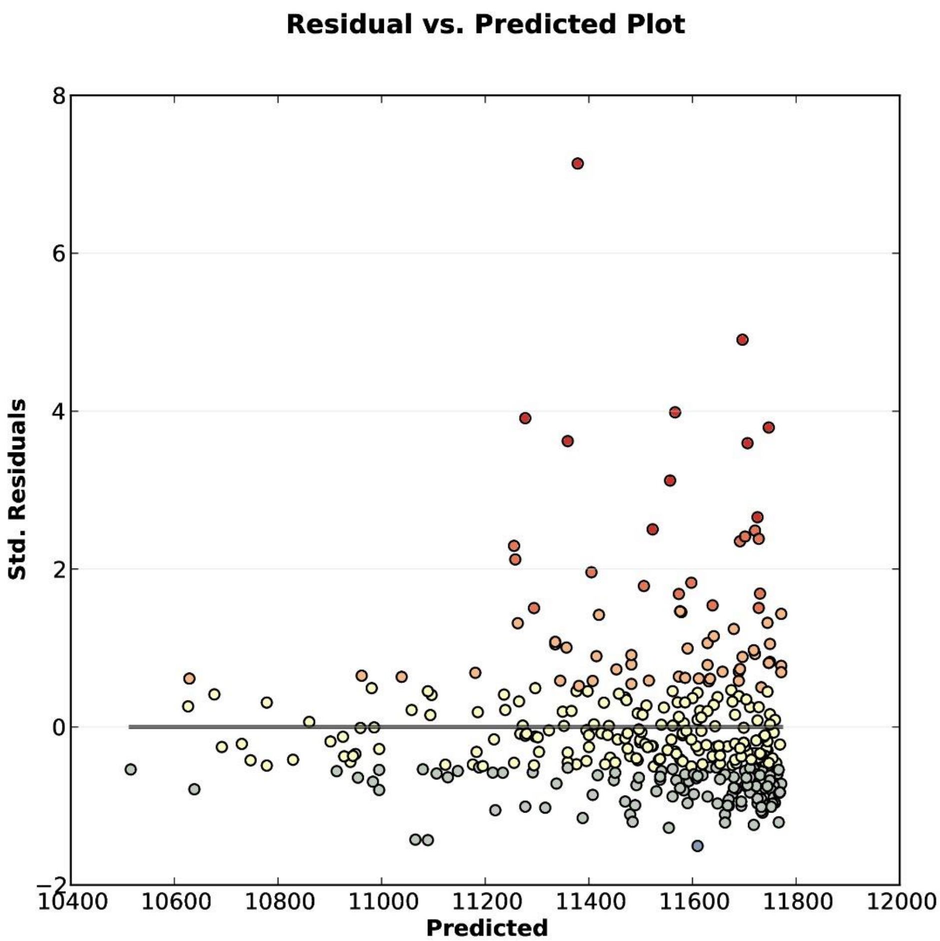


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

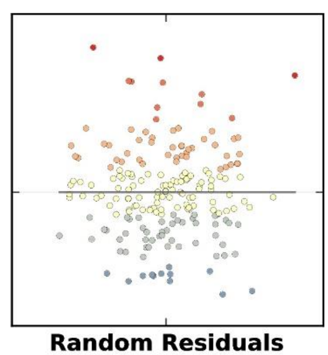
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	DIFFERENCE
Explanatory Variables	AF_AMER
Selection Set	False

Report 13: OLS Report for the Difference between Average Hospital Charges and Average Medicare Payments for Heart Failure (Charge – Payment) and Percent African American

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	10955.352896	1351.524719	8.105921	0.000000*	1474.067073	7.432059	0.000000*
BA_HIGHER	21.436980	50.599917	0.423656	0.672087	57.180663	0.374899	0.707973

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	DIFFERENCE
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7365.374498
Multiple R-Squared [d]:	0.000507	Adjusted R-Squared [d]:	-0.002317
Joint F-Statistic [e]:	0.179485	Prob(>F), (1,354) degrees of freedom:	0.672074
Joint Wald Statistic [e]:	0.140549	Prob(>chi-squared), (1) degrees of freedom:	0.707736
Koenker (BP) Statistic [f]:	0.365939	Prob(>chi-squared), (1) degrees of freedom:	0.545226
Jarque-Bera Statistic [g]:	2013.849471	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

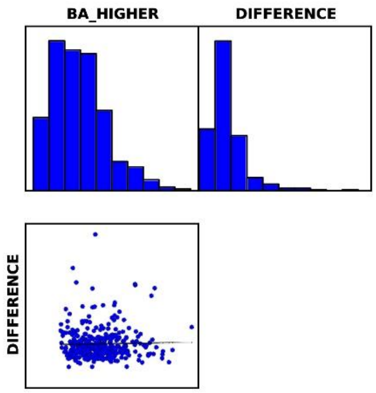
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

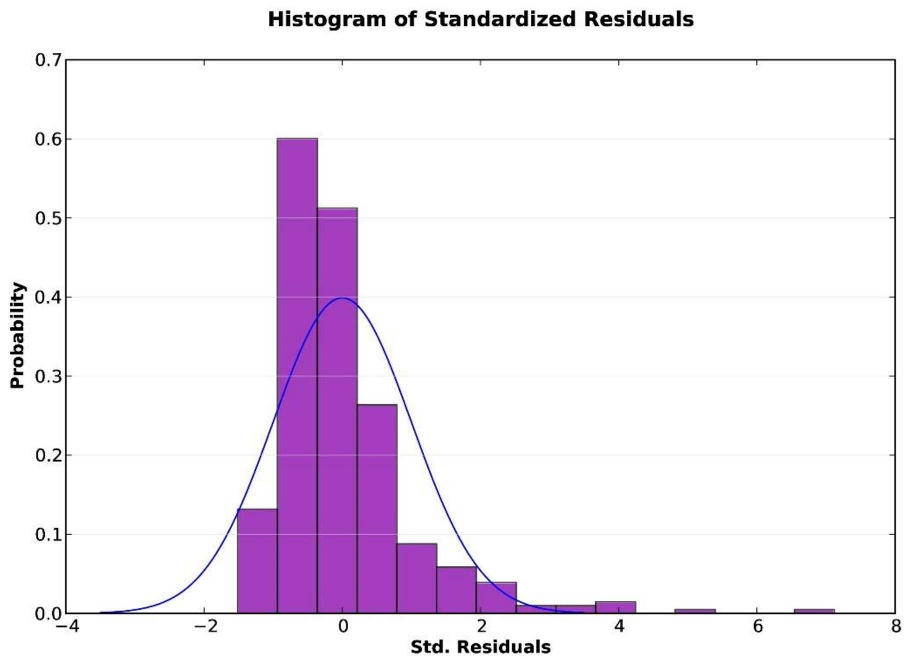
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

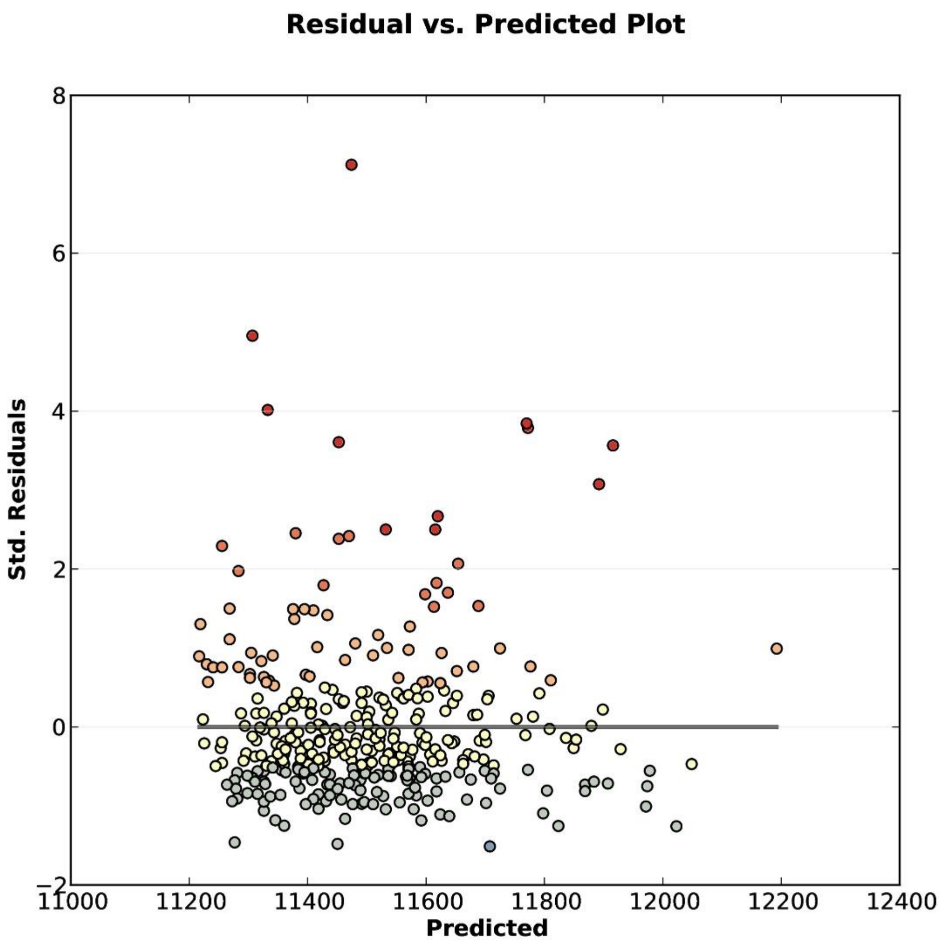


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

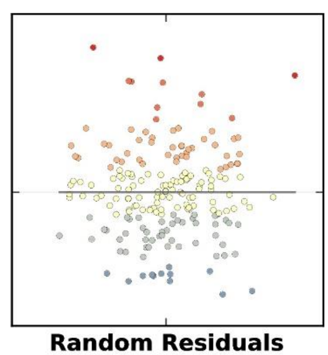
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	DIFFERENCE
Explanatory Variables	BA_HIGHER
Selection Set	False

Report 14: OLS Report for the Difference between Average Hospital Charges and Average Medicare Payments for Heart Failure (Charge – Payment) and Percent with a Bachelor's Degree or Higher

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	13458.762057	967.632762	13.908957	0.000000*	1124.322295	11.970555	0.000000*
DISCHARGES	-50.286280	22.722884	-2.213024	0.027520*	23.385572	-2.150312	0.032195*

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	DIFFERENCE
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7360.663566
Multiple R-Squared [d]:	0.013646	Adjusted R-Squared [d]:	0.010860
Joint F-Statistic [e]:	4.897474	Prob(>F), (1,354) degrees of freedom:	0.027533*
Joint Wald Statistic [e]:	4.623843	Prob(>chi-squared), (1) degrees of freedom:	0.031531*
Koenker (BP) Statistic [f]:	4.995426	Prob(>chi-squared), (1) degrees of freedom:	0.025414*
Jarque-Bera Statistic [g]:	1805.309245	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

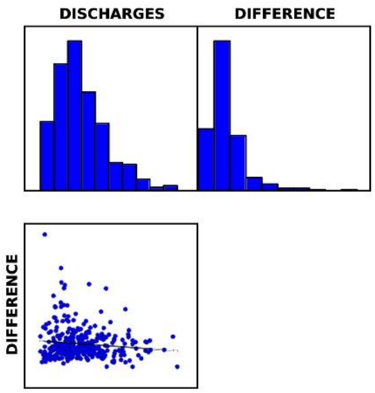
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

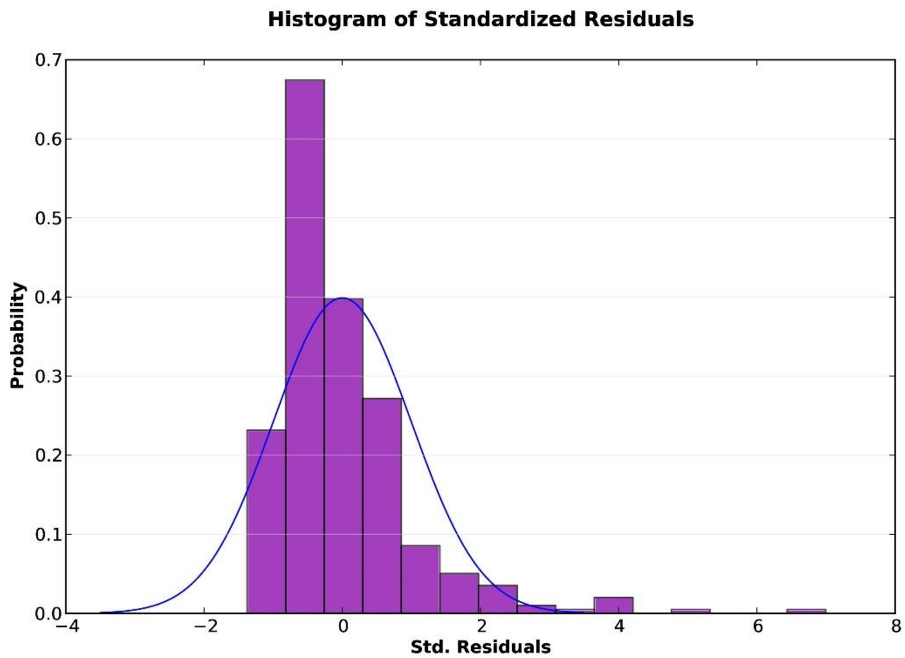
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

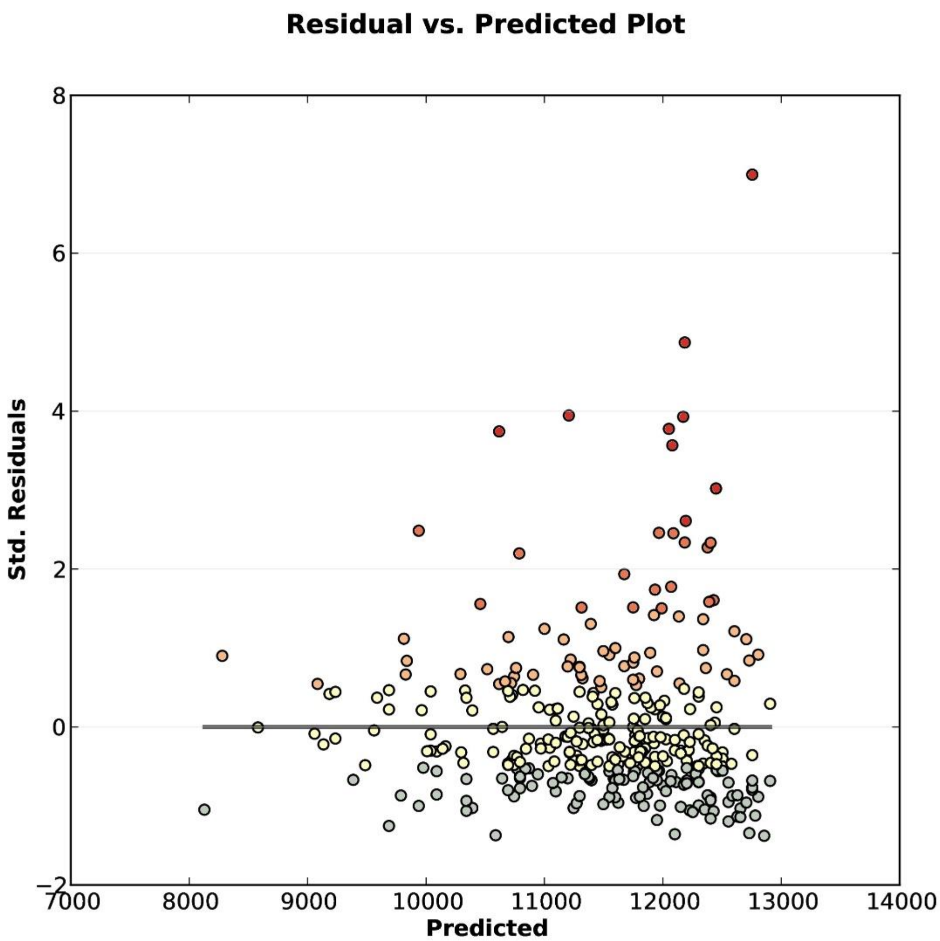


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

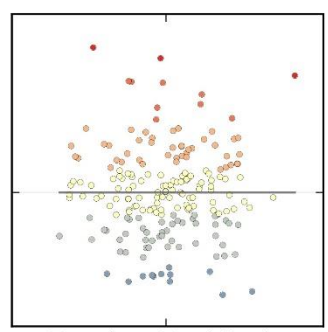
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	DIFFERENCE
Explanatory Variables	DISCHARGES
Selection Set	False

Report 15: OLS Report for the Difference between Average Hospital Charges and Average Medicare Payments for Heart Failure (Charge – Payment) and Average Number of Hospital Discharges for Heart Failure

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	59754.958496	20069.681347	2.977375	0.003115*	19813.833743	3.015820	0.002757*
FEMALE	-950.472074	395.256669	-2.404696	0.016686*	387.609712	-2.452137	0.014672*

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	DIFFERENCE
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7359.786704
Multiple R-Squared [d]:	0.016072	Adjusted R-Squared [d]:	0.013293
Joint F-Statistic [e]:	5.782562	Prob(>F), (1,354) degrees of freedom:	0.016699*
Joint Wald Statistic [e]:	6.012975	Prob(>chi-squared), (1) degrees of freedom:	0.014201*
Koenker (BP) Statistic [f]:	3.213726	Prob(>chi-squared), (1) degrees of freedom:	0.073023
Jarque-Bera Statistic [g]:	1938.819211	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

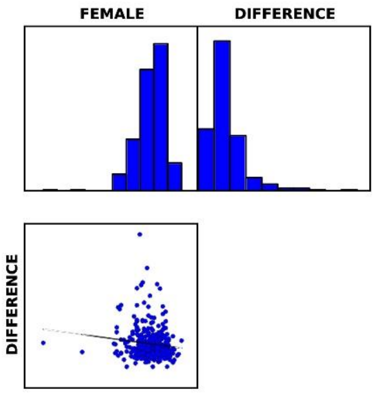
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

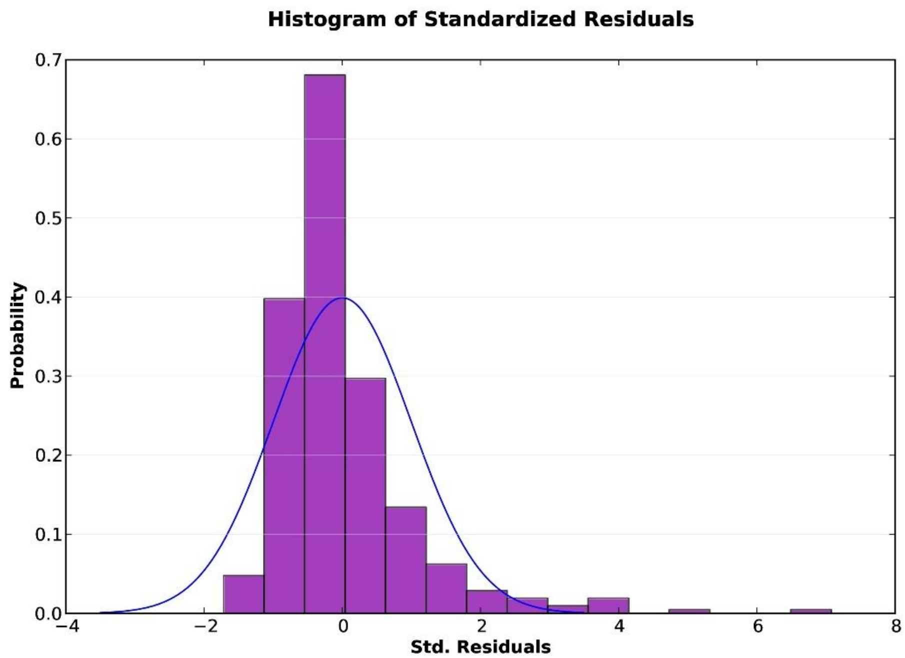
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

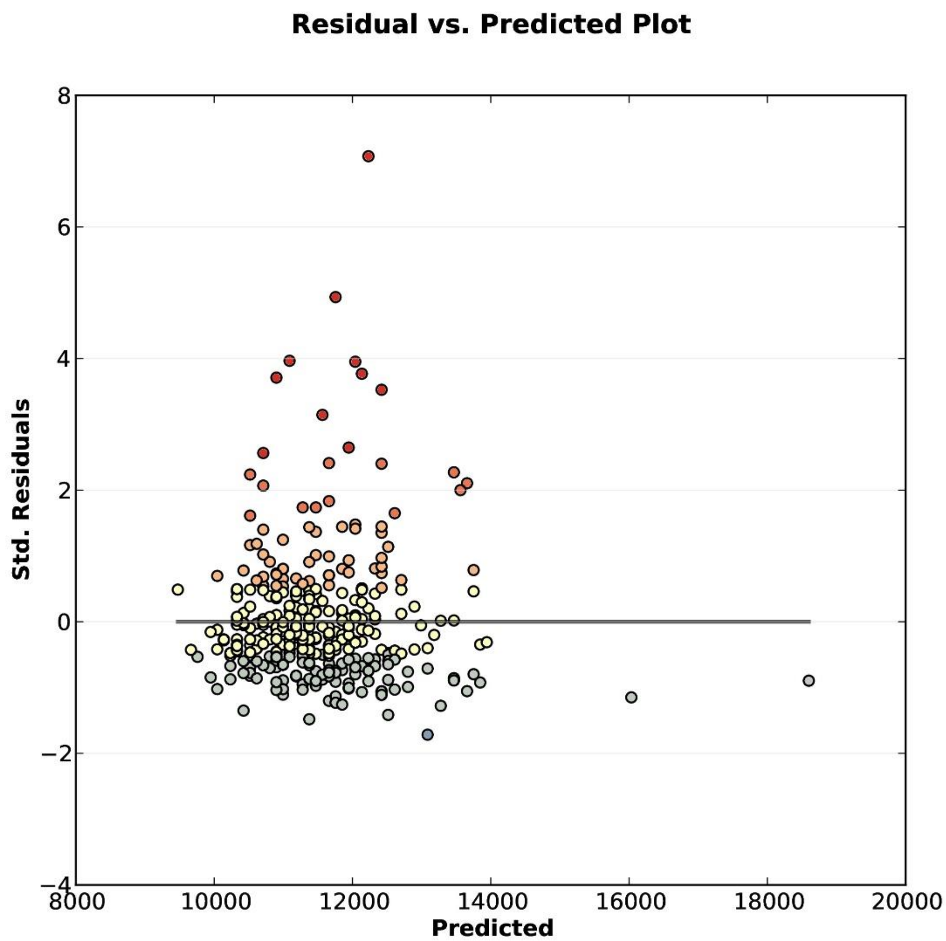


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

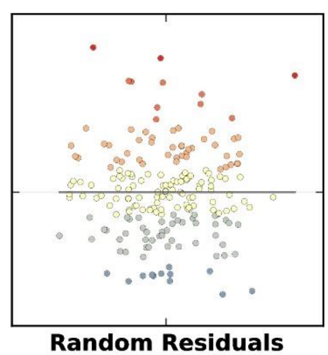
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	DIFFERENCE
Explanatory Variables	FEMALE
Selection Set	False

Report 16: OLS Report for the Difference between Average Hospital Charges and Average Medicare Payments for Heart Failure (Charge – Payment) and Percent Female

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	10489.518574	1559.343526	6.726881	0.000000*	1702.098925	6.162696	0.000000*
POVERTY	66.685749	99.258785	0.671837	0.502123	102.284866	0.651961	0.514848

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	DIFFERENCE
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7365.101325
Multiple R-Squared [d]:	0.001273	Adjusted R-Squared [d]:	-0.001548
Joint F-Statistic [e]:	0.451365	Prob(>F), (1,354) degrees of freedom:	0.502126
Joint Wald Statistic [e]:	0.425053	Prob(>chi-squared), (1) degrees of freedom:	0.514426
Koenker (BP) Statistic [f]:	2.102227	Prob(>chi-squared), (1) degrees of freedom:	0.147085
Jarque-Bera Statistic [g]:	2101.902442	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value (p < 0.05).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant (p < 0.05); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

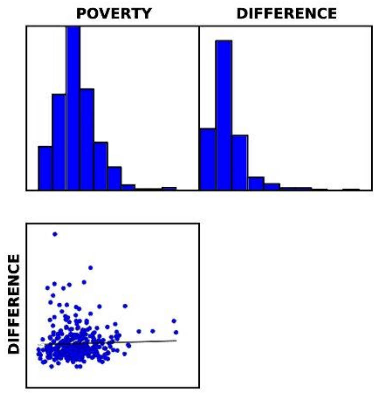
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance (p < 0.05); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant (p < 0.05), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

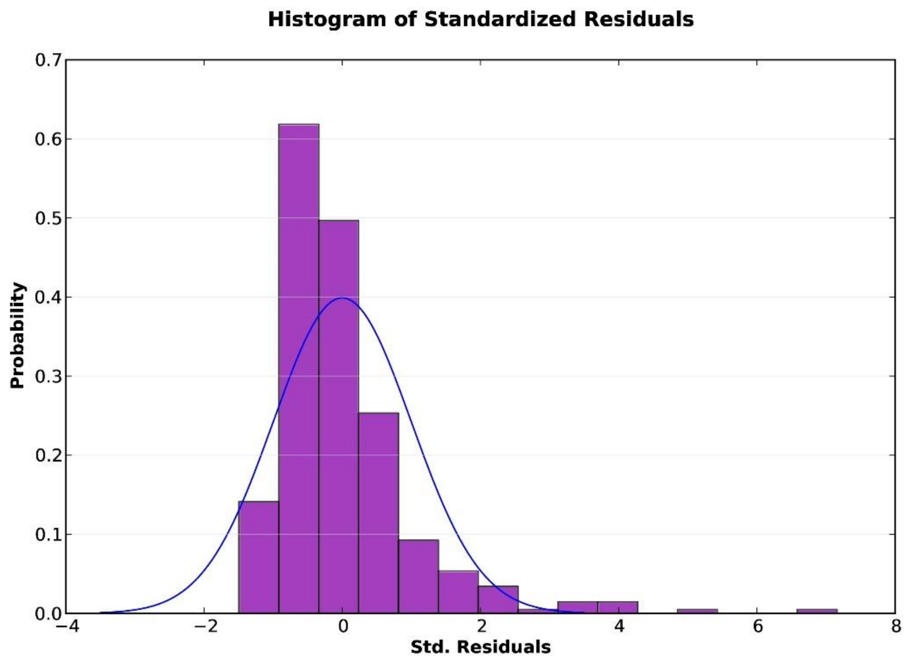
[g] Jarque-Bera Statistic: When this test is statistically significant (p < 0.05) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

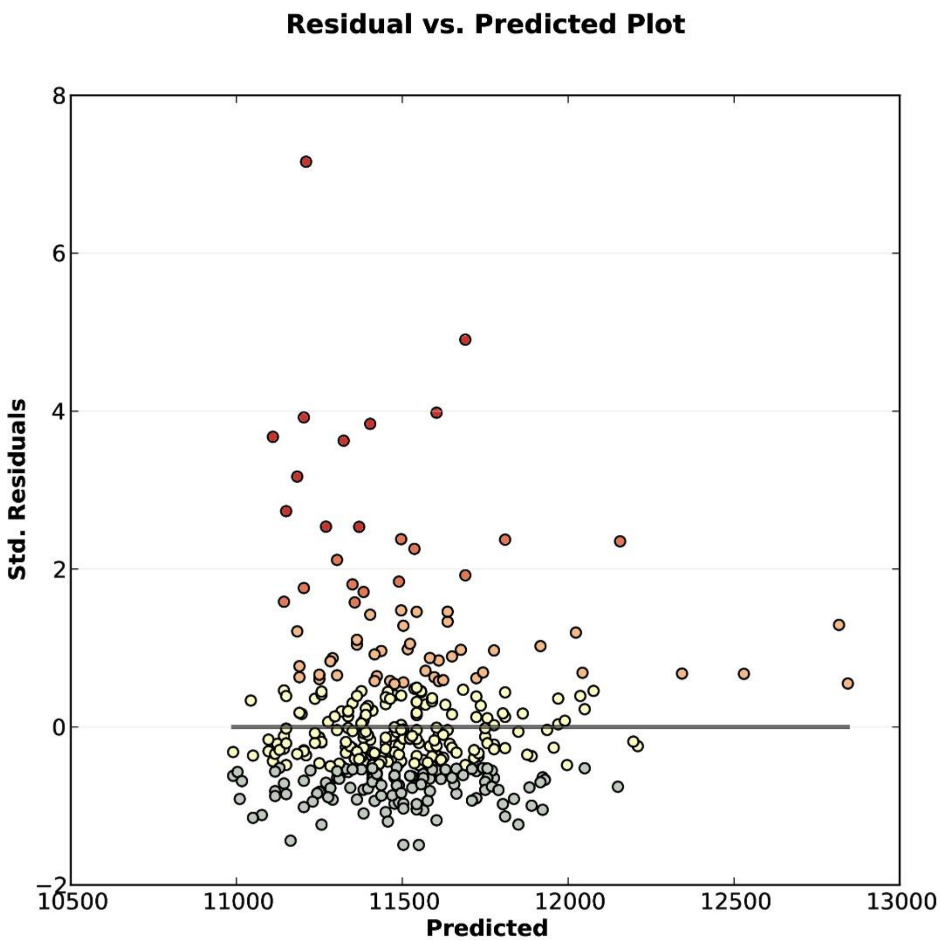


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

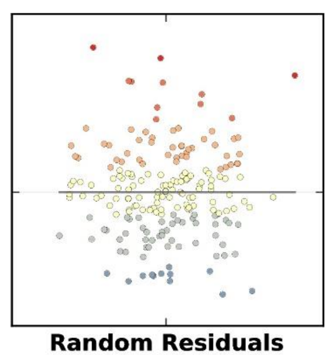
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	DIFFERENCE
Explanatory Variables	POVERTY
Selection Set	False

Report 17: OLS Report for the Difference between Average Hospital Charges and Average Medicare Payments for Heart Failure (Charge – Payment) and Percent in Poverty

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	14722.168861	1651.394195	8.914994	0.000000*	1668.533653	8.823417	0.000000*
SIXFIVE_OL	-240.225031	119.658869	-2.007582	0.045440*	113.612973	-2.114415	0.035167*

OLS Diagnostics

Input Features:	no_ak_hi	Dependent Variable:	DIFFERENCE
Number of Observations:	356	Akaike's Information Criterion (AICc) [d]:	7361.524693
Multiple R-Squared [d]:	0.011257	Adjusted R-Squared [d]:	0.008464
Joint F-Statistic [e]:	4.030387	Prob(>F), (1,354) degrees of freedom:	0.045447*
Joint Wald Statistic [e]:	4.470753	Prob(>chi-squared), (1) degrees of freedom:	0.034480*
Koenker (BP) Statistic [f]:	1.796886	Prob(>chi-squared), (1) degrees of freedom:	0.180089
Jarque-Bera Statistic [g]:	1896.357427	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

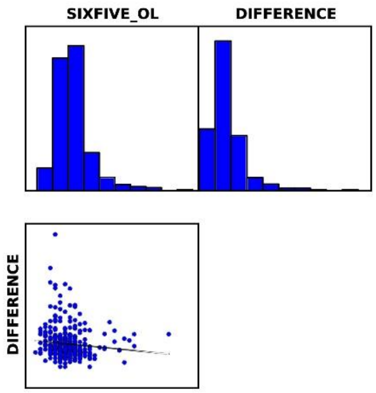
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

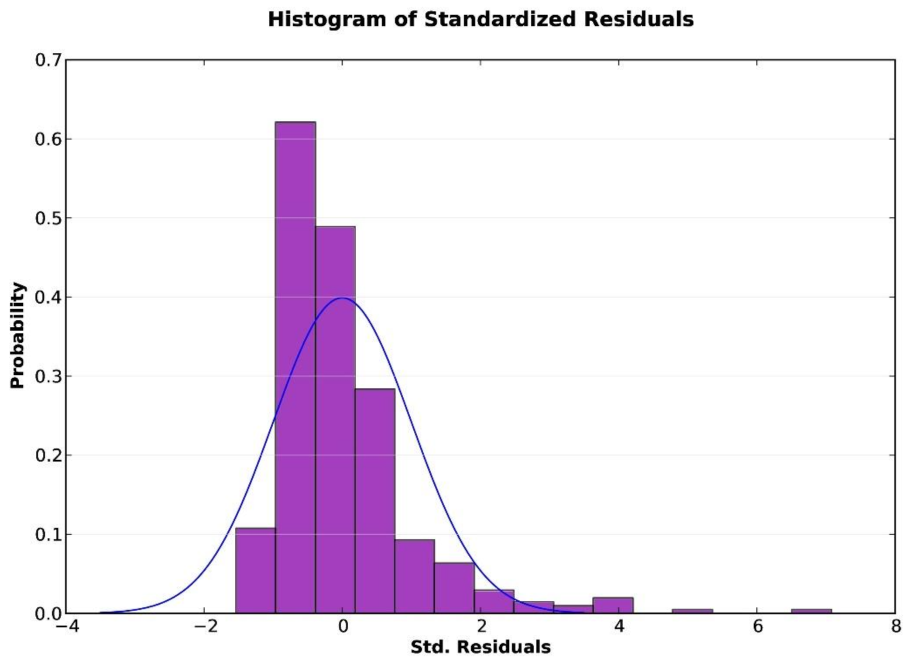
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships

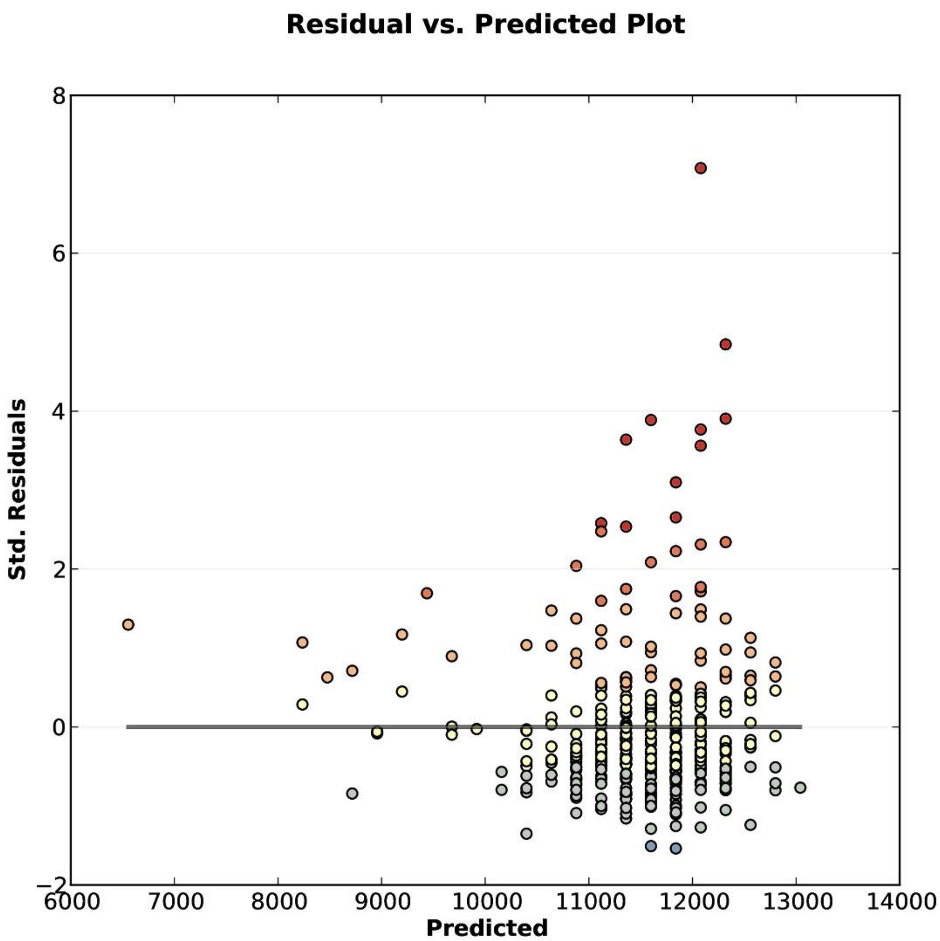


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

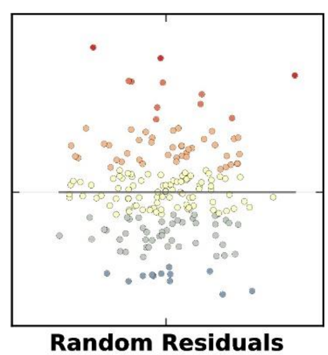
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	no_ak_hi
Unique ID Field	UniqueID
Output Feature Class	None
Dependent Variable	DIFFERENCE
Explanatory Variables	SIXFIVE_OL
Selection Set	False

Report 18: OLS Report for the Difference between Average Hospital Charges and Average Medicare Payments for Heart Failure (Charge – Payment) and Percent Aged Sixty-Five and Older

References

Association of American Medical Colleges. 2007. "Geographic Differences in Health Care Spending." 21 Sept. 2013. <https://www.aamc.org/linkableblob/70266-5/data/geographicdifferences-data.pdf>.

Centers for Disease Control. 2010 “Division for Heart Disease and Stroke Prevention: Data Trends & Maps." *CDC-DHDSP Data Trends & Maps - Location Summary*. 10 Sept. 2013. <http://apps.nccd.cdc.gov/NCVDSS_DTM/LocationSummary.aspx?state=United+States>.

Centers for Medicare and Medicaid Services. "Medicare Provider Charge Data: Inpatient." N.p., June 2013. Web. 12 Aug. 2013. <http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient.html>.

Centers for Medicare and Medicaid Services. "Medicare Physician Fee Schedule." N.p., Apr. 2013. Web. Oct. 2013. <http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/MedcrePhysFeeSchedfctsht.pdf>.

Chan, Leighton, Gary Hart, and David Goodman. "Geographic Access to Health Care for Rural Medicare Beneficiaries." N.p., Apr. 2005. Web. 8 Apr. 2014. <http://depts.washington.edu/uwrhrc/uploads/RHRC_WP97_Dec-27-10.PDF>.

Cutler, David. "Education and Health." N.p., Mar. 2207. Web. 15 Apr. 2014. <http://www.npc.umich.edu/publications/policy_briefs/brief9/>.

Cutler, David, and Louise Sheiner. 1999 “The Geography of Medicare." *The American Economic Review* 89.2: 228-233.

ESRI. "U.S State Boundaries." *ESRI Online Resource*. N.p., 19 July 2011. Web. 27 Dec. 2014. <http://www.arcgis.com/home/item.html?id=540003aa59b047d7a1f465f7b1df1950>.

Fisher, Elliot, Julie Bynum, and Jonathan Skinner. "Slowing the Growth of Health Care Costs - Lessons from Regional Variation — NEJM." *Slowing the Growth of Health Care Costs - Lessons from Regional Variation — NEJM*. N.p., 26 Feb. 2009. Web. 15 Apr. 2014. <http://www.nejm.org/doi/full/10.1056/NEJMp0809794>.

Fisher, Elliot, and Jonathan Skinner. "Making Sense Of Geographic Variations In Health Care: The New IOM Report." *Health Affairs Blog Making Sense Of Geographic Variations In Health Care The New IOM Report Comments*. N.p., 24 July 2013. Web. 15 Apr. 2014. <http://healthaffairs.org/blog/2013/07/24/making-sense-of-geographic-variations-in-health-care-the-new-iom-report/>.

Fuchs, Victor, Mark McClellan, and Jonathan Skinner. 2001. "Area Differences In Utilization Of Medical Care And Mortality Among U.S. Elderly." *National Bureau of Economic Research*

Gage, Barbara, Marilyn Moon, and Sang Chi. 1999. "State-Level Variation in Medicare Spending." *Health Care Financing Review* 21.2: 85-98.

Himmelstein, David, Deborah Thorne, Elizabeth Warren, and Steffie Woolhandler. 2009. "Medical Bankruptcy in the United States, 2007: Results of a National Study." *The American Journal of Medicine* 122.8: 741-746.

Institute of Medicine. 2013. "US Health care Costs - Where Is the Money Going?" *US Health care Costs - Where Is the Money Going?* Institute of Medicine of the National Academies

Kane, Jason. 22 October 2012 "Health Costs: How the U.S. Compares With Other Countries." *PBS*. 13 July 2013. < http://www.pbs.org/newshour/rundown/2012/10/health-costs-how-the-us-compares-with-other-countries.html>

Medicare Payment Advisory Commission. "Report to the Congress: Medicare Payment Policy." 15 Mar. 2013. Web. <http://www.medpac.gov/documents/Mar13_entirereport.pdf>.

Missouri Department of Economic Development. "Cost of Living Annual Average 2013."*Cost of Living Annual Average 2013*. N.p., 2013. Web. 10 Apr. 2014. <http://www.missourieconomy.org/indicators/cost_of_living/index.stm>.

NBC News. "Interactive Map: Health Care Costs Vary Widely across U.S." N.p., 2014. Web. 20 Mar. 2014. <http://www.nbcnews.com/id/31390679/ns/health-health_care/t/interactive-map-health-care-costs-vary-widely-across-us/#.UzVKDfldV8E>.

Newhouse, Joseph, and Alan Garber. "Geographic Access in Medicare Services." N.p., 18 Apr. 2013. Web. 8 Apr. 2014. <http://www.nejm.org/doi/pdf/10.1056/NEJMp1302981>.

Radnofsky, Louise. "Health-Care Costs: A State-by-State Comparison." *The Wall Street Journal*. Dow Jones & Company, 08 Apr. 2013. Web. 05 Feb. 2014. <http://online.wsj.com/news/interactive/IVCostsprint?ref=SB10001424127887323884304578328173966380066>.

Roubideaux, Yvette. "A Review of the Quality of Health Care for American Indians and Alaska." N.p., Sept. 2004. Web. 8 Apr. 2014. <http://www.commonwealthfund.org/usr_doc/roubideaux_qualityhltcare_aians_756.pdf>

Shook, John. 17 November 2009. "The U.S. Versus the World Healthcare Cost Gap." *The U.S. Versus the World Healthcare Cost Gap*. 23 Sept. 2013. <http://www.lean.org/shook/displayobject.cfm?o=1274>.

Tera Insights. 2013. "Exploring Medicare DRG Costs." *Tera Insights*. 24 Sept. 2013. <http://www.terainsights.com/index.php/casestudy/view/2>.

The Center for Evaluative Clinical Sciences. 1998. *The Dartmouth Atlas of Health Care 1998*. Chicago, IL: American Hospital Pub.

The Center for Evaluative Clinical Sciences. 1999. *The Quality of Medical Care in the United States: A Report on the Medicare Program.* Chicago, IL: American Hospital Pub.

The Center for Evaluative Clinical Sciences. 1999. *Appendix on the Geography of Health Care in the United States.* Chicago, IL: American Hospital Pub.

United States Census Bureau. 2013. "Metropolitan and Micropolitan." *Statistical Areas Main*. 24 Sept. 2013. <http://www.census.gov/population/metro/>.

United States Census Bureau. "Growth in Urban Population Outpaces Rest of Nation, Census Bureau Reports." *Growth in Urban Population Outpaces Rest of Nation, Census Bureau*

Reports. N.p., 26 Mar. 2012. Web. 30 Oct. 2014. <http://www.census.gov/newsroom/releases/archives/2010_census/cb12-50.html>.

University of California - Los Angeles. 2013 "GIS at UCLA." *GIS at UCLA*. Web. 14 Nov. 2013. <http://gis.ats.ucla.edu/>.

Welch, W. Pete, Mark E. Miller, H. Gilbert Welch, Elliott S. Fisher, and John E. Wennberg. 1993. "Geographic Variation in Expenditures for Physicians' Services in the United States. *"New England Journal of Medicine* 328.9: 621-627.