CHARACTERIZATION OF HIGHLY EFFICIENT DELIVERY MODELS FOR EYE CARE SERVICES IN VETERAN AFFAIRS

by

Sruthi Kilambi

A master thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Engineering Management

Charlotte

2021

Approved by:

_____
Dr. Ertunga C. Ozelkan

_____
Dr. Charles N. Davis

_____
Dr. Tao Hong

ABSTRACT

Sruthi Kilambi, Characterization of Highly Efficient Delivery Models for Eyecare Services in Veterans Affairs. (Under the direction of DR. ERTUNGA C. OZELKAN)

Veterans Affairs (VA) provides eye care and vision treatment to all eligible veterans to help improve eye and vision health. Over the past years, timely access to eye care has become a challenge with the increase in its demand. Therefore, it is important to improve the efficiency of the processes by identifying and implementing best practices. The main purpose of this study is to identify critical factors to improve VA's eye care and vision clinical performance. For this research, VA aggregate data from FY 2017-2019 are collected for Optometry and Ophthalmology Services for multiple medical centers in different regions in the US and analyzed through statistical models including Multiple Linear Regression, Stepwise Regression, and Principal Component Analysis (PCA). 26 different model variations were compared in an experimental setting to see the robustness and sensitivity of the findings. From all the tested models, Multiple Linear Regression using showed better performance with R-squared values of 0.294, 0.322 and adjusted R-squared 0.264, 0.292 for Optometry and Ophthalmology, respectively. The results indicate that input factors such as "Physician Clinical FTEE", "Clinical Support Staff per Physician Clinical FTEE", "Physician Clinical FTEE per 10KSpecialty Unique" "Adj MDFTEE", "Highest complexity (complex_1a)" and "low complexity (complex_3)" are the most critical factors for Optometry performance. On the other hand, "Residents", "Physician Clinical FTEE", "Physician Clinical FTEE per 10KSpecialty Unique", "Clinical Support Staff per Physician Clinical FTEE" and "seem to be the most critical factors for Ophthalmology performance.

ACKNOWLEDGEMENTS

I wish to express special gratitude to my advisor Dr. Ertunga C. Ozelkan as well as Dr. Charles N. Davis for the advice, support, and guidance on the organization of this research. Special thanks to Dr. Charles N. Davis for his time and availability for providing the research data. I would like to thank Dr. Hong for serving on my committee.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DO	Doctor of Osteopathy
DSS	Decision Support System
DVA	Department of Veterans Affairs
FTE	Full-Time Employee
FY	Fiscal Year
MCG	Medical Center complexity Group
OD	Doctor of Optometry
PPMC	Pearson Product Moment Correlation
RVU	Relative Value Unit
SCS	Specialty Care Services
VAMC	Veteran Affairs Medical Center
VHA	Veterans' Health Administration
VIF	Variance Inflation Factor
VISN	Veterans Integrated Service Networks
VSSC	VHA Support Service Center

CHAPTER1: INTRODUCTION

Congress passed the Mission Act in 2016, which gives veterans greater access to health care in Veteran Affairs (VA) facilities as well as in the outside health community (MISSION Act Strengthens VA Care. 2020, March 17). The Mission Act expands the benefits for caregivers and improves VA's ability to recruit and retain the best medical providers. Since its adoption, the VA has been taking administrative steps to meet the intent of the act. Accordingly, the focus has been on the clinical efficiency and productivity, of the healthcare delivery models, that play a central role in determining how the Mission Act should be implemented.

Clinical decisions must be made regarding which services go out to the health community and which services should stay in-house within VA. It is important that the VA systematically identifies best practices to address low productivity and inefficiency. Delivery models characterized by high costs and "underproduction" must be identified and improved if the VA is to continue throughout the next decade. In this research we investigate the performance of eye-care services, more specifically we analyze optometry and ophthalmology services at various VA medical centers (VAMCs).

VA's eye care is the third busiest service in the Veterans' Healthcare Administration (VHA), behind primary care and mental health. For the Fiscal Year (FY) 2018, over 1.81 million services were recorded in 381 VA medical sites located in urban, rural, and highly rural areas (Petersen, H. 2019). The demand for eye care has grown 24% nationally over the `past 5 years and some local medical facilities have grown by as much as 40%. (Lynch, M.G 2017). The VHA National Eye Care Program is the combination of the national divisions of optometry and ophthalmology and is jointly led by the national directors of optometry and ophthalmology within

Specialty Care Services (SCS). It supports the delivery of eye and vision care services throughout VHA.

For administration purposes, the Veterans Health Administration is divided into areas called Veterans Integrated Service Networks (VISN). Each VISN has a group of facilities in them. For FY 2019, there are 18 VISNs and 140 facilities dealing with optometry and 18 VISNs, and 125 facilities dealing with ophthalmology (VHA Handbook, 2011).

Veterans receive eye care at 134 locations across the nation. These clinics are highly variable in staffing models, the complexity of services, performance measures, student training programs, customer satisfaction, and numbers of Veterans served. As an example, a closer look at the Salisbury VAMC optometry and ophthalmology services show that these represent high-volume clinics. The Salisbury VAMC optometry service provides approximately 45,000 clinical encounters for 32,000 veterans annually. Recently, the Salisbury VAMC optometry service has been mandated to directly manage their service line budget, requiring decisions to be made regarding clinic staffing, measures of productivity, resource utilization, and the enactment of community care. At present, there is little information regarding what constitutes “best practice” models for the delivery of eye care within the VA. Metrics such as patient care, per capita cost of health care, clinical and staff burnout, mortality rate, and life expectancy are common metrics considered while evaluating the performance of health care. It is unclear what combinations of these metrics should be used within the Salisbury VAMC Optometry Service to maximize efficiency (VHA Directive, 2019). Therefore, to understand the characteristics of the "best model" for eye care delivery systems in VA, the following research questions are answered in this study.

1. What are the most influential factors that affect eye care clinical productivity for Optometry and Ophthalmology services in VAMCs?

2. How do top-performing VAMCs compare to the low-performing VAMCs concerning the influential factors for Optometry and Ophthalmology services?
3. How do the influential factors change and compare for Optometry and Ophthalmology services?

The answers to the above questions are valuable for the VAMC eye care service program. The goal is to understand the characteristics of the "best model" for eye care delivery systems and to use this information for any VAMC eye care service program planning.

1.1Summary of Contributions

The main contributions of this research can be summarized as follows:

- **Contribution 1:** Analysis of influential factors that affect operational performance in VAMC Optometry Services.
- **Contribution 2:** Analysis of influential factors that affect operational performance in VAMC Ophthalmology Services.
- **Contribution 3:** Comparison of influential factors on productivity for VAMC Optometry and Ophthalmology services.

Justification: As we will further discuss under the literature review section, there are no studies that study the performance of VAMCs' Optometry and Ophthalmology services.

The rest of the thesis is organized as follows: In Chapter 2, we will provide a literature review, In Chapter 3 we will introduce the methodology and present data and analysis for VA's Optometry services in Chapter 4. Chapter 5 has the data analysis and modeling for VA's Ophthalmology services. Chapter 6 includes a summary and major conclusion for this research.

CHAPTER 2: LITERATURE REVIEW

In this section, we will review the literature most directly related to the work presented here. Therefore, the literature review is divided into 1. Performance Measures in Healthcare with emphasis on Optometry and Ophthalmology and 2. Modeling and Analysis of Performance Measures in Healthcare with emphasis on Optometry and Ophthalmology.

2.1 Performance Measures in Healthcare with emphasis on Optometry and Ophthalmology

Efficiency is a crucial factor discussed in much of healthcare literature (Usherwood, 1987). It may be elementarily defined as the outcome (e.g., number of treated patients) over time. Efficiency is an uncertain measure of outcome, depending on the doctor's aim in treatment and output. This very much depends on the output measures that are thought to be suitable for the specific case. Therefore, the selection of the right measure is a very arduous task involving scrutiny and critical thinking. The belief that there is inefficiency and room for improvement for patient health and productivity is a measure of how productive a corresponding type of hospital is (Kao, C., 2020). Performance measures across the health care system are considered to determine efficiency or productivity. These performance measures vary internationally but are widely classified as system-wide, by disease, and by sub-sector. Literature is available to show how the Nordic countries' health system has been analyzed for its productivity or efficiency (Medin, et. al 2013). This was initiated by comparing the hospital cost efficiency across different Nordic counties. Medin, et al. (2013) presents an example in Finland, where a reform introduced by the policymakers has shown an increase in productivity. The reforms in Finland inspired health care reforms in other countries such as Norway.

There were quite a few papers (Kao, 2020, Usherwood, 1987), which discussed the productivity models in health care. Coleman (2003) discusses a productivity model built for the

VA Connecticut Healthcare System, which includes both clinical work and academic activities. It helps in allocations and optimum use of resources and allows to compare with other (i.e. Non-VA) healthcare systems. The findings show that clinical work and academic activities are important contributors to productivity. In another study, Johnson (2008) studied the effect of resident involvement on the productivity of physicians at Rush University Medical Care, showing physician productivity was negatively affected (decreased by $164,000 in revenue per year) with the resident involvement. The study concluded that residents reduce physician productivity, increase practice cost, and leave physicians less time to spend on patient-related activities. This can be an important point for our research since we have residents as one of the variables. Since there are various specialty hospitals, and some departments may be more productive than others depending on the specialty. A study on Chinese hospitals identifies that "general hospitals composed of all departments", "general hospitals lacking only the Chinese medicine department", and "hospitals with pediatrics and obstetrics/gynecology departments" are the three most productive types (Kao, 2020). They also show that non-for-profit hospitals are more productive than for-profit hospitals. In a study conducted in Iran, Nobakht (2018) assesses nursing perceptions regarding performance indicators and their correlation with productivity within the emergency departments. Their results showed that the nurses were positively impacted in some areas and negatively in some others. For example, a positive impact was observed on the percentage of patients, whose condition was determined within 6 hours and the percentage of the remaining patients, whose condition was resolved within 12 hours. On the contrary, the performance of the nurses had a negative correlation with average triage duration in the studied hospitals (Nobakht, 2018). In our study, we consider both nurse practitioners and other clinical staff as well under "total support staff". To identify the factors that influence health care in-depth, individual and

focus group interviews were conducted in Iran with care providers, managers, policymakers, and payers (Mosadeghrad, 2014). The results of this study show that personal factors of the provider and the patient, and factors pertaining to the healthcare organization, healthcare system, and the broader environment affect healthcare service quality. In research by Letvak S & Buck R (2021), several factors were identified as were identified related to the decrease of work productivity of nurses, including age, total years worked as full-time, quality care provided, job stress score, having had a job injury, and/or a health problem. Yauheniya (2017) reported significant differences in productivity in international comparisons of practices between Italy and Germany. This study showed bed-size category, ownership status, and specialization are significantly related to differences in efficiency performance.

2.2 Modeling and Analysis of Performance Measures in Healthcare with emphasis on Optometry and Ophthalmology

Regression is a commonly utilized technique in the modeling and analysis of healthcare performance. We will review here some of the most related papers. Hao (1994) conducted a multiple linear regression (MLR) analysis to investigate efficiencies and productivity of acute care veteran hospitals and showed that hospitals with membership have higher capital productivity ratios (output/bed) but lower labor productivity ratios (outputs/physician and outputs/nurse) than nonmember hospitals. In their study, input variables like occupancy rate, hospital beds, inpatient and outpatient surgeries emergency room, and outpatient visits were considered to study clinical efficiency. The regression results showed that the relationships between relative efficiency and input variables and the occupancy rate are not significant. Data envelopment analysis was also presented to show how relatively inefficient hospitals can be identified in both membership and

non-membership hospitals, and ways to increase their efficiencies. Smith (1995) presented an MLR analysis using a stepwise regression with backward elimination and with forwarding selection step. They conducted an observational study on physician's productivity, age, race, medical record history, new to clinic and physician, disease type, gender, time of physician arrival and departure, tests record, number of visits records of a patient, and modeled physician productivity as a function of clinical and physician characteristics. The study showed that the variation in "time spent with the patient", is influenced more by the "individual physician" variables. Lynch (2017) used descriptive and regression analysis using both linear and curve-fitting modeling methods. They analyzed data on the number of providers, panel size, number of unique patient visits, number of total clinical visits, and number of support staff to understand the factors that impact eye care productivity. The results showed that ophthalmology and optometry technicians, residents, and nurses were found to have a positive impact on productivity. PAs, nurse practitioners, administrative, and clerical staff had no statistically significant impact on productivity. More specifically, the study has suggested that an ophthalmology technician can improve the productivity of an ophthalmologist by 23.11%.

A survey conducted by Budzi D et al (2010) showed Patients were more satisfied with services provided by NPs who possess certain characteristics acquired from the training.

Most of the outpatients surveyed preferred to see Nurse Practitioners than Physician Assistants and physicians for primary care. Multiple regression analysis was used in this study. This study may permit the administrators to employ more healthcare professionals such as NPs, which may contribute to cost-effective and quality healthcare services.

2.3 Literature Review Conclusions

The literature review showed that data analysis and modeling work in the area of performance of optometry and ophthalmology services is scarce. Narrowing down to VA our literature review shows that to our best knowledge, there is no such modeling and analysis. Therefore, we believe that the research presented here will help closing an important gap in the healthcare literature.

CHAPTER 3: METHODOLOGY AND DATA

In this section, we will present the methodology that is followed in this research and describe the data that were utilized. The methodology consisted of three major steps as summarized in Figure 3 below.

- In Step 1. Data Collection: Input and output variables were defined. After collecting data, the data cleaning process was carried out to deal with missing data entries.
- In Step 2. Descriptive Data Analysis, we performed graphical analysis and descriptive statistics to understand the basic trends of the variables. Also, the correlation was performed to check for multicollinearity.
- In Step 3. Data modeling, we performed different types of regression analysis including multivariate linear regression, stepwise regression as well as Principal Component Analysis to deal with collinearity.

The above steps were repeated for both optometry and ophthalmology services in VAMC. In the remainder of this chapter, we will present the details of each step, elaborating on the methodology and data further. The application of the methodology to optometry and ophthalmology services will be presented separately in Chapters 4 and 5, respectively.

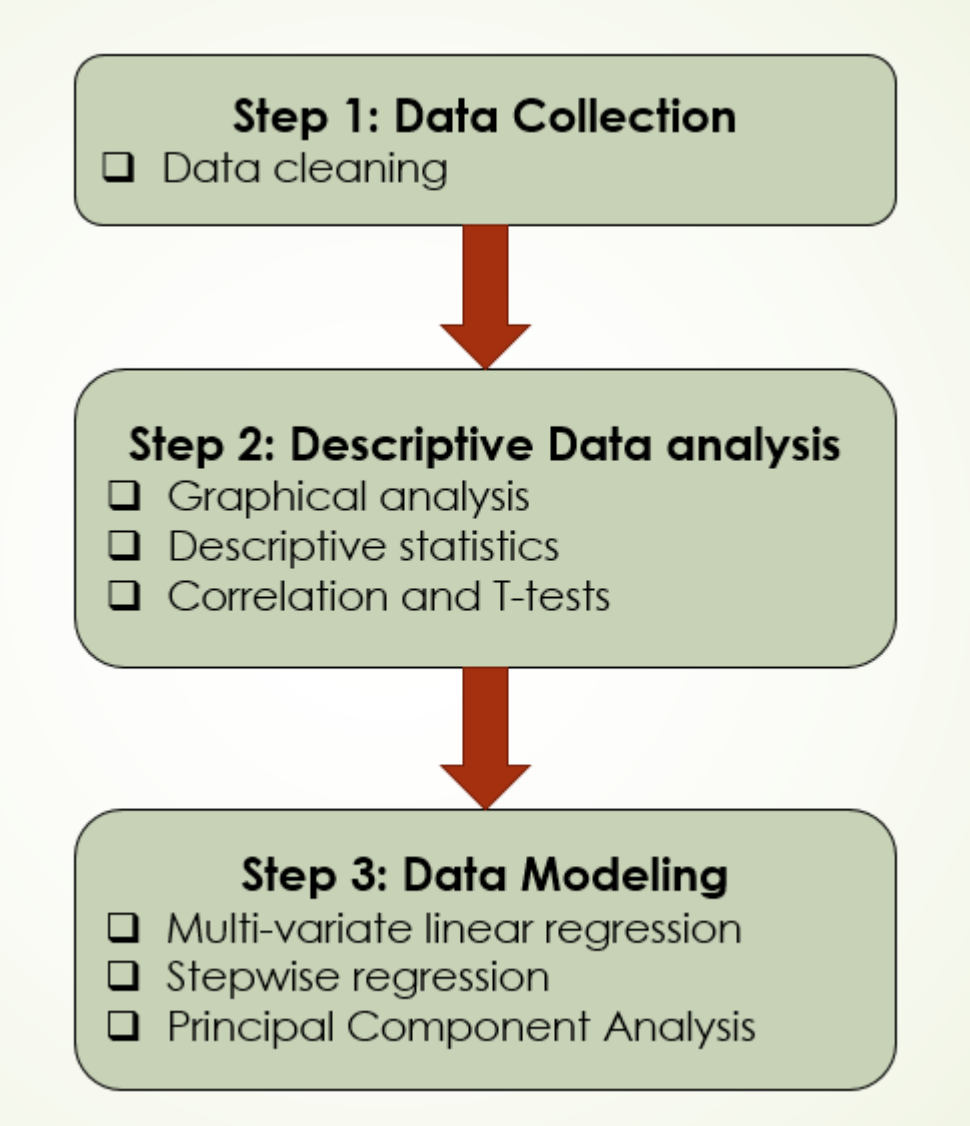


Figure 1: Methodology

3.1 Step 1: Data Collection

The data in this research were extracted by VA from the Department of Veteran Affairs VSSC (VHA Support Service Center) reports such as the Specialty Provider Workforce Report, the Special Productivity Access Report, and the Quadrant Tool Specialty Provider Productivity Standards Performance Report, and other pertinent VSSC reports with facility VISN and national

efficiency information about the 134 Veteran Affairs eye care clinics. This project uses retrospective cross-sectional data from the 2017, 2018, and 2019 fiscal years. All input/output variables were obtained in aggregate form for each facility within a VISN. Hence, no patient or employee identifiers were used for this project.

Input and Output Variables

The list of input and output variables for both ophthalmology and optometry data is shown in Table 1.

Table 1: List of input/output variables

INPUT VARIABLES	OUTPUT VARIABLE
Residents	Productivity
MCG	
Physician Clinical FTEE	
Adj MDFTE	
Specialty Unique Patients	
Core Unique Patients	
Physician Clinical FTEE per 10K Specialty Unique	
Physician Clinical FTEE per 100K Core Facility Unique	
Specialty Unique Encounters	
Associate Providers per Physician Clinical FTEE	
Admin Support staff per10k Physician Clinical FTEE	
Associate Providers per 100K Core Facility Unique	
Admin Support Staff per Physician Clinical FTEE	
Clinical Support Staff per Physician Clinical FTEE	
Total Support Staff per Physician Clinical FTEE	

We will next describe each input and output variable, starting first with the output variable below:

Output Variable: In this study, Clinical Productivity was selected as the efficiency measure. For VHA productivity is defined as the ratio of total Physician "Work Relative Value Unit" (wRVU) for the entire specialty provider and Physicians Clinical time in terms of Full-Time Equivalent Employees (FTEE):

$$Productivity = \frac{wRVU}{Physician\ Clinical\ FTEE}$$

Relative Value Unit (RVU) is a measure of the difficulty and expense of a professional service. The number of RVUs associated with each CPT (Current Procedural Terminology) code is determined by the Centers for Medicare and Medicaid Services (CMS). RVUs are primarily designed for reimbursement purposes but have been widely employed to measure physician work effort/workload as well. The total RVU consists of three components: physician work (wRVU), practice expense (peRVU), and malpractice expense (mpRVU). RVU tables may be obtained from the VA Office of Productivity, Efficiency & Staffing staff upon request. For productivity measurement, only the wRVU is utilized. For purposes of physician productivity measurement, only the specialty group practice physician clinical work component of the RVU (wRVU) value was utilized (VHA Handbook Productivity and Staffing Guidance for Specialty Provider Group Practice, 2015). We also would like to note that the productivity metric used in this report at the VISN, Medical Center complexity Group (MCG), and/or Facility level is a measurement of productivity for VA paid Physician Staff Only, that is it excludes In-House FTEE & Contract Physician Clinical FTEE and RVU work. Full-Time Equivalent Employee (FTEE) is a unit that indicates the workload of a full-time employee in VHA, which helps to compare their workloads across various contexts. It can also be defined as a staffing parameter equal to the amount of time assigned to one full-time employee. It may be composed of several part-time employees whose total time commitment equals that of a full-time employee. One FTE equals 40 hours per week.

A detailed explanation of each input variable is provided below:

- **Medical Center Complexity Group (MCG)** - Depending on the level of complexity the facilities in VHA are divided into 5 groups which are explained in the table below.

Table 2: Medical center complexity groups.

MCG	Complexity Level	Facility Description
1a	Highest complexity	Facility with high volume, high risk patients, most complex clinical programs, and large research and teaching programs
1b	Highest complexity	Facilities with medium-high volime, high risk patients, many complex clinical programs, and medium-large research and teaching programs
1c	Mid-High complexity	Facilities with medium-high volume, medium risk patients, some complex clinical programs, and medium sized research and teaching programs
2	Medium complexity	Facilites with medium volum, low risk patients, few complex clinical programs, and small or no research and teacjing programs
3	Low complexity	Facilities with low volum, low risk patients, few or no complex clinical programs, and small or no research and teaching programs

- **Residents** are doctors in training who have a doctorate. Ophthalmology residents must complete a general postgraduate year (PGY) approved by the Accreditation Council for Graduate Medical Education (ACGME) and a minimum of 3 years of postgraduate training in an ACGME-accredited ophthalmology training program to become eligible for certification by the American Board of Ophthalmology. Optometry Residents are Post Graduate Year 1 (PGY 1) trainees in a primary eye or specialty vision care residency (VHA DIRECTIVE, 2019).
- **Physician Clinical FTEE** is the total amount of Physician working hours expressed as FTEE (Full-Time Employee Equivalent) that is assigned to clinical duties, excluding administration, teaching, and research time as defined in the VA DSS (Decision Support System) databases. Annual leave, sick leave, and leave without pay hours are excluded from the reported FTEE.
- **Adjusted MD FTEE** is the total amount of VA paid physicians who worked FTEE that is assigned to clinical duties in the specialty, excluding administration, teaching, and research as defined in DSS databases.
- **Specialty Unique Patients** are the unique patients seen in a specialty. In other words, the total unique patients treated by a physician within said specialty.

- **Core Unique Patients** are "All Patients" counted as unique patients at the VISN, MCG, and Facility summary levels. Female unique patients are used in the OB/GYN Specialty report.
- **Physician Clinical FTEE per 10K Specialty Unique** is the number of FTEE for 10,000 Patients.
- **Physician Clinical FTEE per 100K Specialty Unique** is the number of FTEE for 100,000 Specialty Unique Patients.
- **Specialty Unique Encounters** is the number of cases requiring unique specialized medical services.
- **Associate Providers per Physician Clinical FTEE** is the ratio of Nurse Practitioner RVU to the Physician Clinical FTEE.
- **Associate Providers per 10K Specialty Unique** is the Nurse Practitioner RVU per 10,000 Specialty Unique
- **Associate Providers per 100K Core Facility Unique** is the Nurse Practitioner RVU for 100,000 Core Facility Unique
- **Admin Support Staff per Physician Clinical FTEE** is the staff who support the administration activities like clerks, receptionists. This variable gives the ratio of the work RVU of Admin staff to physician Clinical FTEE.
- **Clinical Support Staff per Physician Clinical FTEE** are the ones who support the administrative activities of Nurse Practitioners. This variable gives the ratio of the work RVU of clinical staff to physician Clinical FTEE.
- **Total Support Staff per Physician Clinical FTEE** is the ratio of the staff who perform support duties that facilitate the work of clinical staff throughout the Medical Center to Physician Clinical FTEE.

3.1.1 Data Cleaning

There are several data outliers and missing data values in both optometry and ophthalmology data. Therefore, data cleaning and preparation were performed here as explained in detail in Chapters 4 and 5 for optometry and ophthalmology data, respectively.

3.2 Step 2: Descriptive Data Analysis

As we have many independent variables for one output variable, some exploratory data analysis is performed before building the regression model. This enables us to understand the relationship between input and output variables. For this purpose, we have performed graphical analysis, computed descriptive statistics, and analyzed correlation, which will be explained next.

3.2.1 Graphical Analysis

A graphical representation of data can help to understand the data distribution. We have used normal distribution graphs to study the mean, standard deviation and to observe the symmetry of the data. Scatter plots were used to visually detect the correlation between the input and output variables. These graphs will be presented for optometry and ophthalmology data in Chapters 4 and 5.

3.2.2 Descriptive Statistic

Along with the graphical analysis, descriptive statistics were computed for preliminary analysis of all the variables. For this, we have documented the mean, median, mode, minimum and maximum values, skewness, and kurtosis for the optometry and ophthalmology data.

3.2.3 Correlation Analysis and T-test

Some of the data points in our data may be correlated, and therefore besides the visual analysis we performed a more formal Persons Correlation Analysis. Typical ranges and interpretation for

different Person's correlation values r values are given in Figure 2. In addition, T-tests were also conducted to check the statistically significant correlation values.

Range of Corelation Coefficient Values	Level of Correlations	Range of Corelation Coefficient Values	Level of Correlations
0.80 to 1.00	Very Strong Positive	-1.00 to -0.80	Very Strong Negative
0.60 to 0.79	Strong Positive	-0.79 to -0.60	Strong Negative
0.40 to 0.59	Moderate Positive	-0.59 to -0.40	Moderate Negative
0.20 to 0.39	Weak Positive	-0.39 to -0.20	Weak Negative
0.00 to 0.19	Very Weak Positive	-0.19 to -0.01	Very Weak Negative

Table 3: Range and Interpretation for Persons Correlation values (Meghanathan N, 2016).

3.3 Step 3: Data Modeling

For data modeling, we built 26 models that are systematically summarized in Table 4 to compare several approaches and to select the best approach. In building these models, we changed the regression type, inclusion of PCA, focus on high/low/all performers, and split sampling for testing and validation of the proposed models for both optometry and ophthalmology. Hence, 13 models were built for optometry and 13 for ophthalmology for the analysis. We have utilized SPSS statistical software package to run these models. The corresponding results and analysis of these models will be presented for optometry and ophthalmology in Chapters 4 and 5, respectively.

Table 4: Analyzed Models-Experimental Design for Sensitivity

No.	Model Code	Model Name	Regression Type	PCA	Performance	Split Sampling	Data
1	M11	MLR	MLR	No	All	No	Optometry
2	M12	MLR-High	MLR	No	High	No	Optometry
3	M13	MLR-Low	MLR	No	Low	No	Optometry
4	M14	MLR-Split	MLR	No	All	Yes	Optometry
5	M21	MLR-PCA100	MLR	100%	All	No	Optometry
6	M22	MLR-PCA100-High	MLR	100%	High	No	Optometry
7	M23	MLR-PCA100-Low	MLR	100%	Low	No	Optometry
8	M24	MLR-PCA100-Split	MLR	100%	All	Yes	Optometry
9	M31	MLR-PCA80	MLR	80%	All	No	Optometry
10	M32	MLR-PCA80-High	MLR	80%	High	No	Optometry
11	M33	MLR-PCA80-Low	MLR	80%	Low	No	Optometry
12	M41	Step-PCA80	Step	80%	All	No	Optometry
13	M42	Step-PCA100	Step	100%	All	No	Optometry
14	M51	MLR	MLR	No	All	No	Opthalmology
15	M52	MLR-High	MLR	No	High	No	Opthalmology
16	M53	MLR-Low	MLR	No	Low	No	Opthalmology
17	M54	MLR-Split	MLR	No	All	Yes	Opthalmology
18	M61	MLR-PCA100	MLR	100%	All	No	Opthalmology
19	M62	MLR-PCA100-High	MLR	100%	High	No	Opthalmology
20	M63	MLR-PCA100-Low	MLR	100%	Low	No	Opthalmology
21	M64	MLR-PCA100-Split	MLR	100%	All	Yes	Opthalmology
22	M71	MLR-PCA80	MLR	80%	All	No	Opthalmology
23	M72	MLR-PCA80-High	MLR	80%	High	No	Opthalmology
24	M73	MLR-PCA80-Low	MLR	80%	Low	No	Opthalmology
25	M81	Step-PCA80	Step	80%	All	No	Opthalmology
26	M82	Step-PCA100	Step	100%	All	No	Opthalmology

Next, we will provide a brief description of the utilized modeling techniques below:

3.3.1 Multiple Linear Regression (MLR)

MLR is a regression model where several independent variables are used to predict the outcome of a dependent variable (Kenton, W. 2020). MLR Model formulation can be written as follows:

$$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \mathcal{E},$$

Where y: is the predicted outcome or the dependent variable.

β_0: is the y-intercept,

$\beta_i X_i$: is the regression coefficient (β_i) of the ith independent variable (X_i),

$\mathcal{E}$: is the error term.

For this regression to be valid, there are several assumptions to be validated (Chatterjee and Hadi, 2012; Williams, 2013):

1. Linearity in the parameters: The model that relates the response y to the predictors $X_1, \ldots X_n$ is assumed to be linear in the regression parameters.
2. Errors: the errors are assumed to be random and independent identically distributed.
3. Normality: Classical MLR assumes further that errors are normally distributed for hypothesis testing purposes.
4. Homoscedasticity: MLR assumes that variance of residuals should remain the same for different values of the outcome. The standardized residuals versus predicted values plot help to understand if they are equally distributed.
5. Multi collinearity: MLR assumes that inputs are not correlated. This assumption is needed mathematically for carrying out the estimation of the regression parameters. If a correlation exists between two or more than two predictor variables, then it is said to have multicollinearity. In severe cases (such as a perfect correlation between two or more predictors), multicollinearity can mean that no unique least-squares solution to a regression analysis can be computed (Belsley, 1980; Slinker, 1985). Less severe multicollinearity often gives an unstable estimate of the coefficients of the predictor variables. That is, the standard errors and confidence intervals for the coefficient estimates will be inflated (Belsley et al., 1980).

MLR is based on ordinary least squares (OLS), the model fit is such that the sum-of-squares of the difference of observed and predicted values is minimized (Kenton, W. 2020). It assumes that the total variation in the data behavior will be partially explained by the regression model and partially by the random error.

$$SS_T = SS_R + SS_E,$$

Where SS_R – is the variation explained by the regression model, SS_E is the variation explained by error, and SS_T is the total variation. An important performance measure for regression analysis is the proportion of variation explained by regression compared to the overall variation in the problem analyzed. This proportion is referred to as the coefficient of determination R^2 and computed as follows:

$$R^2 = \frac{SS_R}{SS_T}$$

R^2 can be viewed as a "goodness of fit" measure varying between 0 and 1, which by definition with values closer to 1 showing a better fit. A modified, perhaps a more reliable version of this measure is the adjusted coefficient of determination of simply Adjusted-R^2, which recognizes the fact that inclusion of additional inputs (whether they are significantly useful or not) increases R^2, may cause overfitting. Hence, Adjusted-R^2 adjusts to reflect the "true contribution" of the inputs to the performance of the regression model based on the number of inputs k and number of observations n as follows:

$$\text{Adjusted-}R^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$$

3.3.2 Stepwise Regression

Stepwise regression analysis selects the variables in a step-by-step manner (Hintze, J. L., Dr. 2007). The independent variables are added or removed one by one depending on their statistical significance. In other words, it adds either the most significant variable or removes the least significant ones.

There are different ways to select the variables:

- **Forward selection:** This procedure is simple to define, the technique starts with no independent variables in the model, then selects the variable, which gives the highest

R-squared values. At each step, the independent variables which can increase the R-squared value are added and finally stops when none of the variables are statistically significant. This procedure is a good choice when there is a multicollinearity issue.

- **Backward selection:** This technique uses all the independent variables in the model and then removes the least significant ones in each step. This continues until no non-significant variables remain.

3.3.3 Principal Component Analysis (PCA)

Principal component regression is a two-step multivariate calibration method where in the first step, Principal Component Analysis, (PCA) is performed where the measure variables are converted to factors and in the second step a Multiple Linear regression between factors obtained in the PCA step and the dependent variable (y) to be modeled (Maesschalck, 1999). The step-by-step procedure to obtain Principal Components is explained next.

1. **Standardization:** It is performed to standardize the range of continuous initial variables for equal contribution to the analysis. As PCA is sensitive regarding the variance of initial variables they are transformed to comparable scales to avoid biased results.

2. **Co-variance matrix:** This step is to check if there is any relation between the input variables.

3. **Computing Eigenvectors and Eigen Values** of the covariance matrix for Principle Components: These values are computed from the covariance matrix to determine the principal components. Principal components are new variables that are constructed as a linear combination of initial variables. They are uncorrelated by definition (Jaadi, Z., 2021). Principal components help to reduce dimensionality without losing the information.

For interpreting the extracted components in PCA, we perform rotation. Yaremko, Harari, Harrison, and Lynn (1986), defined rotation as follows: "In factor or principal-components analysis, rotation of the factor axes (dimensions) identified in the initial extraction of factors, to obtain simple and interpretable factors." An orthogonal rotation method assumes that the components in the analysis are uncorrelated. Gorsuch (1983, pp. 203-204) lists four different orthogonal methods: equamax, orthomax, quartimax, and varimax. Varimax rotations are used to maximize the sum of the variance of the squared loadings, where loadings mean correlation between variables and factors. In simple words, varimax rotation is a "statistical technique used at one level of factor analysis as an attempt to clarify the relationship among factors" (Allen, M. 2017). After the components are obtained from PCA, the Varimax rotation is applied to adjust the coordinates of these components to maximize the variance among the shared items. By maximizing the variance, the results show how the data correlate with each principal component. We have used SPSS software to perform this rotation.

CHAPTER 4: OPTOMETRY DATA MODELING AND ANALYSIS

In Chapter 3, we have presented an overview of the methodology and the data that went into modeling analysis. In this chapter, we will explain the application of the methodology for the optometry data. After providing a brief introduction to optometry services below, we will proceed with the discussion of the methodological steps, more specifically the data collection (step1), descriptive data analysis (step2), and data modeling (step 3) for optometry analysis.

Optometry is a field of science that uses suitable instruments or appliances to examine the eyes for defects in vision and disorders. They provide corrective lenses or may advise to consultant an ophthalmologist depending on the severity of the condition. VA optometry provides three-quarters of all eye care in the VA where optometrists are performing at 96% of these sites (VHA Handbook, 2011). An optometrist is a Doctor of Optometry (OD) who is authorized to give primary and special eye and vision care administrations. An optometrist is medicinally prepared to examine, analyze, treat, and oversee illnesses and disorders of the visual framework, the eye, and related structures, and analyze related systemic conditions. This incorporates, but is not constrained to diagnosis, treatment, and the management of diabetic retinopathy, glaucoma, macular degeneration, and other eye illnesses; provision of refractions for eyeglass solutions, eyeglasses, medicinally important contact lenses, as well as low-vision and brain injury vision restoration administrations. Optometrists get 4 years of Doctoral-level degree after their baccalaureate training. Residency preparation is a 1-year past achievement of the optometry degree, and association preparation is for 1 to 2 years past the finishing of the residency.

4.1 Step 1: Data Collection- Optometry

An aggregated data was collected from the Optometry centers across the United States for the fiscal years 2017, 2018, and 2019. We have a total of 417 data entries for each variable, of which some are missing. All the variables except MCG are continuous data variables. MCG is a categorical data type. To include it in our regression model, we have coded MCG data variables into 5 different data variables called complex_1a, complex_1b, complex_1c, complex_2, and complex_3, all having 0 and 1 value, indicating whether the condition exists or not.

The list of all the data variables used for regression models are tabulated below:

Table 5: List of Data Variables- Optometry

S.NO	NAME OF MEASURE	INPUT/OUTPUT	OPTOMETRY	
			VALID	MISSING
1	Residents	Input	220	197
2	Physician Clinical FTEE	Input	413	4
3	Adj MDFTE	Input	399	18
4	Specialty Unique Patients	Input	413	4
5	Core Unique Patients	Input	417	0
6	Physician Clinical FTEE per 10K Specialty Unique	Input	413	4
7	Physician Clinical FTEE per 100K Core Facility Unique	Input	417	0
8	Specialty Unique Encounters	Input	413	4
9	Associate Providers per Physician Clinical FTEE	Input	302	115
10	Admin Support staff per10k Physician Clinical FTEE	Input	302	115
11	Associate Providers per 100K Core Facility Unique	Input	306	111
12	Admin Support Staff per Physician Clinical FTEE	Input	409	8
13	Clinical Support Staff per Physician Clinical FTEE	Input	397	20
14	Total Support Staff per Physician Clinical FTEE	Input	409	8
15	complex_1a	Input	417	0
16	complex_1b	Input	417	0
17	complex_1c	Input	417	0
18	complex_2	Input	417	0
19	complex_3	Input	417	0
20	Productivity	Output	399	18

We have different types of data types in this relatively large data set with different scales. To make it an “apples to apples” comparison, we brought all the variables to the same scale by applying standardization (or Z-score Normalization), where all input variables are rescaled by removing the respective mean by dividing by the respective standard deviation.

From Table 5, we can observe that for residents out of 417 data values 197 data values are missing, which accounts for almost 50% of the data. For variables like Associate Providers per Physician Clinical FTEE, Admin Support staff per10k Physician Clinical FTEE, Associate Providers per 100K Core Facility Unique around 30% of the data values are missing. If we drop these missing data values for our data set, it decreases our sample size. So, we must either completely ignore these variables as a whole or fill in the missing values. As our project aim is to analyze the relation between input and output variables, disregarding these missing data variables will not serve our goal. Hence, we have filled the missing data values using data cleaning methods as described next.

4.1.1 Data Cleaning

Most of the facilities have residency students working with them. But for the facilities located in the rural areas, there may not be any academic institutions near those locations. Hence, we have the missing data value for residents. After investigating this further, we concluded that no residency students were attending the facility for that year. Therefore, we can consider the missing values for residents as 'zeros'.

The variables Associate Providers per Physician Clinical FTEE, Admin Support staff per10k Physician Clinical FTEE, Associate Providers per 100K Core Facility Unique, deal with the values of clerical and support staff in the facilities. Optometry facilities deal with primary care services where they are not provided with support staff in most of the facilities. Hence these missing data values can also be considered as ‘zeros’ for these input variables.

For the other variables with missing data values like productivity, Adj MDFTEE, and Physician Clinical FTEE, we needed to perform further cleaning. There are many data cleaning methods available. In our data, the missing values do not show any pattern so we can substitute

these using the mean replacement method, where the missing data values are replaced by the mean of the data set (Little and Rubin, 2002).

The missing data values for the remaining data variables such as Specialty Unique Patients, Physician Clinical FTEE per 10K Specialty Unique, Specialty Unique Encounters, Admin Support Staff per Physician Clinical FTEE, Total Support Staff per Physician Clinical FTEE, accounts for less than 5% of the data. Hence, we have dropped those data entries.

4.2 Step 2: Descriptive Data Analysis- Optometry

4.2.1 Graphical Analysis

Histogram graphs for all the continuous input variables of optometry are computed and presented in Appendix B. We observe data outliers for productivity and other variables. There seems to be a large variation in data values as we are considering the data from eye care facilities spread over the entire United States. Hence outliers indicate that there are best or worst performing facilities, which requires further investigation. A graphical representation to estimate the individual effect of the input variable on the output variable is attached in Appendix C. The graphs show most of the variables have a relatively weak relationship with productivity.

4.2.2 Descriptive Statistics

For preliminary analysis of all the variables, we have computed the mean, minimum and maximum values, skewness, and kurtosis for all the variables. Table 6 shows descriptive statistics for Optometry data.

The general rule for skewness is that if the skewness value is greater than +1 or less than -1, it is an indication of a skewed distribution. This is true in the case of all the variables except Physician Clinical FTEE per10K Specialty Unique, Physician Clinical FTEE per 100K Core Facility Unique, Clinical Support Staff per Physician Clinical FTEE. Similarly, for kurtosis, if the

values are greater than +1 then the distribution is too peaked if it's less than -1 then the distribution is too flat (Hair et al., 2017). Expect for complex_1a, complex_1c all the variables in our data have peaked distributions.

Table 6: Descriptive statistics

S.No	Variable	N	Range	Minimum	Maximum	Mean	Skewness	Kurtosis
1	Residents	417	11.000	0.000	11.000	1.583	1.633	3.017
2	SpecialtyUniquePatients	417	46801.000	0.000	46801.000	11254.835	1.891	4.832
3	CoreUniquePatients	417	130662.000	12449.000	143111.000	51205.177	1.002	0.505
4	PhysicianClinicalFTEEper10KSpecialtyUniques	417	10.213	0.000	10.213	4.415	0.199	2.931
5	PhysicianClinicalFTEEper100KCoreFacilityUniques	417	29.297	0.000	29.297	10.147	0.421	0.149
6	SpecialtyUniqueEncounters	417	69888.000	0.000	69888.000	15911.643	1.837	4.193
7	AssociateProvidersperPhysicianClinicalFTEE	417	16.170	0.000	16.170	0.120	13.142	177.978
8	AssociateProvidersper10KSpecialtyUniques	417	40.001	-0.001	40.000	0.347	12.472	162.485
9	AssociateProvidersper100KCoreFacilityUniques	417	6.527	-0.002	6.525	0.191	7.968	89.547
10	AdminSupportStaffperPhysicianClinicalFTEE	417	2.779	0.000	2.779	0.463	1.881	7.686
11	ClinicalSupportStaffperPhysicianClinicalFTEE	417	3.523	0.000	3.523	0.911	0.763	0.753
12	TotalSupportStaffperPhysicianClinicalFTEE	417	5.735	0.000	5.735	1.374	1.085	2.633
13	complex_1a	417	1.000	0.000	1.000	0.281	0.980	-1.044
14	complex_1b	417	1.000	0.000	1.000	0.151	1.956	1.833
15	complex_1c	417	1.000	0.000	1.000	0.230	1.286	-0.347
16	complex_2	417	1.000	0.000	1.000	0.139	2.093	2.394
17	complex_3	417	1.000	0.000	1.000	0.199	1.513	0.290
18	PhysicianClinicalFTEE_1	417	26.981	0.000	26.981	5.019	2.231	8.083
19	AdjMDFTE_1	417	26.254	0.640	26.894	4.911	2.565	10.553
20	ProductivityMeasure_1	417	17819.029	1410.596	19229.625	5327.577	2.027	17.470

4.2.3 Correlations and T-test

T-test was performed to see whether the correlation between dependent variable productivity and the independent variables is statistically significant at the selected confidence interval (Table 7). The results show that except for Residents, |t stat|> t critical value (1.653), hence we reject the Null hypothesis and there is a statistically significant correlation between productivity and the factors considered.

Table 7: One sample T-test

test value=1.653	
Predictors	t
Residents	-0.683
SpecialtyUniquePatients	28.839
CoreUniquePatients	38.509
PhysicianClinicalFTEEper10KSpecialtyUniques	47.352
PhysicianClinicalFTEEper100KCoreFacilityUniques	33.972
SpecialtyUniqueEncounters	27.261
AssociateProvidersperPhysicianClinicalFTEE	-27.599
AdminSupportstaffper10kPhysicianClinicalFTEE	-9.716
AssociateProvidersper100KCoreFacilityUniques	-63.604
AdminSupportStaffperPhysicianClinicalFTEE	-74.728
ClinicalSupportStaffperPhysicianClinicalFTEE	-25.318
TotalSupportStaffperPhysicianClinicalFTEE	-6.974
complex_1a	-62.304
complex_1b	-85.538
complex_1c	-68.934
complex_2	-89.232
complex_3	-74.272
PhysicianClinicalFTEE_1	18.917
AdjMDFTE_1	19.56

Pearson correlation matrix was also computed to find the correlation coefficients between all variables. The results shown in Table 8 indicate that there is a strong correlation between input variables showing high chances of Multi-collinearity in the data. Multi Collinearity violates the assumptions of MLR, making it difficult to run the regression analysis.

Table 8: Data Correlations

	Residents	Physician Clinical FTEE	Adj MDFTE	Specialty Unique Patients	Core Unique Patients	Physician Clinical FTEE per 10K Specialty Uniques	Physician Clinical FTEE per 100K Core Facility Uniques	Specialty Unique Encounters	Associate Providers per Physician Clinical FTEE	Admin Support staff per10k Physician Clinical FTEE	AssociateProviders per 100K Core Facility Uniques	Admin Support Staff per Physician Clinical FTEE	Clinical Support Staff per Physician Clinical FTEE	Total Support Staff per Physician Clinical FTEE	complex_1a	complex_1b	complex_1c	complex_2	complex_3
Residents	1																		
Physician Clinical FTEE	.626**	1.00																	
Adj MDFTE	.615**	.985**	1.00																
Specialty Unique	.592**	.910**	.896**	1.00															
Core Unique Patients	.254**	.623**	.593**	.663**	1.00														
Physician Clinical FTEE	.191**	.268**	.247**	0.00	0.07	1.00													
Physician Clinical FTEE	.522**	.547**	.515**	.408**	-.163**	.479**	1.00												
Specialty Unique	.673**	.921**	.902**	.977**	.615**	0.06	.469**	1.00											
Associate Providers per	-0.07	-.126*	-.164**	-.127**	-0.04	-.136**	-.178**	-.121*	1.00										
Admin Support staff per10k	-0.08	-.137**	-.165**	-.140**	-0.04	-.120*	-.195**	-.133**	.992**	1.00									
AssociateProviders per 100K	-.114*	-0.07	-0.08	-0.05	-0.09	-0.09	0.00	-0.06	0.03	0.04	1.00								
Admin Support Staff per Clinical	-0.07	-0.09	-0.09	-0.03	-.139**	-.139**	-0.03	-0.03	-0.05	-0.04	.398**	1.00							
Support Staff	-.208**	-.204**	-.226**	-.135**	-.197**	-.151**	-.129**	-.139**	-.105*	-.105*	.227**	.521**	1.00						
Total Support Staff per	-.179**	-.185**	-.204**	-.110*	-.200**	-.166**	-.105*	-.113*	-.098*	-0.09	.325**	.780**	.940**	1.00					
complex_1a	.259**	.432**	.400**	.443**	.667**	.128**	-0.06	.418**	-0.05	-0.05	-0.05	-.155**	-.244**	-.241**	1.00				
complex_1b	0.00	-0.05	-0.03	-0.04	.101*	-.125*	-.200**	-0.05	0.00	0.02	-0.03	-.145**	-0.06	-.102*	-.263**	1.00			
complex_1c	0.00	-0.01	-0.04	0.00	-.119*	0.04	.127**	0.02	-0.05	-0.06	-0.05	0.03	0.08	0.07	-.342**	-.231**	1.00		
complex_2	-.118*	-.164**	-.155**	-.199**	-.281**	-0.01	0.04	-.196**	.160**	.153**	0.06	0.06	0.07	0.08	-.251**	-.170**	-.220**	1.00	
complex_3	-.190**	-.295**	-.259**	-.298**	-.472**	-0.06	0.08	-.268**	-0.03	-0.03	0.09	.218**	.183**	.221**	-.311**	-.210**	-.273**	-.200**	1

**. Correlation is significant at the 0.01 level (2-tailed)

*. Correlation is significant at the 0.05 level (2-tailed).

4.3 Step 3: Data Modeling- Optometry

To find the best fit model for our data we have tested the data with different models which are listed below in Table 9.

Table 9: Overview of Data Modeling

No.	Model Code	Model Name	Regression Type	PCA	Performance	Split Sampling	Data
1	M11	MLR	MLR	No	All	No	Optometry
2	M12	MLR-High	MLR	No	High	No	Optometry
3	M13	MLR-Low	MLR	No	Low	No	Optometry
4	M14	MLR-Split	MLR	No	All	Yes	Optometry
5	M21	MLR-PCA100	MLR	100%	All	No	Optometry
6	M22	MLR-PCA100-High	MLR	100%	High	No	Optometry
7	M23	MLR-PCA100-Low	MLR	100%	Low	No	Optometry
8	M24	MLR-PCA100-Split	MLR	100%	All	Yes	Optometry
9	M31	MLR-PCA80	MLR	80%	All	No	Optometry
10	M32	MLR-PCA80-High	MLR	80%	High	No	Optometry
11	M33	MLR-PCA80-Low	MLR	80%	Low	No	Optometry
12	M41	Step-PCA80	Stepwise	80%	All	No	Optometry
13	M42	Step-PCA100	Stepwise	100%	All	No	Optometry
14	M43	Step-PCA100-High	Stepwise	100%	High	No	Optometry
15	M44	Step-PCA100-Low	Stepwise	100%	Low	No	Optometry
16	M45	Step-PCA100-Split	Stepwise	100%	All	Yes	Optometry

4.3.1 Multiple Linear Regression Analysis-MLR (M11)

ANOVA analysis for the M11 regression model is shown in Table 10. While the Adjusted R squared of 0.264 is not very high (Table 11), as seen in Table 10, the regression is significant at $\propto = 0.05 = 5\%$, in fact with very low Sig. (Significance) values close to zero, indicating close to 100% confidence.

Table 10: ANOVA Table for M11

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	122.9	17	7.185	9.756	0.000
Residual	293.853	399	0.736		
Total	416	416			

Table 11: Regression Statistics for M11

Regression Statistics	
R	0.542
R Squared	0.294
Adjusted R Squared	0.264
Standard Error	0.858
observations	416

Based on Table 12, T-tests show the following as statistically significant input variables:

- Residents
- PhysicianClinicalFTEEper10KSpecialtyUniques
- PhysicianClinicalFTEEper100KCoreFacilityUniques
- Specialty Unique Encounters
- Admin Support staff per 10k Physician Clinical FTEE
- Total Support staff per Physician Clinical FTEE
- Physician Clinical FTEE

The other variables did not show any statistical significance. Controlling physician clinical FTEE, per 10K Specialty Unique, Associate Providers per 10K Core Facility Unique, complex_1c, the regression coefficient [B= 0.315, 95%, Sig. <0.05] suggest that for each unit of residents, the productivity increases approximately by 0.315 units. Similarly, productivity decreases by 0.049 units for a one-unit change in Specialty unique patients. Table 12 also shows variation inflation factors (VIFs), which indicates collinearity among variables. A VIF less than 10 is acceptable (Myers, R.H. 1990) in the regression model, if there are values more than 10, it shows there is a problem of multicollinearity for our independent variables such as Specialty Unique patients, specialty Unique encounters, Associate Providers per Physician

Clinical FTEE, Admin Support staff per10k Physician Clinical FTEE, Physician Clinical FTEE.

Table 12: Regression Coefficients for M11

Predictors	Coefficient	Std. Error	t	Sig.	VIF
(Constant)	-1.70E-15	0.043	0	1	
Residents	0.187	0.072	2.58	0.01	2.881
SpecialtyUniquePatients	-0.252	0.138	-1.834	0.068	10.381
CoreUniquePatients	-0.012	0.105	-0.119	0.905	6.066
PhysicianClinicalFTEEper10KSpecialtyUniques	-0.104	0.051	-2.051	0.041	1.411
PhysicianClinicalFTEEper100KCoreFacilityUniques	-0.206	0.104	-1.987	0.048	5.9
SpecialtyUniqueEncounters	1.009	0.185	5.442	0	18.853
AssociateProvidersperPhysicianClinicalFTEE	0.163	0.102	1.606	0.109	5.666
AdminSupportstaffper10kPhysicianClinicalFTEE	-0.277	0.118	-2.352	0.019	7.632
Associateprovidersper100KCoreFacilityUniques	0.172	0.097	1.776	0.077	5.134
AdminSupportStaffperPhysicianClinicalFTEE	-0.143	0.075	-1.905	0.058	3.116
TotalSupportStaffperPhysicianClinicalFTEE	0.246	0.084	2.924	0.004	3.88
complex_1a	-0.135	0.251	-0.54	0.59	34.586
complex_1b	-0.151	0.198	-0.765	0.445	21.475
complex_1c	-0.111	0.225	-0.492	0.623	27.81
complex_2	-0.194	0.176	-1.105	0.27	16.999
complex_3	-0.239	0.163	-1.463	0.144	14.648
PhysicianClinicalFTEE	-0.695	0.21	-3.308	0.001	24.26
AdjMDFTE	-0.025	0.115	-0.213	0.831	7.273

Figure 5 shows a 4-in-1 plot for validating MLR assumptions. The normal plot shows that most of the data points fall on a straight-line justifying normality. We can see some outliers as discussed before under the descriptive analysis. A similar observation is made with the histogram, which is fairly symmetric. The scatter plot of residuals vs predicted value shows no clear pattern in the distribution, hence we can conclude independence and randomness of the errors. The variance is also almost constant. Therefore, the data is homoscedastic. The last assumption in MLR is multicollinearity which can be checked in several ways. As indicated earlier, some VIF values shown in Table 12 are more than 10 indicating there is a certain extent of Multicollinearity, which we aim to address using PCA.

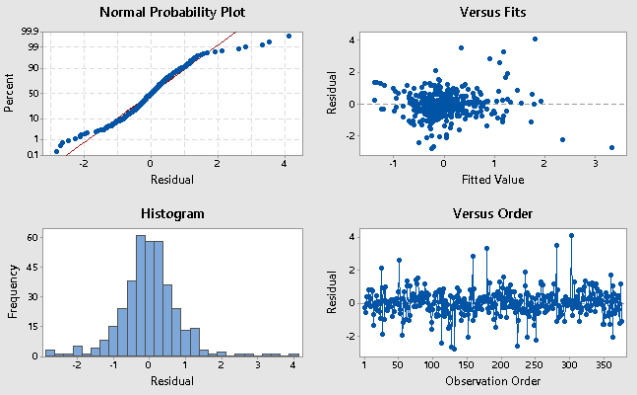


Figure 2: 4-in-1 Graph for MLR

4.3.2 MLR-High (M12)

One of the research objectives is to understand the influential factors for high and low-performing VAMCs. M12 model is a high performer analysis for the optometry services. Here we defined high performance as "performance being more than or equal to 1 standard deviation above the mean". Therefore, to find out which variables contribute to the high-level performance at different VAMCs, we have computed the z-score values of productivity and grouped the data for a z-score greater than the value 1. The results are shown in Tables 13-15. The R-square value shown in Table 14 indicates 85.4% of the variance is explained by the input variables. Table 32 shows that the regression is significant. The significant factors turn out to be:

- Core Unique Patients
- Physician Clinical FTEE per10K Specialty Unique
- Physician Clinical FTEE per 100K Core Facility Unique
- Specialty Unique Encounters
- Complex_1a
- complex_1b
- Physician Clinical FTEE

Table 13: ANOVA table for M12

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	69.798	17	4.106	7.902	0
Residual	11.951	23	0.52		
Total	81.749	40			

Table 14 Regression Statistics for M12

Regression Statistics	
R	0.924
R squared	0.854
Adjusted R squared	0.746
Standard Error	0.721
Observations	40

Table 15: Regression Coefficients for M12

Predictors	Coefficients	Std. Error	t	Sig.
(Constant)	6.053	3.093	1.957	0.063
Residents	0.024	0.23	0.106	0.916
SpecialtyUniquePatients	-0.623	0.782	-0.797	0.434
CoreUniquePatients	-2.293	0.545	-4.204	0
PhysicianClinicalFTEEper10KSpecialtyUniques	-2.933	0.541	-5.421	0
PhysicianClinicalFTEEper100KCoreFacilityUniques	-1.442	0.513	-2.81	0.01
SpecialtyUniqueEncounters	-3.863	0.754	-5.122	0
AssociateProvidersperPhysicianClinicalFTEE	28.964	86.744	0.334	0.741
AdminSupportstaffper10kPhysicianClinicalFTEE	13.529	49.4	0.274	0.787
AssociateProvidersper100KCoreFacilityUniques	-1.707	1.133	-1.506	0.146
AdminSupportStaffperPhysicianClinicalFTEE	-0.557	0.387	-1.438	0.164
ClinicalSupportStaffperPhysicianClinicalFTEE	0.459	0.311	1.476	0.154
complex_1a	1.167	0.505	2.312	0.03
complex_1b	-0.501	0.173	-2.892	0.008
complex_2	-0.409	0.2	-2.048	0.052
complex_3	0.305	0.214	1.427	0.167
PhysicianClinicalFTEE	6.374	1.347	4.732	0
AdjMDFTE	2.161	1.801	1.2	0.242

4.3.3 MLR-Low (M13)

Similar to the high performers, we defined low performance as "performance being less than or equal to 1 standard deviation below the mean". Therefore, to find out which variables contribute to the high-level performance at different VAMCs, we have computed the z-score values of productivity and grouped the data for z-score less than the value -1. Again, PCA was performed to remove collinearity from the data. The results are shown in Tables 16-19. The R-square value shown in Table 17 indicates 95.3% of the variance is explained by the input variables. Table 16 shows that the regression is significant. The significant factors turn out to be:

- Specialty Unique Patients
- Physician Clinical FTEE per 100K Core Facility Unique
- Adj MDFTE

Table 16: ANOVA for model M13

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	10.628	16	0.664	17.93	0
Residual	0.519	14	0.037		
Total	11.147	30			

Table 17: Regression Statistics for M13

Regression Statistics	
R	0.976
R squared	0.953
Adjusted R squared	0.9
Standard Error	0.192
Observations	30

Table 18: Regression Coefficients for M13

Predictors	Coefficients	Std. Error	t	Sig.
(Constant)	-1.05	0.205	-5.121	0
Residents	-0.548	0.494	-1.109	0.286
Specialty Unique Patients	1.926	0.41	4.697	0
Core Unique Patients	0.096	0.152	0.636	0.535
Physician Clinical FTEE per 10K Specialty Uniques	0.048	0.071	0.674	0.511
Physician Clinical FTEE per 100K Core Facility Uniques	-0.511	0.235	-2.179	0.047
Associate Providers per Physician Clinical FTEE	-0.12	0.121	-0.993	0.338
Associate Providers per 10K Specialty Uniques	0.127	0.129	0.98	0.344
AssociateProviders per 100K Core Facility Uniques	-0.097	0.125	-0.779	0.449
Admin Support Staff per Physician Clinical FTEE	-0.07	0.069	-1.014	0.328
Clinical Support Staff per Physician Clinical FTEE	0.001	0.05	0.013	0.99
complex_1a	-0.283	0.233	-1.213	0.245
complex_1b	-0.086	0.1	-0.861	0.404
complex_1c	-0.034	0.136	-0.248	0.807
complex_2	-0.021	0.109	-0.192	0.85
PhysicianClinicalFTEE	0.197	0.349	0.565	0.581
AdjMDFTE	-0.839	0.173	-4.86	0

4.3.4 MLR-Split (M14)

We are interested to check how well our models perform on a new data set and to know how well we can predict an observation in absolute terms. A common way to do this is to compare the Error Estimates like Mean Absolute Deviation (MAD, mean of absolute error), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percent Error (MAPE). For this, we randomly split the data in 60% and 40% ratios as Training and Test data sets and have run our regression model on Training data. The results are presented below from tables 19-24. The table 25, 26, and 27 show the regression models build on the training data. Using this regression equation, we try predicting the Productivity values for our Test data set and estimated the errors and calculated the R sq and adjusted R- sq values for the test data sample. From table 28 we can see the R – squared value and adj R – squared value for the test data sample as 0.459 and 0.398, respectively.

Table 19:ANOVA Table for Training Data

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	98.53	17	5.796	6.777	0
Residual	195.85	229	0.855		
Total	294.38	246			

Table 20: Regression Statistics for Training Data

Regression Statistics	
R	0.579
R squared	0.335
Adjusted R squared	0.285
Standard Error	0.925
Observations	246

Table 21: Regression Coefficients for Training Data

Predictors	Coefficients	Std. Error	t	Sig.
(Constant)	0.02	0.061	0.321	0.748
Residents	0.356	0.101	3.544	0
SpecialtyUniquePatients	0.517	0.347	1.488	0.138
CoreUniquePatients	0.179	0.172	1.045	0.297
PhysicianClinicalFTEEper10KSpecialtyUniques	-0.228	0.115	-1.985	0.048
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.231	0.165	1.396	0.164
SpecialtyUniqueEncounters	-0.124	0.339	-0.365	0.715
AssociateProvidersperPhysicianClinicalFTEE	0.236	1.35	0.174	0.862
AdminSupportstaffper10kPhysicianClinicalFTEE	-0.307	1.502	-0.204	0.838
AssociateProvidersper100KCoreFacilityUniques	-0.101	0.089	-1.14	0.255
AdminSupportStaffperPhysicianClinicalFTEE	0.123	0.08	1.532	0.127
ClinicalSupportStaffperPhysicianClinicalFTEE	0.111	0.072	1.544	0.124
complex_1b	0.058	0.074	0.784	0.434
complex_1c	0.1	0.096	1.044	0.298
complex_2	0.047	0.096	0.496	0.62
complex_3	0.138	0.111	1.238	0.217
PhysicianClinicalFTEE	0.486	0.315	1.544	0.124
AdjMDFTE	-1.173	0.242	-4.844	0

Table 22: Performance Matrix for Training data

ERROR	ESTIMATE
Mean Absolute Deviation	0.541
Mean Square Error	0.793
Root Mean Square Error	0.890
Mean Absolute Percennt Error	250.437

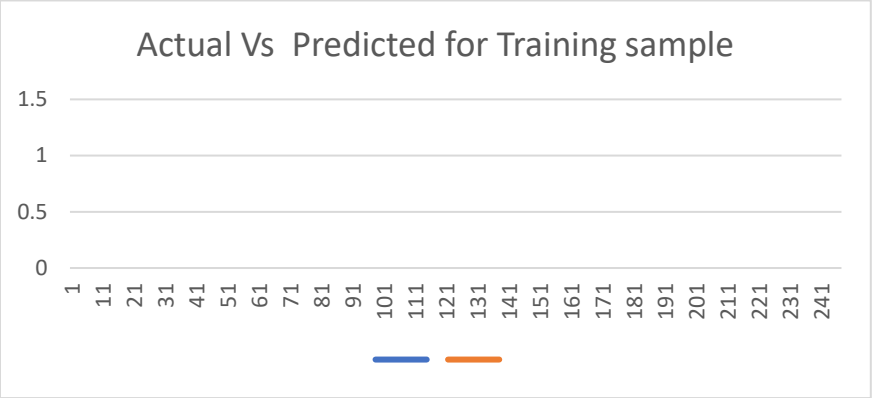


Figure 3: Actual Vs Predicted Graph for training data.

Table 23 Regression Statistics for Test Data

Regression Statistics	
R squared	0.459
Adjusted R squared	0.3987
SSE	64.765
SST	119.757
Observations	170

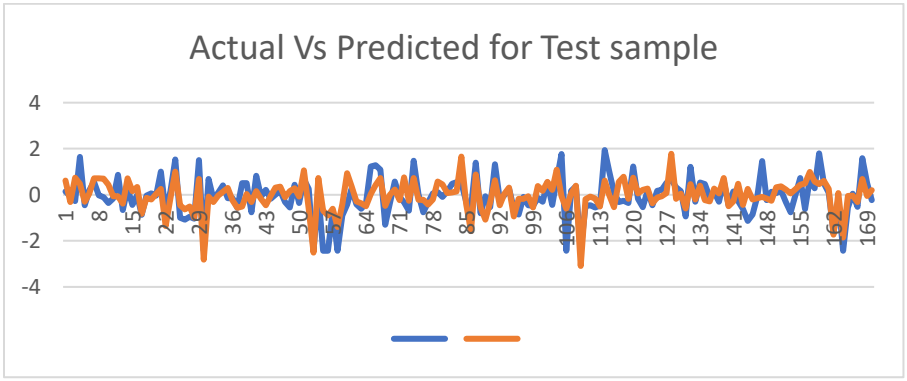


Figure 4: Actual Vs Predicted Graph for test data.

Table 24: Performance Matrix for Test data

ERROR	ESTIMATE
Mean Absolute Deviation	0.475
Mean Square Error	0.381
Root Mean Square Error	0.617
Mean Absolute Percennt Error	806.540

4.3.5 Principal Component Analysis-MLR-PCA100 (M21):

As indicated before, MLR modeling resulted in VIF values of more than 10. This means there is a high possibility of having multicollinearity in the data. To overcome this problem, principal component analysis (PCA) is performed to replace the 19 inter-related variables with independent components. We have used the Varimax rotation to simplify the factor structure and to make the interpretation easy. To retain 100% variance, we choose all 19 PCA components in the M21 and performed MLR analysis. The results are shown in Tables 25-28. Since Sig. ~0 (Table 25), the regression is significant at$\propto = 0.05 = 5\%$. The results in Table 26, show R-squared value is 0.295 and the adjusted R-squared value is 0.263. From Table 27, we can see that factors 4,5,8,9, 12 are statistically significant variables with Sig. <0.05. The component matrix for the significant factors is tabulated below and the most influential variables are chosen. We have considered all the absolute values of weighted averages above and highlighted values above 0.500. The contributing variables for the significant factors can be summarized as:

- **Factor_4**: Physician Clinical FTEE per 100K Core specialty Unique
- **Factor_5**: complex_1c, complex_3
- **Factor_8**: Physician Clinical FTEE per 10K specialty Unique
- **Factor_9**: Associate Providers per 100K Core specialty Unique

Table 25: ANOVA table for M21

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	122.574	18	6.81	9.236	0
Residual	293.426	398	0.737		
Total	416	416			

Table 26:Regression statistics for M21

Regression Statistics	
R	0.543
R squared	0.295
Adjusted R squared	0.263
Standard Error	0.858
Observations	416

Table 27: Regression Coefficients for M21

Predictors	Coeffiecients	Std. Error	t	Sig.	VIF
(Constant)	0.001	0.042	0.013	0.99	
FACTOR_1	0.007	0.042	0.167	0.867	1.017
FACTOR_2	0.067	0.042	1.585	0.114	1.008
FACTOR_3	-0.016	0.042	-0.371	0.711	1.005
FACTOR_4	-0.132	0.062	-2.13	0.034	2.165
FACTOR_5	-0.087	0.042	-2.063	0.04	1.002
FACTOR_6	-0.038	0.046	-0.818	0.414	1.187
FACTOR_7	-0.086	0.044	-1.957	0.051	1.085
FACTOR_8	-0.419	0.043	-9.82	0	1.03
FACTOR_9	-0.116	0.044	-2.618	0.009	1.116
FACTOR_10	0.02	0.042	0.463	0.643	1
FACTOR_11	0.067	0.042	1.588	0.113	1.005
FACTOR_12	0.276	0.051	5.433	0	1.461
FACTOR_13	0.062	0.043	1.434	0.152	1.054
FACTOR_14	-0.078	0.043	-1.824	0.069	1.023
FACTOR_15	-0.027	0.042	-0.631	0.528	1.003
FACTOR_16	0.058	0.042	1.376	0.169	1.016
FACTOR_17	0	0.051	-0.003	0.997	1.457
FACTOR_18	-260168.886	342126.682	-0.76	0.447	3.631

Table 28: Component Matrix for M21

Component Matrix					
	FACTOR_4	FACTOR_5	FACTOR_8	FACTOR_9	FACTOR_12
Residents	0.263	0.029	0.062	0.014	0.209
SpecialtyUniquePatients	0.030	0.072	0.154	0.081	0.032
CoreUniquePatients	0.472	0.007	0.129	0.104	0.176
PhysicianClinicalFTEEper10KSpecialtyUniques	0.317	0.182	0.590	0.402	0.104
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.624	0.056	0.067	0.030	0.261
SpecialtyUniqueEncounters	0.047	0.063	0.138	0.059	0.011
AssociateProvidersperPhysicianClinicalFTEE	0.495	0.043	0.052	0.101	0.027
AdminSupportstaffper10kPhysicianClinicalFTEE	0.483	0.040	0.089	0.101	0.019
AssociateProvidersper100KCoreFacilityUniques	0.053	0.167	0.498	0.639	0.072
AdminSupportStaffperPhysicianClinicalFTEE	0.049	0.061	0.081	0.031	0.073
ClinicalSupportStaffperPhysicianClinicalFTEE	0.086	0.111	0.031	0.347	0.011
TotalSupportStaffperPhysicianClinicalFTEE	0.044	0.057	0.009	0.242	0.021
PhysicianClinicalFTEE	0.104	0.010	0.006	0.020	0.005
AdjMDFTE	0.165	0.006	0.003	0.049	0.032
complex_1a	0.372	0.432	0.055	0.036	0.194
complex_1b	0.331	0.279	0.310	0.134	0.077
complex_1c	0.180	0.812	0.054	0.132	0.047
complex_2	0.384	0.110	0.286	0.045	0.097
complex_3	0.193	0.526	0.149	0.017	0.154

4.3.6 MLR-PCA100-High (M22)

Similar to MLR-High (M12), we have built a regression model for High performers and applied PCA as there is multi-collinearity in the data. We have performed PCA/factor analysis of the grouped data and the results are shown in tables 29-32. The R-squared value for this model is 0.856. It means 85.6% of the variance is explained by the input variables. Table 29 show that Factor_1, 2, 6, 7, 13, 14, 15 are statistically significant with sig. < 0.05. The significant factors with absolute weight values greater than 0.5 are highlighted in Table 30. The most influential input variables are summarized below:

- Factor_1: Residents, Specialty Unique Patients, Core Unique Patients, Specialty Unique Encounters, Admin Support Staff per Physician Clinical FTEE, Clinical Support Staff per Physician Clinical FTEE, Total Support Staff per Physician Clinical FTEE, PhysicianClinicalFTEE_1, AdjMDFTE_1, complex_1a

- Factor_2: Associate Providers per Physician Clinical FTEE, Admin Support staff per 10k Physician Clinical FTEE, Associate Providers per 100K Core Facility Unique, complex_1b
- Factor_6: complex_2, complex_3
- Factor_7: complex_1a

Table 29: ANOVA Table for MLR-PCA100-High

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	96.296	18	5.35	7.567	0
Residual	16.262	23	0.707		
Total	112.558	41			

Table 30: Regression Statistics for MLR-PCA100-High

Regression Statistics	
R	0.925
R squared	0.856
Adjusted R squared	0.742
Standard Error	0.840
Observations	41

Table 31: Regression Coefficients for MLR-PCA100-High

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	2.039	0.13	15.662	0	
Factor_1	-0.396	0.162	-2.446	0.023	1.517
Factor_2	0.418	0.158	2.64	0.015	1.453
Factor_3	-0.092	0.138	-0.671	0.509	1.1
Factor_4	-0.244	0.135	-1.809	0.084	1.052
Factor_5	-0.201	0.16	-1.257	0.221	1.479
Factor_6	0.347	0.144	2.41	0.024	1.199
Factor_7	0.623	0.14	4.445	0	1.141
Factor_8	-0.223	0.141	-1.583	0.127	1.156
Factor_9	0.118	0.134	0.881	0.387	1.034
Factor_10	-0.042	0.14	-0.298	0.768	1.13
Factor_11	0.17	0.137	1.245	0.226	1.081
Factor_12	0.022	0.132	0.169	0.867	1.017
Factor_13	-0.277	0.132	-2.098	0.047	1.011
Factor_14	0.869	0.167	5.189	0	1.627
Factor_15	0.614	0.171	3.596	0.002	1.689
Factor_17	0.001	0.14	0.005	0.996	1.138
Factor_18	187511.818	1024267.808	0.183	0.856	3.49
Factor_19	-158999.113	301946.211	-0.527	0.604	4.96

Table 32: Component Matrix for MLR PCA100 High

Predictors	Factor_1	Factor_2	Factor_6	Factor_7	Factor_13	Factor_14	Factor_15
Residents	0.67	0.047	0.007	0.122	0.026	0.005	0.004
SpecialtyUniquePatients	0.862	0.229	0.013	0.259	0.002	0.099	0.043
CoreUniquePatients	0.739	0.016	0.133	0.193	0.141	0.031	0.023
PhysicianClinicalFTEEper10KSpecialtyUniques	0.393	0.028	0.051	0.36	0.061	0.015	0.054
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.488	0.125	0.187	0.154	0.108	0.029	0.001
SpecialtyUniqueEncounters	0.847	0.303	0.015	0.221	0.106	0.082	0.036
AssociateProvidersperPhysicianClinicalFTEE	0.012	0.91	0.075	0.014	0.008	0.016	0.012
AdminSupportstaffper10kPhysicianClinicalFTEE	0.039	0.879	0.097	0.133	0.013	0.024	0.026
AssociateProvidersper100KCoreFacilityUniques	0.129	0.792	0.021	0.189	0.013	0.006	0.011
AdminSupportStaffperPhysicianClinicalFTEE	0.676	0.368	0.142	0.166	0.016	0	0.008
ClinicalSupportStaffperPhysicianClinicalFTEE	0.575	0.376	0.008	0.029	0.001	0.001	0.002
TotalSupportStaffperPhysicianClinicalFTEE	0.665	0.405	0.061	0.086	0.006	0	0.001
PhysicianClinicalFTEE	0.919	0.215	0.002	0.035	0.039	0.004	0.083
AdjMDFTE	0.89	0.244	0.05	0.2	0.012	0.014	0.02
complex_1a	0.669	0.174	0.27	0.536	0.003	0	0.01
complex_1b	0.07	0.665	0.415	0.29	0.008	0.004	0.004
complex_1c	0.199	0.315	0.081	0.29	0.015	0.001	0.001
complex_2	0.275	0.114	0.738	0.209	0.003	0.002	0.002
complex_3	0.426	0.278	0.503	0.296	0.005	0.005	0.008

4.3.7 MLR-PCA100-Low (M23)

Similar to the MLR-Low (M13) performers, we performed MLR-PCA100-Low. Again, PCA was performed to remove collinearity from the data. The results are shown in Tables 33-36 R-square value shown in Table 34 indicates 0.85 of the variances is explained by the input variables. Table 33 shows that the regression is significant. The significant factors turn out to be Factor_5. Based on the component matrix in Table 36, we list the influential input variables as follows

- Factor_5: Admin Support Staff per Physician Clinical FTEE, complex_3

Table 33: ANOVA Table for M23

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	1.963	18	0.109	3.766	0.012
Residual	0.347	12	0.029		
Total	2.31	30			

Table 34: Regression Statistics for M23

Regression Statistics	
R	0.922
R squared	0.85
Adjusted R squared	0.642
Standard Error	0.171
Observations	30

Table 35: Regression Coefficients for M23

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-1.297	0.031	-42.417	0	
Factor_1	0.075	0.041	1.821	0.094	1.751
Factor_2	-0.093	0.043	-2.146	0.053	1.943
Factor_3	-0.006	0.032	-0.202	0.843	1.061
Factor_4	0.005	0.041	0.129	0.9	1.707
Factor_5	-0.185	0.032	-5.788	0	1.053
Factor_6	0.016	0.037	0.435	0.671	1.397
Factor_7	0.023	0.032	0.722	0.484	1.08
Factor_8	-0.02	0.034	-0.606	0.555	1.181
Factor_9	0.043	0.048	0.888	0.392	2.437
Factor_10	-0.019	0.038	-0.481	0.639	1.536
Factor_11	-0.023	0.035	-0.661	0.521	1.293
Factor_12	0.006	0.05	0.13	0.898	2.55
Factor_13	0.05	0.038	1.315	0.213	1.509
Factor_14	0.004	0.031	0.131	0.898	1.001
Factor_15	-0.043	0.06	-0.722	0.484	3.702
Factor_16	-0.011	0.033	-0.345	0.736	1.119
Factor_17	-0.048	0.07	-0.687	0.505	5.137
Factor_19	-90074.079	200012.08	-0.45	0.66	15.457

Table 36: Component Matrix for M23

Predictors	Factor_5
Residents	0.01
SpecialtyUniquePatients	0.047
CoreUniquePatients	0.182
PhysicianClinicalFTEEper10KSpecialtyUniques	0.147
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.075
SpecialtyUniqueEncounters	0.035
AssociateProvidersperPhysicianClinicalFTEE	0.135
AdminSupportstaffper10kPhysicianClinicalFTEE	0.205
AssociateProvidersper100KCoreFacilityUniques	0.072
AdminSupportStaffperPhysicianClinicalFTEE	0.514
ClinicalSupportStaffperPhysicianClinicalFTEE	0.083
TotalSupportStaffperPhysicianClinicalFTEE	0.116
PhysicianClinicalFTEE	0.043
AdjMDFTE	0.045
complex_1a	0.102
complex_1b	0.339
complex_1c	0.031
complex_2	0.318
complex_3	0.722

4.3.8 PCA-Split Sampling- MLR-PCA100-Split (M24)

Similar to MLR-split (M14), we were interested to check how well our PCA models perform on a new data set and to know how well we can predict an observation in absolute terms. For this, we

split the data in 70% and 30% ratios as Training and Test data sets and have run our regression model on Training data. The results are presented below from tables 37-42. The table 37, 38, and 39 show the regression models build on the training data. Using this regression equation, we try predicting the Productivity values for our Test data set and estimated the errors and calculated the R sq and adjusted R- sq values for the test data sample. From table 38 we can see the R – squared value and adj R – squared value for the test data sample as 0.331 and 0.214, respectively.

Table 37: ANOVA for Training data

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	123.644	18	6.869	12.167	0
Residual	155.816	276	0.565		
Total	279.46	294			

Table 38: Regression Statistics for Training data

Regression Statistics	
R	0.665
R squared	0.442
Adjusted R squared	0.406
Standard Error	0.751
Observations	294

Table 39: Regression Coefficients for Training data

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	0.069	0.044	1.574	0.117	
Factor_1	0.04	0.044	0.904	0.367	1.002
Factor_2	0.144	0.044	3.25	0.001	1.022
Factor_3	-0.21	0.048	-4.38	0	1.192
Factor_4	-0.12	0.044	-2.717	0.007	1.023
Factor_5	-0.014	0.044	-0.306	0.76	1.014
Factor_6	-0.037	0.044	-0.841	0.401	1.026
Factor_7	-0.04	0.044	-0.906	0.366	1.001
Factor_8	-0.43	0.044	-9.67	0	1.029
Factor_9	-0.063	0.046	-1.37	0.172	1.113
Factor_10	0.092	0.044	2.081	0.038	1.014
Factor_11	0.137	0.047	2.908	0.004	1.148
Factor_12	0.211	0.044	4.764	0	1.018
Factor_13	0.21	0.05	4.229	0	1.289
Factor_14	-0.175	0.06	-2.902	0.004	1.885
Factor_15	0.051	0.082	0.618	0.537	3.503
Factor_16	0.083	0.048	1.749	0.081	1.182
Factor_17	-0.106	0.157	-0.675	0.5	12.903
Factor_19	-146405.082	152820	-0.958	0.339	17.365

Table 40: Performance Matrix for Training data

ERROR	ESTIMATE
Mean Absolute Deviation	0.462
Mean Square Error	0.528
Root Mean Square Error	0.727
Mean Absolute Percennt Error	744.140

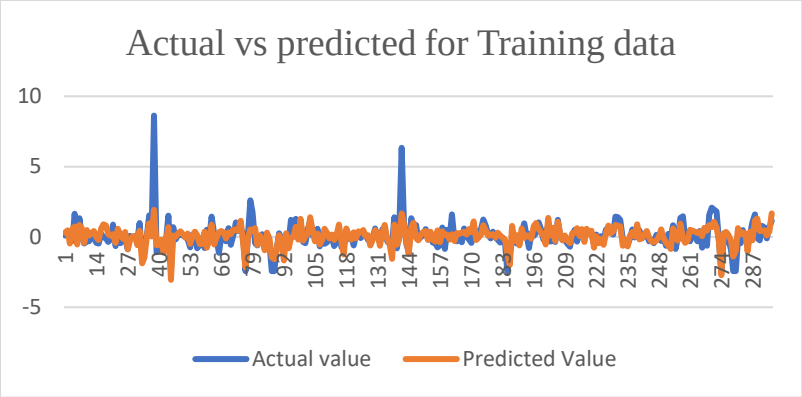


Figure 5:Predicted values Vs Actual values Graph for Training data

Table 41: Error Estimates for Testing data

ERROR	ESTIMATE
Mean Absolute Deviation	0.840
Mean Square Error	1.614
Root Mean Square Error	1.270
Mean Absolute Percennt Error	312.330

Table 42: Regression Statistics for Test data

Regression stastics	
SSE	131.802
SST	196.903
R-squared	0.331
adj R sq	0.214

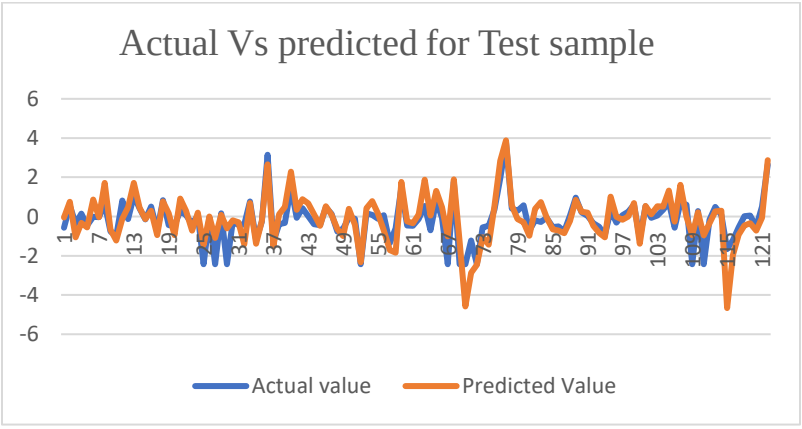


Figure 6: Predicted values Vs Actual values Graph for Testing data

4.4 Other Models

The other models M31-M33 and M41-M42 are performed similarly. For Models M31- M33, we have applied Kaiser's rule (Ledesma & Valero-Mora, 2007) which recommends retaining the factors whose eigenvalue is greater than 1 which explains 80% of variation. Therefore, in these models, we have selected the 7 significant factors that result in 80% of the variation and used them as inputs in an MLR model. For Models M41 and M42, we have used Stepwise regression instead of MLR for PCA80 and PCA100 variations. The results of these models are attached in Appendix D.

4. 5 Summary of Optometry Modeling Results

The summary of Regression values from M11 to M42 for optometry are summarized in Table 35 below.

Table 43: Optometry Results

Model Code	Model Name	R-Squared	Adj.R-Squared
M11	MLR	0.29	0.26
M12	MLR-High	0.85	0.75
M13	MLR-Low	0.95	0.90
M14	MLR-Split	Train: 0.335 Test:0.459	Train: 0.285 0.40
M21	MLR-PCA100	0.30	0.26
M22	MLR-PCA100-High	0.86	0.74
M23	MLR-PCA100-Low	0.85	0.64
M24	MLR-PCA100-Split	Train:0.442 Test: 0.331	Train: 0.406 Test: 0.214
M31	MLR-PCA80	0.10	0.09
M32	MLR-PCA80-High	0.35	0.22
M33	MLR-PCA80-Low	0.57	0.34
M41	Step-PCA80	0.10	0.09
M42	Step-PCA100	0.27	0.26

Non-PCA conclusions:

- Physician Clinical FTEE per 100K Core Facility Unique is a common significant input variable.
- Except for Physician Clinical FTEE per 100K Core Facility Unique MLR high and low did not have any significant variables in common.

The Factors which had a significant impact on productivity for High performers and low performers for MLR models are tabulated below in Table 37.

Table 44:High and low performers

High Performers	Low Performers
Physician Clinical FTEE per 100K Core Facility Uniques	Physician Clinical FTEE per 100K Core Facility Uniques
Physician Clinical FTEE per 10K Specialty Uniques	Specialty Unique Patients
Core Unique Patients	AdjMDFTE
Specialty Unique Encounters	
complex_1a	
complex_1b	
PhysicianClinicalFTEE	

Except for Physician Clinical FTEE per 100K Core specialty Unique, there were no common input variables.

PCA Results:

- Admin Support Staff per Physician Clinical FTEE is the common significant variable for PCA100 models.
- Clinical Support Staff per Physician Clinical FTEE and Total Support Staff per Physician Clinical FTEE are the common significant variable for PCA80 models.

Overall Conclusions:

The most influential input variables for productivity are identified in Table 36 by considering the occurrence of all the input variables based on their number of statistically significant appearances in all models. We can observe that Physician Clinical FTEE, Clinical

Support Staff per Physician Clinical FTEE, Physician Clinical FTEE per 10k Specialty Unique have appeared as significant input variables in 5 different models. Similarly, Adj MDFTEE, Highest complexity level (complex_1a), Lowest complexity level (complex_3) have appeared in 4 different models.

Table 45: Ranking of Input variables for Optometry.

Input Variable	No. of models Appeared	Models Appeared
PhysicianClinicalFTEE	5	M11, M12, M22, M32, M33
ClinicalSupportStaffperPhysicianClinicalFTEE	5	M31, M32, M33, M22, M41
PhysicianClinicalFTEEper10KSpecialtyUniques	5	M11, M12, M21, M33, M42
AdjMDFTE	4	M32, M33, M22, M13
complex_1a	4	M12, M22, M33, M32
complex_3	4	M21, M31, M41, M42

CHAPTER 5: OPHTHALMOLOGY DATA MODELING AND ANALYSIS

In this chapter, we will explain the application of the methodology for the ophthalmology data. After providing a brief introduction to ophthalmology services below, we will proceed with the discussion of the methodological steps for ophthalmology analysis.

Ophthalmology is a branch of science which deals with structure, function, and diseases of the eye. Ophthalmology service is an integral part of the Veterans Health Administration (VHA) health care team (VHA Handbook, 2011). The service consists of over 1200 ophthalmologists who provide a broad spectrum of medical and surgical care to enrolled Veterans at 136 facilities throughout the United States. A Medical Doctor (MD) or Doctor of Osteopathy (DO) who is licensed to independently provide primary, specialty, surgical, and laser eye care services is called an Ophthalmologist. An ophthalmologist is medically prepared and qualified to analyze and treat all eye and visual issues, provide absolute eye care, and analyze general illness of the body. These medicines incorporate, however, are not constrained to cataract surgery, diabetic retinopathy laser treatment, glaucoma treatment, and macular degeneration infusions. After baccalaureate instruction, ophthalmologists complete 4 years of clinical school,1 year of internship, 3 years of an ophthalmology residency including the administration of complex visual conditions and medical procedure, and frequently 1 to 2 years of extra fellowship training in a specific specialty.

5.1 Step 1: Data Collection- Ophthalmology

An aggregated data was collected from all the Ophthalmology centers for the fiscal years 2017, 2018, and 2019. We have a total of 375 data entries for each variable, of which some are missing. All the variables except MCG are continuous data variables.

The data variables used in Optometry and Ophthalmology were similar, so we have followed a similar process like in Chapter 4 to include the MCG categorical variable in our regression model by coding MCG data variable into 5 different data variables with 0 and 1 values, which indicates whether the condition exists or not. The list of all the data variables used for regression models is tabulated below.

Table 46: List of Data Variables- Ophthalmology

S.NO	NAME OF MEASURE	INPUT/OUTPUT	OPHTHALMOLOGY	
			VALID	MISSING
1	Residents	Input	264	111
2	Physician Clinical FTEE	Input	371	4
3	Adj MDFTE	Input	331	44
4	Specialty Unique Patients	Input	372	3
5	Core Unique Patients	Input	375	0
6	Physician Clinical FTEE per 10K Specialty Unique	Input	375	0
7	Physician Clinical FTEE per 100K Core Facility Unique	Input	375	0
8	Specialty Unique Encounters	Input	372	3
9	Associate Providers per Physician Clinical FTEE	Input	153	222
10	Admin Support staff per10k Physician Clinical FTEE	Input	153	222
11	Associate Providers per 100K Core Facility Unique	Input	153	222
12	Admin Support Staff per Physician Clinical FTEE	Input	355	20
13	Clinical Support Staff per Physician Clinical FTEE	Input	354	21
14	Total Support Staff per Physician Clinical FTEE	Input	356	19
15	complex_1a	Input	375	0
16	complex_1b	Input	375	0
17	complex_1c	Input	375	0
18	complex_2	Input	375	0
19	complex_3	Input	375	0
20	Productivity	Output	330	45

The data distribution of Ophthalmology and Optometry variables were very similar. Therefore, we have used the same methods for data preparation for Ophthalmology data as well. Once again, to bring all the variables to the same scale Standardization (or Z-score Normalization) technique was performed, where all the variables were rescaled to make sure that the mean and standard deviation to be 0 and 1 respectively.

5.1.1 Data Cleaning

The missing data values are very high for ophthalmology data sets when compared to optometry data sets. The data variables like residents, Adj MDFTEE, productivity, etc., have

missing data values. We can observe from Table 33, for ophthalmology data Associate Providers per Physician, Clinical FTEE, Admin Support staff per10k Physician Clinical FTEE, Associate Providers per 100K Core Facility Unique out of 375 data values only 153 data values are available, and 222 data values are missing. These values account for more than 50% of the data. Similarly, out of 375 data values for residents, 111 data values are missing. This accounts for almost 30% of data. If we drop these missing data values for our data set, it decreases our sample size. So, we must either completely ignore these variables as a whole or fill in the missing values. As our project aim is to analyze the relation between input and output variables, disregarding these missing data variables will not serve our goal. For residents, Associate Providers per Physician Clinical FTEE, Admin Support staff per10k Physician Clinical FTEE, Associate Providers per 100K Core Facility Unique as zeros. Similar to the optometry case, the missing data values the remaining data variables for ophthalmology accounts for less than 5% of the data. Hence, we have dropped those data values.

5.2 Step 2: Descriptive Data Analysis- Ophthalmology

5.2.1 Graphical Analysis

Similar to the optometry analysis, Histograms and scatter plots were generated for each input and output variable as shown in Appendix C. Similar to Optometry data histograms, could observe data outliers for productivity and other variables. These are not because of data entry errors or other reasons. There is a huge variation in data values as we are considering the data from eye care facilities spread over the entire United States. Hence removing the outliers does not talk completely about the best or worst performing facilities.

Also, individual scatter plots for productivity vs input variables are attached in Appendix C. The graphs show that most of the variables have a very weak relation to productivity.

5.2.2 Descriptive Statistics

For preliminary analysis of data, we perform descriptive statistics. It helps us to identify data mean, skewness, kurtosis, Maximum and minimum values. Table 47 below shows descriptive statistics for Ophthalmology data. The general rule for skewness is that if the skewness value is greater than +1 or less than -1, it is an indication of a skewed distribution. This is true in the case of all the variables except Residents, Core Unique patients, Physician Clinical FTEE per 100K Core Facility Unique, Total Support Staff per Physician Clinical FTEE. Similarly, for kurtosis, if the values are greater than +1 then the distribution is too peaked, if it is less than -1 then the distribution is too flat (Hair et al., 2017). Expect for Core Unique patients, Physician Clinical FTEE per 100K Core Facility Unique, most of the variables in our data have peaked distribution (greater than 1).

Table 47: Descriptive Statistics -Ophthalmology

Descriptive Statistics								
S.No	Input/output variables	N	Range	Minimum	Maximum	Mean	Statistic	Kurtosis
1	Residents	375	4.742	-1.155	3.587	0.000	1.000	-0.018
2	Specialty Unique Patients	375	6.100	-0.983	5.117	0.000	1.000	7.471
3	Core Unique Patients	375	5.064	-1.775	3.289	0.000	1.000	0.380
4	Physician Clinical FTEE per 10K Specialty Uniques	375	19.383	-0.312	19.071	0.000	1.000	356.413
5	Physician Clinical FTEE per 100K Core Facility Uniques	375	5.695	-1.685	4.011	0.000	1.000	0.939
6	Specialty Unique Encounters	375	5.395	-1.155	4.240	0.000	1.000	2.111
7	Associate Providers per Physician Clinical FTEE	375	7.811	-0.418	7.393	0.000	1.000	17.369
8	Admin Support staff per10k Physician Clinical FTEE	375	8.164	-0.362	7.802	0.000	1.000	21.340
9	AssociateProviders per 100K Core Facility Uniques	375	9.450	-0.381	9.068	0.000	1.000	32.650
10	Admin Support Staff per Physician Clinical FTEE	375	8.776	-1.309	7.467	0.000	1.000	11.344
11	Clinical Support Staff per Physician Clinical FTEE	375	10.103	-1.388	8.715	0.000	1.000	18.241
12	Total Support Staff per Physician Clinical FTEE	375	8.840	-1.513	7.327	0.000	1.000	11.480
13	PhysicianClinicalFTEE	375	6.796	-1.263	5.533	0.000	1.000	4.009
14	AdjMDFTE	375	5.524	-1.376	4.147	0.000	1.000	2.095
15	ProductivityMeasure	375	8.958	-3.059	5.899	0.000	1.000	5.458
16	complex_1a	375	1.000	0.000	1.000	0.312	0.464	-1.343
17	complex_1b	375	1.000	0.000	1.000	0.168	0.374	1.186
18	complex_1c	375	1.000	0.000	1.000	0.256	0.437	-0.744
19	complex_2	375	1.000	0.000	1.000	0.139	0.346	2.421
20	complex_3	375	1.000	0.000	1.000	0.117	0.322	3.721

5.2.3 Correlations and T-test

T-test was performed to investigate the significance of the correlation between the dependent variable productivity and the independent variables.

The results shown in Table 48 indicate that |t stat|> t critical value except for Clinical support staff (0.216 < 1.653), hence we reject the Null hypothesis and there is a statistically significant correlation between productivity and the factors considered. From Table 49, we can see that there is a strong correlation between input variables showing high chances of multicollinearity in the data, which is similar to the optometry data correlation results. Multi-collinearity violates the assumptions of MLR, making it difficult to run the regression analysis.

Table 48: One sample t-test

t critical value = 1.653	
Predictors	t value
Residents	8.568
Specialty Unique Patients	19.029
Core Unique Patients	38.707
Physician Clinical FTEE per 10K Specialty Uniques	5.027
Physician Clinical FTEE per 100K Core Facility Uniques	22.002
Specialty Unique Encounters	22.365
Associate Providers per Physician Clinical FTEE	-264.068
Admin Support staff per10k Physician Clinical FTEE	-20.790
AssociateProviders per 100K Core Facility Uniques	-39.993
Admin Support Staff per Physician Clinical FTEE	-31.961
Clinical Support Staff per Physician Clinical FTEE	0.216
Total Support Staff per Physician Clinical FTEE	9.100
PhysicianClinicalFTEE_1	10.688
AdjMDFTE_1	11.091
complex_1a	-55.975
complex_1b	-76.815
complex_1c	-61.905
complex_2	-84.739
complex_3	-92.283

Table 49: Data Correlations

	Residents	Physician Clinical FTEE	Adj MDFTE	Specialty Unique Patients	Core Unique Patients	Physician Clinical FTEE per 10K Specialty Uniques	Physician Clinical FTEE per 100K Core Facility Uniques	Specialty Unique Encounters	Associate Providers per Physician Clinical FTEE	Admin Support staff per10k Physician Clinical FTEE	AssociateProviders per 100K Core Facility Uniques	Admin Support Staff per Physician Clinical FTEE	Clinical Support Staff per Physician Clinical FTEE	Total Support Staff per Physician Clinical FTEE	complex_1a	complex_1b	complex_1c	complex_2	complex_3	ProductivityMeasure_1
Residents	1																			
Physician Clinical FTEE	.591**	1																		
Adj MDFTE	.525**	.627**	1																	
Specialty Unique Patients	-0.089	-0.1	-0.054	1																
Core Unique Patients	.573**	.628**	.194**	-0.097	1															
Physician Clinical FTEE per 10K Specialty Uniques	.713**	.935**	.720**	-0.101	.656**	1														
Physician Clinical FTEE per 100K Core Facility Uniques	.156**	0.091	.232**	-0.018	0.022	.145**	1													
Specialty Unique Encounters	.140**	0.011	.184**	0.005	0.036	0.086	.888**	1												
Associate Providers per Physician Clinical FTEE	.276**	.231**	.265**	-0.014	.225**	.296**	.817**	.854**	1											
Admin Support staff per10k Physician Clinical FTEE	0.096	0.074	.184**	0.095	-0.007	0.071	.185**	.107*	.142**	1										
AssociateProviders per 100K Core Facility Uniques	.102*	.118*	.263**	.424**	-0.04	.145**	0.1	0.029	0.073	.566**	1									
Admin Support Staff per Physician Clinical FTEE	.111*	.115*	.264**	.355**	-0.033	.135**	.141**	0.06	.106*	.781**	.957**	1								
Clinical Support Staff per Physician Clinical FTEE	.581**	.585**	.649**	-0.057	.346**	.683**	.196**	.165**	.292**	0.081	0.082	0.091	1							
Total Support Staff per Physician Clinical FTEE	.160**	0.024	0.06	-0.027	0.098	0.048	-0.081	-0.089	-0.051	-0.032	.109*	0.071	-.303**	1						
complex_1a	-.206**	-.211**	-.190**	-0.036	-.111*	-.240**	-0.034	-0.03	-0.096	0.057	-0.034	-0.005	-.395**	-.264**	1					
complex_1b	-.324**	-.267**	-.323**	.152**	-.183**	-.327**	-0.041	-0.026	-0.092	-0.04	-0.078	-0.073	-.270**	-.180**	-.235**	1				
complex_1c	-.366**	-.276**	-.356**	0.004	-.257**	-.336**	-0.087	-0.057	-.123*	-0.083	-0.081	-0.09	-.246**	-.164**	-.214**	-.146**	1			
complex_2	.645**	.875**	.688**	-0.088	.752**	.915**	.159**	.137**	.342**	0.033	0.056	0.054	.632**	0.045	-.227**	-.287**	-.318**	1		
complex_3	.519**	.809**	.597**	-.103*	.618**	.824**	.159**	.107*	.315**	-0.005	0.005	0.002	.599**	-0.014	-.307**	-.223**	-.191**	.901**	1	
ProductivityMeasure_1	.298**	.174**	.269**	-0.061	-0.053	.260**	.116*	0.029	0.098	.126*	.243**	.228**	.234**	0.041	0.024	-.189**	-.214**	0.07	0.044	1

**. Correlation is significant at the 0.01 level.

*. Correlation is significant at the 0.05 level (2-tailed)

5.3 Step 3: Data Modeling- Ophthalmology

To find the best fit model for our data we have tested the data with different models similar to the optometry case, which are listed below.

Table 50: Overview of Data Modeling

No.	Model Code	Model Name	Regression Type	PCA	Performance	Split Sampling	Data
1	M51	MLR	MLR	No	All	No	Opthalmology
2	M52	MLR-High	MLR	No	High	No	Opthalmology
3	M53	MLR-Low	MLR	No	Low	No	Opthalmology
4	M54	MLR-Split	MLR	No	All	Yes	Opthalmology
5	M61	MLR-PCA100	MLR	100%	All	No	Opthalmology
6	M62	MLR-PCA100-High	MLR	100%	High	No	Opthalmology
7	M63	MLR-PCA100-Low	MLR	100%	Low	No	Opthalmology
8	M64	MLR-PCA100-Split	MLR	100%	All	Yes	Opthalmology
9	M71	MLR-PCA80	MLR	80%	All	No	Opthalmology
10	M72	MLR-PCA80-High	MLR	80%	High	No	Opthalmology
11	M73	MLR-PCA80-Low	MLR	80%	Low	No	Opthalmology
12	M81	Step-PCA80	Step	80%	All	No	Opthalmology
13	M82	Step-PCA100	Step	100%	All	No	Opthalmology

5.3.1 Multiple Linear Regression Analysis -MLR (M51)

ANOVA analysis for the M51 regression model is shown in Table 52. While the Adjusted R squared of 0.292 is not very high (Table 52), as seen in Table 51, the regression is significant at $\propto = 0.05 = 5\%$, in fact with very low Sig. (Significance) values close to zero, indicating close to 100% confidence. Based on Table 53, T-tests show that there is a statistically significant association for productivity and the following variables with $p<0.05$:

- Residents
- Physician Clinical FTEE per 10K specialty Unique
- Physician Clinical FTEE per 100K Core Facility Unique
- Admin Support staff per 10k Physician Clinical FTEE
- Total Support Staff per Physician Clinical FTEE
- Physician Clinical FTEE

Table 51: ANOVA Table for M51

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	131.439	18	7.302	10.717	0.000
Residual	242.561	356	0.681		
Total	374	374			

Table 52: Regression Statistics for M51

Regression Statistics	
R	0.571
R Squared	0.326
Adjusted R Squared	0.292
Standard Error	0.841
observations	374

Table 53: Regression Coefficients for M51

Predictors	Coefficient	Std. Error	t	Sig.	VIF
(Constant)	-1.70E-15	0.043	0	1	
Residents	0.187	0.072	2.58	0.01	2.881
SpecialtyUniquePatients	-0.252	0.138	-1.834	0.068	10.381
CoreUniquePatients	-0.012	0.105	-0.119	0.905	6.066
PhysicianClinicalFTEEper10KSpecialtyUniques	-0.104	0.051	-2.051	0.041	1.411
PhysicianClinicalFTEEper100KCoreFacilityUniques	-0.206	0.104	-1.987	0.048	5.9
SpecialtyUniqueEncounters	1.009	0.185	5.442	0	18.853
AssociateProvidersperPhysicianClinicalFTEE	0.163	0.102	1.606	0.109	5.666
AdminSupportstaffper10kPhysicianClinicalFTEE	-0.277	0.118	-2.352	0.019	7.632
AssociateProvidersper100KCoreFacilityUniques	0.172	0.097	1.776	0.077	5.134
AdminSupportStaffperPhysicianClinicalFTEE	-0.143	0.075	-1.905	0.058	3.116
TotalSupportStaffperPhysicianClinicalFTEE	0.246	0.084	2.924	0.004	3.88
complex_1a	-0.135	0.251	-0.54	0.59	34.586
complex_1b	-0.151	0.198	-0.765	0.445	21.475
complex_1c	-0.111	0.225	-0.492	0.623	27.81
complex_2	-0.194	0.176	-1.105	0.27	16.999
complex_3	-0.239	0.163	-1.463	0.144	14.648
PhysicianClinicalFTEE	-0.695	0.21	-3.308	0.001	24.26
AdjMDFTE	-0.025	0.115	-0.213	0.831	7.273

Similar to optometry data Figure 7 shows a 4-in-1 plot for validating MLR assumptions. The normal plot shows that most of the data points fall on straight-line justifying normality. We can see some outliers as discussed before under the descriptive analysis. A similar observation is made with the histogram, which is fairly symmetric. The scatter plot of residuals vs predicted value

shows no clear pattern in the distribution, hence we can conclude independence and randomness of the errors. The variance is also almost constant. Therefore, the data is homoscedastic. The last assumption in MLR is multicollinearity which can be checked in several ways. Some VIF values shown in Table 44 are more than 10 indicating there is a certain extent of Multicollinearity, which we aim to address using PCA.

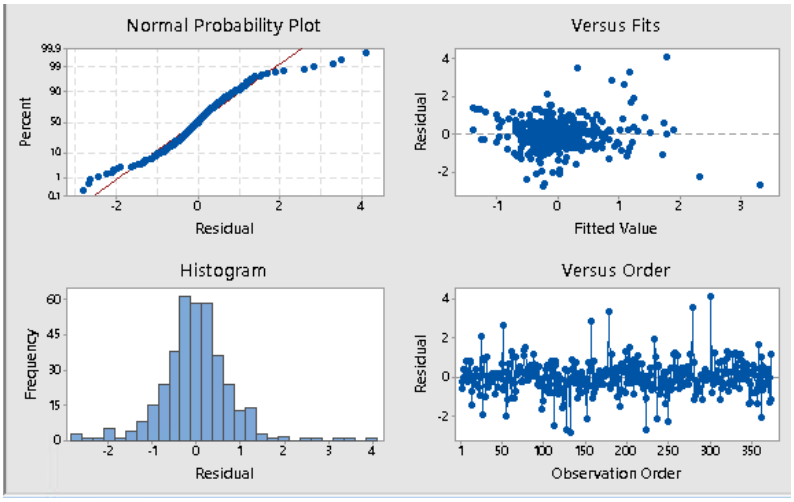


Figure 7: 4-in-graphs for MLR

5.3.2 MLR-High (M52):

As we have seen high-performing models in chapter 4, the M52 model is a high performer analysis for the ophthalmology services. Similar to Optometry data models M12 and M22, we have defined high performance as "performance being more than or equal to 1 standard deviation above the mean". Therefore, to find out which variables contribute to the high-level performance at different VAMCs, we have computed the z-score values of productivity and grouped the data for 1 standard deviation above the mean, which means z-score greater than the value 1.

The results are tabulated in tables 54 to 56. The R-squared value for this model is 0.797. It means 79.7% of the variance is explained by the input variables.

Table 56 shows the most influential input variables significant at $\alpha=0.05$, sig. <0.05 are summarized below:

- Physician Clinical FTEE,
- Physician Clinical FTEE per 100K Core Facility Unique,
- Associate Providers per Physician Clinical FTEE,
- Clinical Support Staff per Physician Clinical FTEE,
- AdjMDFTE_1

Table 54: ANOVA Table for M52

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	35.82	16	2.239	4.917	0.001
Residual	9.106	20	0.455		
Total	44.927	36			

Table 55: Model Summary for M52

Regression Statistics	
R	0.893
R squared	0.797
Adjusted R squared	0.635
Standard Error	0.674
Observations	36

Table 56: Regression Coefficients for M22

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	0.412	0.464	0.888	0.385	
Residents	0.144	0.421	0.342	0.736	12.403
PhysicianClinicalFTEE	3.164	1.375	2.3	0.032	89.086
SpecialtyUniquePatients	-0.884	0.558	-1.585	0.129	26.764
CoreUniquePatients	-0.915	0.514	-1.779	0.09	26.44
PhysicianClinicalFTEEper10KSpecialtyUniques	-5.607	4.606	-1.217	0.238	6.195
PhysicianClinicalFTEEper100KCoreFacilityUniques	-1.093	0.502	-2.176	0.042	11.844
SpecialtyUniqueEncounters	0.654	0.799	0.818	0.423	53.692
AssociateProvidersperPhysicianClinicalFTEE	1.883	0.84	2.243	0.036	47.249
AdminSupportstaffper10kPhysicianClinicalFTEE	-2.066	1.249	-1.654	0.114	41.805
AssociateProvidersper100KCoreFacilityUniques	-0.776	0.652	-1.19	0.248	20.032
AdminSupportStaffperPhysicianClinicalFTEE	-0.241	0.2	-1.21	0.24	3.941
ClinicalSupportStaffperPhysicianClinicalFTEE	0.743	0.2	3.723	0.001	3.014
complex_1b	-0.171	0.215	-0.795	0.436	3.634
complex_1c	-0.186	0.25	-0.743	0.466	4.09
complex_2	-0.295	0.295	-0.999	0.329	6.901
AdjMDFTE_1	-1.978	0.654	-3.023	0.007	20.715

5.3.3 MLR-Low (M53)

Similar to low-performing models, M13 and M23 of optometry, we have computed the same for ophthalmology data in M53. We defined low performance as "performance being less than or equal to 1 standard deviation below the mean". Therefore, to find out which variables contribute to the high-level performance at different VAMCs, we have computed the z-score values of productivity and grouped the data for z-score less than the value -1. The results are shown in Tables 57-59 R-square value shown in Table 59 indicates 79.9% of the variance is explained by the input variables. Table 57 shows that the regression is significant. The significant factors turn out to be:

- Specialty Unique Patients,
- Physician Clinical FTEE per 10K Specialty Unique
- AdjMDFTE_1
- Physician Clinical FTEE

Table 57: ANOVA table for M53

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	8.741	17	0.514	5.393	0
Residual	2.193	23	0.095		
Total	10.934	40			

Table 58: Regression statistics for M53

Regression Statistics	
R	0.894
R squared	0.799
Adjusted R squared	0.651
Standard Error	0.308
Observations	40

Table 59: Regression Coefficients for M53

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-1.671	0.332	-5.031	0	
Residents	-0.059	0.173	-0.339	0.738	15.313
PhysicianClinicalFTEE	2.498	1.159	2.156	0.042	665.615
SpecialtyUniquePatients	-1.932	0.663	-2.916	0.008	171.119
CoreUniquePatients	0.313	0.165	1.898	0.07	5.089
PhysicianClinicalFTEEper10KSpecialtyUniques	-0.725	0.209	-3.466	0.002	1.909
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.226	0.156	1.448	0.161	14.864
SpecialtyUniqueEncounters	-0.136	1.222	-0.112	0.912	460.435
AssociateProvidersperPhysicianClinicalFTEE	0.164	0.155	1.059	0.301	4.455
AdminSupportstaffper10kPhysicianClinicalFTEE	0.006	0.117	0.048	0.962	5.969
AssociateProvidersper100KCoreFacilityUniques	-0.017	0.306	-0.057	0.955	6.173
AdminSupportStaffperPhysicianClinicalFTEE	-0.067	0.087	-0.778	0.444	3.472
TotalSupportStaffperPhysicianClinicalFTEE	0.185	0.096	1.933	0.066	3.072
complex_1a	0.602	0.436	1.382	0.18	25.735
complex_1b	-0.097	0.122	-0.799	0.433	3.992
complex_1c	0.081	0.095	0.86	0.399	3.958
complex_2	-0.076	0.069	-1.111	0.278	3.337
AdjMDFTE_1	-1.24	0.486	-2.552	0.018	93.724

5.3.4 MLR-Split (M54)

We are interested to check how well our models perform on a new data set and to know how well we can predict an observation in absolute terms. A common way to do this is to compare the Errors like Mean Absolute Deviation (MAD, mean of absolute error), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percent Error (MAPE). For this, we randomly split the data into 60 and 40 ratios as Training and Test data sets and have run our regression model on the training data. The results are presented below in tables 60-65. Using this regression equation from training data we try predicting the Productivity values for our Test data set and estimate the errors to compare the training and test data.

Table 60: Model Summary for Training Data

Regression Statistics	
R	0.626
R squared	0.392
Adjusted R squared	0.34
Standard Error	0.805
Observations	229

Table 61: ANOVA table for Training Data

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	88.054	18	4.892	7.543	0
Residual	136.835	211	0.649		
Total	224.89	229			

Table 62: Regression Coefficients for Training Data

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-0.024	0.054	-0.443	0.658	
Residents	0.059	0.088	0.665	0.507	2.96
PhysicianClinicalFTEE	-1.266	0.338	-3.748	0	35.547
SpecialtyUniquePatients	-0.236	0.184	-1.284	0.201	9.282
CoreUniquePatients	0.058	0.147	0.394	0.694	7.702
PhysicianClinicalFTEEper10KSpecialtyUniques	-0.054	0.055	-0.986	0.325	1.723
PhysicianClinicalFTEEper100KCoreFacilityUniques	-0.046	0.155	-0.297	0.767	8.1
SpecialtyUniqueEncounters	1.298	0.235	5.519	0	16.688
AssociateProvidersperPhysicianClinicalFTEE	0.087	0.11	0.79	0.43	3.913
AdminSupportstaffper10kPhysicianClinicalFTEE	-0.183	0.167	-1.093	0.276	6.808
AssociateProvidersper100KCoreFacilityUniques	0.242	0.138	1.751	0.081	4.939
AdminSupportStaffperPhysicianClinicalFTEE	0.019	0.079	0.239	0.812	2.118
ClinicalSupportStaffperPhysicianClinicalFTEE	0.116	0.084	1.375	0.171	2.925
complex_1a	0.009	0.307	0.029	0.977	31.485
complex_1b	-0.063	0.24	-0.264	0.792	21.697
complex_1c	-0.051	0.27	-0.19	0.85	24.349
complex_2	-0.124	0.211	-0.591	0.555	18.103
complex_3	-0.173	0.195	-0.887	0.376	14.314
AdjMDFTE_1	0.123	0.145	0.848	0.397	6.446

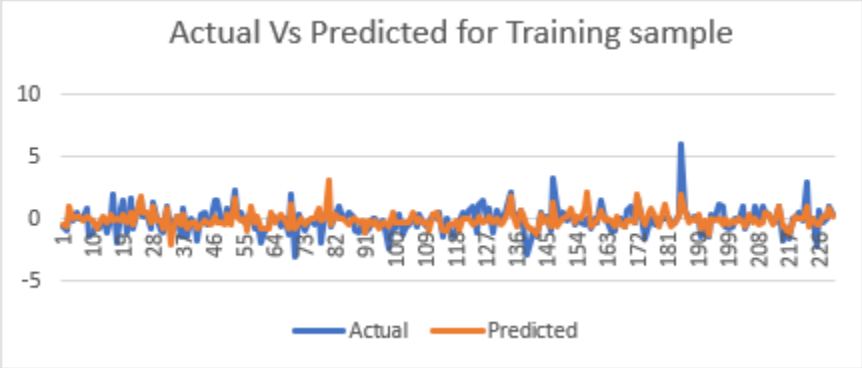


Figure 8: Actual Vs Predicted for Training data

Table 63: Performance Matrix for Training data

ERROR	ESTIMATE
Mean Absolute Deviation	0.552
Mean Square Error	0.595
Root Mean Square Error	0.771

Table 64: Regression Statistics for Test Data

Regression Statistics	
R squared	0.276
Adjusted R squared	0.172
SSE	105.930
SST	146.323
Observations	145

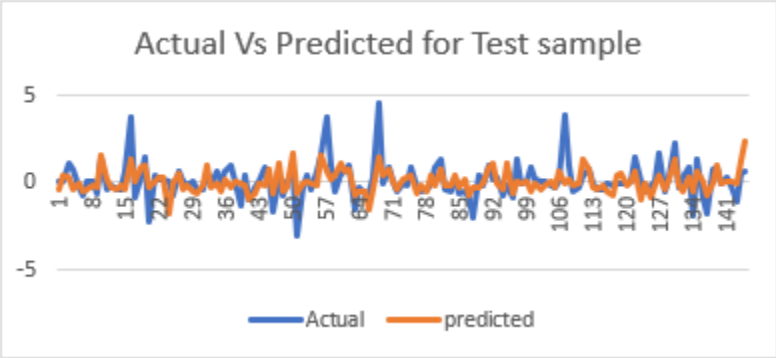


Figure 9 Actual Vs Predicted for Training data

Table 65 Regression Statistics for Test data

ERROR	ESTIMATE
Mean Absolute Deviation	0.587
Mean Square Error	0.731
Root Mean Square Error	0.855

5.3.5 MLR-PCA100 (M61)

.

As indicated before for optometry models, MLR modeling resulted in VIF values of more than 10. This means there is a high possibility of having multicollinearity in the data. To overcome this problem, principal component analysis (PCA) is performed to replace the 19 inter-related

variables with independent components. We have used the Varimax rotation to simplify the factor structure and to make the interpretation easy and retained 100% of the variability in the PCA-MLR model. For this, we choose all 19 PCA components in Model 61 and performed MLR analysis. The results are shown in tables.

Since Sig. ~0 (Table 67), the regression is significant at $\propto = 0.05 = 5\%$. The results in Table 66, show R-squared value is adjusted R-squared are 0.33 and 0.294. From table 51 we can see that the factors 3, 6,7,8,9, 11, 13, 15, and 17 are statistically significant variables with Sig.<0.05. The component matrix for the significant factors is, tabulated below individually and the most influential variables are chosen We have considered all the absolute values of weighted averages above and highlighted values above 0.500. The contributing variables for the significant factors can be summarized as:

- **Factor_3**: Associate Providers per Physician Clinical FTEE, Admin Support staff per 10k Physician Clinical FTEE, Associate Providers per 100K Physician Clinical FTEE, Clinical Support staff per Physician Clinical FTEE
- **Factor_6**: complex_2, complex_3
- **Factor_8**: Physician Clinical FTEE per 10K specialty Unique

The other factors do not have the absolute values of weighted averages above 0.5.

Table 66: Model Summary for M61

Regression Statistics	
R	0.574
R Squared	0.33
Adjusted R Squared	0.294
Standard Error	0.840
observations	374

Table 67: ANOVA for M61

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	132.725	19	6.986	10.278	0.000
Residual	241.275	355	0.68		
Total	374	374			

Table 68: Regression coefficient for M61

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	1.00E-03	0.043	0.018	0.985	
FACTOR_1	0.083	0.044	1.873	0.062	1.068
FACTOR_2	0.068	0.045	1.526	0.128	1.103
FACTOR_3	0.269	0.048	5.65	0	1.248
FACTOR_4	0.016	0.043	0.37	0.712	1.002
FACTOR_5	-0.036	0.044	-0.815	0.416	1.057
FACTOR_6	0.202	0.05	4.053	0	1.368
FACTOR_7	0.214	0.044	4.841	0	1.076
FACTOR_8	-0.094	0.044	-2.13	0.034	1.062
FACTOR_9	0.282	0.052	5.472	0	1.466
FACTOR_10	-0.036	0.045	-0.811	0.418	1.104
FACTOR_11	-0.228	0.043	-5.234	0	1.041
FACTOR_12	-0.005	0.048	-0.105	0.916	1.28
FACTOR_13	-0.113	0.043	-2.639	0.009	1
FACTOR_14	0.083	0.049	1.704	0.089	1.299
FACTOR_15	0.129	0.045	2.845	0.005	1.131
FACTOR_16	-0.016	0.09	-0.176	0.86	4.427
FACTOR_17	0.204	0.045	4.512	0	1.126
FACTOR_18	-0.011	0.044	-0.244	0.807	1.053
FACTOR_19	-345339.24	251021.599	-1.376	0.17	6.91

Table 69: Component Matrix for M61

Component Matrix									
	FACTOR_3	FACTOR_6	FACTOR_7	FACTOR_8	FACTOR_9	FACTOR_11	FACTOR_13	FACTOR_15	FACTOR_17
Residents	0.071	0.027	0.069	0.054	0.491	0.31	0.11	0.11	0.005
SpecialtyUniquePatients	0.168	0.016	0.102	0.074	0.206	0.123	0.06	0.06	0.018
CoreUniquePatients	0.109	0.072	0.397	0.266	0.236	0.095	0.14	0.14	0.05
PhysicianClinicalFTEEper10KSpecialtyUniques	0.26	0.235	0.468	0.556	0.058	0.05	0.017	0.017	0.005
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.059	0.163	0.469	0.289	0.206	0.148	0.163	0.163	0.056
SpecialtyUniqueEncounters	0.132	0.006	0.032	0.005	0.08	0.091	0.055	0.055	0.011
AssociateProvidersperPhysicianClinicalFTEE	0.722	0.016	0.013	0.001	0.056	0.049	0.031	0.031	0.013
AdminSupportstaffper10kPhysicianClinicalFTEE	0.788	0.005	0.039	0.047	0.009	0.014	0.085	0.085	0.011
AssociateProvidersper100KCoreFacilityUniques	0.689	0.002	0.119	0.026	0.001	0.071	0.11	0.11	0.006
AdminSupportStaffperPhysicianClinicalFTEE	0.305	0.043	0.111	0.449	0.077	0.042	0.013	0.013	0
ClinicalSupportStaffperPhysicianClinicalFTEE	0.509	0.045	0.065	0.05	0.046	0.068	0.031	0.031	0.004
TotalSupportStaffperPhysicianClinicalFTEE	0.493	0.049	0.01	0.121	0.008	0.036	0.019	0.019	0.003
complex_1a	0.027	0.073	0.275	0.153	0.28	0.227	0.012	0.012	0.015
complex_1b	0.224	0.121	0.079	0.068	0.092	0.04	0.001	0.001	0.01
complex_1c	0.021	0.199	0.205	0.091	0.111	0.074	0.024	0.024	0.008
complex_2	0.087	0.675	0.093	0.277	0.105	0.144	0.014	0.014	0.007
complex_3	0.083	0.752	0.303	0.133	0.039	0.169	0.041	0.041	0.006
PhysicianClinicalFTEE	0.049	0.047	0.148	0.036	0.14	0.024	0.028	0.028	0.146
AdjMDFTE	0.005	0.225	0.137	0.072	0.251	0.052	0.273	0.273	0.051

5.3.6 MLR-PCA100-High (622)

As seen earlier for optometry high-level performers (M12, M22) and ophthalmology M53, we follow a similar approach and performed M622, a high performer analysis for the ophthalmology services. we have computed the z-score values of productivity and grouped the data for 1 standard deviation above the mean, which means a z-score greater than the value 1.

As there is multicollinearity in the data, we have performed a factor analysis of the grouped data z-score greater than the value 1.

The results are tabulated in tables 70 to 73. The R-squared value for this model is 0.809. It means 80.9% of the variance is explained by the input variables.

Table 70: Model Summary for M27

Regression Statistics	
R	0.899
R squared	0.809
Adjusted R squared	0.638
Standard Error	0.672
Observations	36

Table 71: ANOVA table for M27

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	36.339	17	2.138	4.73	0.001
Residual	8.587	19	0.452		
Total	44.927	36			

Table 72: Regression coefficient for M27

Predictors	Coefficients	Std. Error	t	Sig.
(Constant)	0.687	0.514	1.337	0.197
Factor_ 2	-0.869	0.625	-1.39	0.181
Factor_ 3	-0.543	1.368	-0.397	0.696
Factor_ 4	-0.001	0.293	-0.002	0.998
Factor_ 5	0.037	0.183	0.2	0.843
Factor_ 6	0.696	0.548	1.27	0.219
Factor_ 7	-0.593	1.699	-0.349	0.731
Factor_ 8	-4.325	4.451	-0.972	0.343
Factor_ 9	0.151	0.281	0.536	0.598
Factor_ 10	-0.109	0.211	-0.518	0.611
Factor_ 11	-0.171	0.22	-0.778	0.446
Factor_ 12	0.036	0.169	0.21	0.836
Factor_ 13	-0.55	0.264	-2.088	0.05
Factor_ 14	-0.346	0.301	-1.147	0.266
Factor_ 15	0.794	0.428	1.856	0.079
Factor_ 16	0.198	0.192	1.03	0.316
Factor_ 17	0.429	0.236	1.817	0.085
Factor_ 19	-2164729.237	2019639	-1.072	0.297

Table 73: Component Matrix for M27

Predictors	Factor_13
Residents	0.112
PhysicianClinicalFTEE	0.043
SpecialtyUniquePatients	0.079
CoreUniquePatients	0.12
PhysicianClinicalFTEEper10KSpecialtyUniques	0.018
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.144
SpecialtyUniqueEncounters	0.062
AssociateProvidersperPhysicianClinicalFTEE	0.029
AdminSupportstaffper10kPhysicianClinicalFTEE	0.078
AssociateProvidersper100KCoreFacilityUniques	0.098
AdminSupportStaffperPhysicianClinicalFTEE	0.015
ClinicalSupportStaffperPhysicianClinicalFTEE	0.034
TotalSupportStaffperPhysicianClinicalFTEE	0.02
complex_1a	0.017
complex_1b	0.003
complex_1c	0.027
complex_2	0.011
complex_3	0.038
AdjMDFTE_1	0.294

5.3.7 PCA-Low-Level Performers (M63)

To find out which variables contribute to the low-level performance, we have computed the z-score values of productivity and grouped the data for 1 standard deviation below the mean, which means z-score less than or equal to 1.

As there is multicollinearity in the data, we have performed a factor analysis of the grouped data. The results are tabulated in the table below. The R-squared value for this model is 0.802. From the table, we observe that all the input variables Factor_1, Factor_4, Factor_6, Factor_8, Factor_12 is statistically significant at $\alpha=0.05$.

The most weighted variables for significant factors are:

- Factor_1: Residents, Physician Clinical FTEE, Specialty Unique Patients, Core Unique Patients, Physician Clinical FTEE per 100K Core Facility Unique, Specialty Unique Encounters, Associate Providers per 100K Core Facility Unique, complex_1a, AdjMDFTE_1
- Factor_4: AdminSupportstaffper10kPhysicianClinicalFTEE, complex_1c, complex_2
- Factor_6: complex_1b, complex_1c

Table 74: Model Summary for M27

Regression Statistics	
R	0.895
R squared	0.802
Adjusted R squared	0.639
Standard Error	0.313
Observations	40

Table 75: ANOVA table for M27

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	8.765	18	0.487	4.939	0
Residual	2.169	22	0.099		
Total	10.934	40			

Table 76: Regression coefficient for M27

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-1.613	0.049	-32.885	0	
Factor_1	0.175	0.05	3.515	0.002	1.009
Factor_2	0.059	0.05	1.189	0.247	1.004
Factor_3	0.104	0.052	1.976	0.061	1.116
Factor_4	0.134	0.053	2.548	0.018	1.124
Factor_5	0.065	0.052	1.251	0.224	1.094
Factor_6	-0.124	0.053	-2.332	0.029	1.138
Factor_7	0.106	0.051	2.078	0.05	1.047
Factor_8	-0.243	0.05	-4.87	0	1.013
Factor_9	0.031	0.056	0.559	0.582	1.269
Factor_10	0.036	0.052	0.69	0.498	1.107
Factor_11	-0.095	0.052	-1.834	0.08	1.097
Factor_12	0.175	0.05	3.46	0.002	1.032
Factor_13	0.034	0.05	0.684	0.501	1.026
Factor_14	0.112	0.062	1.804	0.085	1.573
Factor_15	0.035	0.056	0.634	0.533	1.266
Factor_17	-0.085	0.17	-0.499	0.623	11.689
Factor_18	-463037.648	890341	-0.52	0.608	21.288
Factor_19	-5651.143	156668	-0.036	0.972	20.853

Table 77: Component Matrix for M27

Predictors	Factor_1	Factor_4	Factor_6	Factor_8	Factor_12
Residents	0.678	0.222	0.024	0.256	0.006
PhysicianClinicalFTEE	0.969	0.049	0.011	0.069	0.041
SpecialtyUniquePatients	0.96	0.032	0.007	0.053	0.086
CoreUniquePatients	0.508	0.373	0.346	0.002	0.126
PhysicianClinicalFTEEper10KSpecialtyUniques	0.332	0.079	0.321	0.451	0.034
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.852	0.253	0.099	0.07	0.08
SpecialtyUniqueEncounters	0.975	0.018	0.011	0.057	0.068
AssociateProvidersperPhysicianClinicalFTEE	0.031	0.381	0.016	0.084	0.224
AdminSupportstaffper10kPhysicianClinicalFTEE	0.107	0.513	0.014	0.038	0.207
AssociateProvidersper100KCoreFacilityUniques	0.58	0.044	0.023	0.178	0.012
AdminSupportStaffperPhysicianClinicalFTEE	0.011	0.094	0.17	0.427	0.042
ClinicalSupportStaffperPhysicianClinicalFTEE	0.073	0.29	0.032	0.138	0.022
TotalSupportStaffperPhysicianClinicalFTEE	0.059	0.179	0.094	0.073	0.001
complex_1a	0.836	0.053	0.09	0.133	0.043
complex_1b	0.038	0.236	0.759	0.085	0.071
complex_1c	0.011	0.548	0.684	0.011	0.02
complex_2	0.089	0.684	0.002	0.237	0.008
complex_3	0.356	0.051	0.121	0.241	0.034
AdjMDFTE_1	0.954	0.089	0.007	0.076	0.035

5.3.8 PCA-Testing and Validation- MLR-PCA100-TEST (M64)

We are interested to check how well our models perform on a new data set and to know how well we can predict an observation in absolute terms. A common way to do this is to compare the Errors like Mean Absolute Deviation (MAD, mean of absolute error), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percent Error (MAPE). For this, we randomly split the data in 55% and 45% ratios as Training and Test data sets and have run our regression model on the training data. The results are presented below in tables 78-3. Using this regression equation from training data we try predicting the Productivity values for our Test data set and estimate the errors to compare the training and test data.

Table 78: ANOVA for Training data

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	80.647	18	4.48	8.104	0.000
Residual	80.721	146	0.553		
Total	161.368	164			

Table 79: Regression statistics for Training data

Regression Statistics	
R	0.707
R Squared	0.5
Adj. R Squared	0.438
Standard Error	0.743
observations	164

Table 80: Regression Coefficients for Training data

Predictors	Cofficients	Std. Error	t	Sig.
(Constant)	-0.041	0.059	-0.699	0.485
Factor_1	-0.088	0.06	-1.457	0.147
Factor_2	0.131	0.059	2.224	0.028
Factor_3	0.33	0.061	5.405	0
Factor_5	0.074	0.061	1.21	0.228
Factor_6	0.214	0.058	3.664	0
Factor_7	0.305	0.074	4.119	0
Factor_8	-0.386	0.062	-6.259	0
Factor_9	0.195	0.059	3.279	0.001
Factor_10	0.159	0.059	2.69	0.008
Factor_11	-0.108	0.059	-1.828	0.07
Factor_12	-0.041	0.06	-0.682	0.496
Factor_13	-0.001	0.059	-0.021	0.983
Factor_14	0.065	0.062	1.047	0.297
Factor_15	0.239	0.065	3.707	0
Factor_16	-0.188	0.06	-3.125	0.002
Factor_17	0.052	0.06	0.867	0.387
Factor_18	-1508430.029	670671.1	-2.249	0.026
Factor_19	-217036.763	317121.7	-0.684	0.495

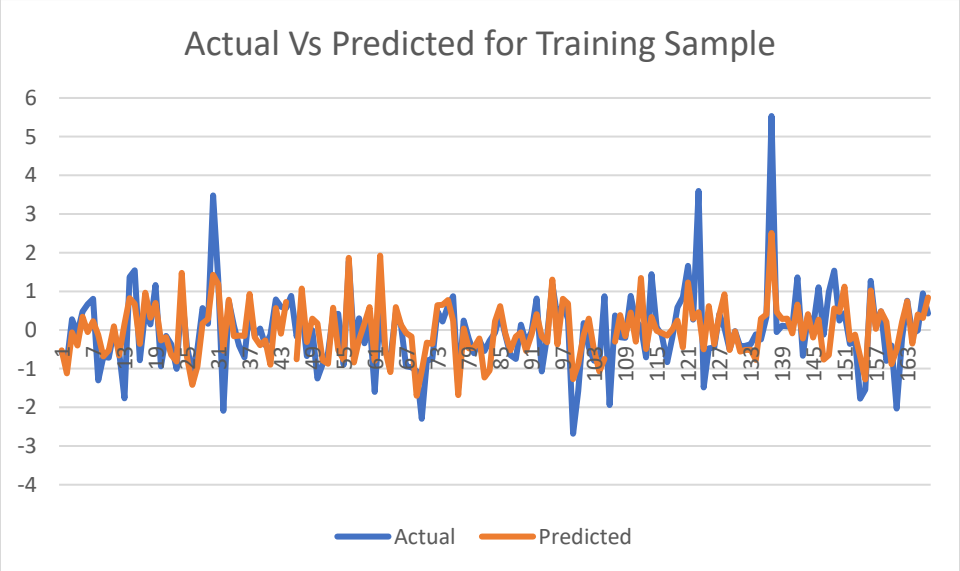


Figure 10: Actual Vs Predicted for Training data

Table 81: Error Estimates for Training data

ERROR	ESTIMATE
Mean Absolute Deviation	0.491
Mean Square Error	0.502
Root Mean Square Error	0.708
Mean Absolute Percennt Error	186.722

Table 82: Error Estimates for Test data

ERROR	ESTIMATE
Mean Absolute Deviation	0.688
Mean Square Error	0.827
Root Mean Square Error	0.910
Mean Absolute Percennt Error	208.346

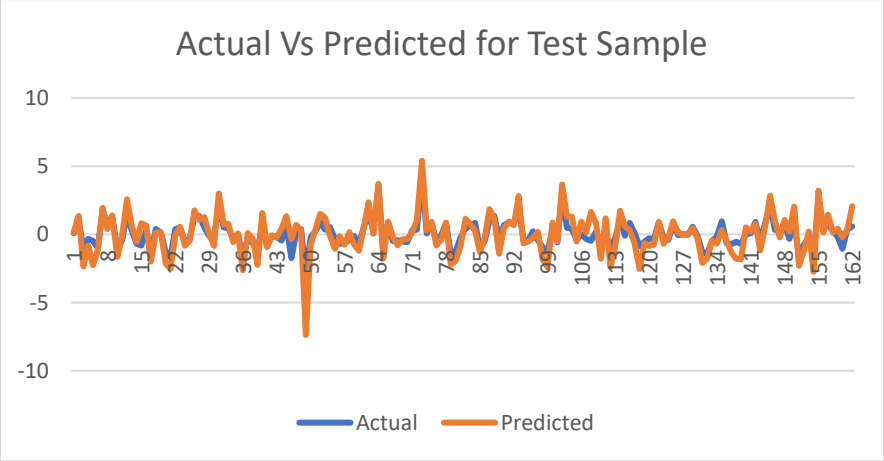


Figure 11 Actual Vs Predicted for Training data

Table 83 Regression Statistics for Test data

Regression Statistics	
R squared	0.135
Adjusted R squared	0.026
SSE	134.040
SST	155.025
Observations	162

5.4 Other Models

The other models M71-M73 and M81-M82 are performed similarly. For Models M71- M73, we have applied Kaiser's rule (Ledesma & Valero-Mora, 2007) which recommends retaining the factors whose eigenvalue is greater than 1 which explains 80% of variation. Therefore, in these models, we have selected the 7 significant factors that result in 80% of the variation and used them as inputs in an MLR model. For Models M71 and M72, we have used Stepwise regression instead of MLR for PCA80 and PCA100 variations. The results of these models are attached in Appendix D.

5.5. Summary of the Ophthalmology Modeling Results

Out of the 13 ophthalmology models, M51 and M61 seem to have similar R-squared values of 0.33 and Adj. R-squared value approximately 0.29 (M52, M53, and M62, M63). We can observe the highest R squared and Adj. R-squared for High and Low performing Models. But in these cases, only high and low performing data points were considered.

Table 84: Ophthalmology results

Model Code	Model Name	R-Squared	Adj.R-Squared
M51	MLR	0.32	0.29
M52	MLR-High	0.80	0.64
M53	MLR-Low	0.80	0.65
M54	MLR-Split	Train: 0.392 Test: 0.276	Train: 0.340 Test: 0.172
M61	MLR-PCA100	0.33	0.29
M62	MLR-PCA100-High	0.81	0.64
M63	MLR-PCA100-Low	0.80	0.64
M64	MLR-PCA100-Split	Train:0.488 Test:0.135	Train: 0.438 Test: 0.0264
M71	MLR-PCA80	0.11	0.10
M72	MLR-PCA80-High	0.41	0.29
M73	MLR-PCA80-Low	0.34	0.23
M81	Step-PCA80	0.10	0.09
M82	Step-PCA100	0.36	0.34

Non-PCA Results:

- Physician Clinical FTEE is a common significant input variable.
- Except for Physician Clinical FTEE, MLR high and low did not have any significant variables in common.
- To estimate what input factors, contribute to the High and low performance. We have tabulated the significant variables from models M26 and M27 below in Table 69.

Table 85: High and low performers

High Performers	Low Performers
Physician Clinical FTEE	Physician Clinical FTEE
Physician Clinical FTEE per 100K Core Facility Unique	Specialty Unique Patients
Associate Providers per Physician Clinical FTEE	Physician Clinical FTEE per 10K Specialty Unique
Clinical Support Staff per Physician Clinical FTEE	
AdjMDFTE	

PCA Results:

- There was no common significant input variable for PCA100 and PCA80 models.

For finding the most influential factors for productivity for ophthalmology services, we have ranked the variables based on their number of significant appearances across all the models and listed them in Table 68. We can observe that Residents, Physician Clinical FTEE, Physician Clinical FTEE per 10K specialty Unique, Clinical Support staff per Physician Clinical FTEE, are the most important influential factors for productivity for the ophthalmology data which were occurred in most of the models.

Table 86 Ranking of Input variables for Ophthalmology.

Input Variable	No. of models	Models Appeared
Residents	6	M51,M71,M63, M73, M81, M82
PhysicianClinicalFTEE	6	M23,M22,M21,M63, M53, M82
ClinicalSupportStaffperPhysicianClinicalFTEE	6	M51, M61, M42, M62, M81, M82
PhysicianClinicalFTEEper10KSpecialtyUniques	5	M41, M51, M53, M62, M82

CHAPTER6: SUMMARY AND CONCLUSIONS

Eyecare is primary care for Veterans Health Affairs, and its demand is increasing nationwide. With good clinical practice guidelines being adopted in more and more facilities, some important changes in clinical efficiency or productivity can be expected. In this study, we have focused mainly on the clinical productivity of optometry and ophthalmology services and investigated the main factors affecting their performance. For this purpose, we build different multiple regression models to test the sensitivity and robustness of results. The main conclusions are summarized below:

The most influential factors that affect eye care clinical productivity

- Physician Clinical FTEE is the most influential Input variable for both Optometry and Ophthalmology.

High performing compared to the Low performing:

- Except for Physician Clinical FTEE per 100K Core Facility Unique, high, and low performers of Optometry did not have any other common input variable.
- Except for Physician Clinical FTEE, High and low performers of Ophthalmology did not have any other common input variable.

Comparison for Optometry and Ophthalmology services:

- Physician Clinical FTEE
- Physician Clinical FTEE per 10K Specialty Unique.
- Clinical Support Staff per Physician Clinical FTEE

Were the common most significant input variables observed in both Optometry and Ophthalmology data.

6.1 Limitations of Research

While the results of the study are encouraging, there are some limitations to future work. The subject of study was a wide data range of productivity and staff across all the facilities in the US. This study must be taken care of, as there is a huge variation in the data. On the other hand, the results are generalized as the study was conducted across 700 facilities (approximate data for both optometry and ophthalmology in 1 year), and to a specific facility, the results may vary. But the methodology used in this study can be applied to a specific facility to observe the findings. Moreover, the data available was for a period of three years (2017-2019). It would have been a better sample if we had the data for at least 5 years.

6.2 Future Research ideas and directions

The following research ideas and directions can be suggested:

- Future research can address other research questions like how the same input variables can impact the output variables like cost, patient satisfaction, etc.
- In the literature review, we have reviewed a few papers which mentioned wait lines, and the number of beds, hospital size. More data can be collected on these variables which can help us to find their impact on productivity.
- Also, we can run different regression models assuming the data variables to be quadratic or polynomial data or using logistic or Probit regression methods. This can be an interesting alternative analysis method to test the data in future analysis.

REFERENCES

Allen, M. (2017). *The sage encyclopedia of communication research methods* (Vols. 1-4). Thousand Oaks, CA: SAGE Publications, Inc doi: 10.4135/9781483381411

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression diagnostics: Identifying influential data and sources of collinearity. New York, NY: John Wiley & Sons.

Bevans, R. (2020). Multiple Linear Regression: A Quick and Simple Guide. Retrieved January 27, 2021, from https://www.scribbr.com/statistics/multiple-linear-regression/8

Budzi, D., Lurie, S., Singh, K., & Hooker, R. (2010). Veterans perceptions of care by nurse practitioners, physician assistants, and physicians: A comparison from satisfaction surveys. Journal of the American Academy of Nurse Practitioners, 22(3), 170–176. DOI: 10.1111/j.1745-7599.2010. 00489.x

Chatterjee, S., & Hadi, A. S. (2012). Regression analysis by example (5th ed.). Hoboken, NJ: John Wiley & Sons.

Coleman, D. L., Moran, E., Serfilippi, D., Mulinski, P., Rosenthal, R., Gordon, B., & Mogielnicki, R. P. (2003). Measuring physicians' productivity in a Veterans' Affairs Medical Center, DOI: 10.1097/00001888-200307000-00007

Evans, R. H. (1996). An Analysis of Criterion Variable Reliability in Conjoint Analysis. Perceptual and Motor Skills, 82(3), 988–990. doi: 10.2466/PMS.1996.82.3.988

Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data, 1888. Readings in the History of Psychology. 336–346. doi: 10.1037/11304-039

Gilman SC, Aron DC, Cannon GW, Chang BK, Godleski L, Golden RM, Henley SS, Holland GJ, Kaminetzky CP, (2010). Measuring the Intensity of Resident Supervision in the Department of Veterans Affairs: The Resident Supervision Index. (2010). doi: 10.1097/ACM.0b013e3181d5a954

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates

Hao, S. H.-S., & Pegels, C. C. (1994). Evaluating relative efficiencies of Veterans Affairs medical centers using data envelopment, ratio, and multiple

regression analysis. Journal of Medical Systems, 1994,18(2), 55–67. DOI: 10.1007/bf00999452

Hintze, J. L., Dr. (2007). Chapter 311 stepwise regression - ncss-wpengine.netdna-ssl.com. Retrieved February 10, 2021, from https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Stepwise_Regression.pdf.

Jaadi, Z. (2019). A Step-by-Step Explanation of Principal Component Analysis. Retrieved January 26, 2021, from https://builtin.com/data-science/step-step-explanation-principal-component-analysis

Johnson, T., Shah, M., Rechner, J., & King, G. (2008). Evaluating the effect of resident involvement on physician productivity in an academic general internal medicine practice. doi: 10.1097/ACM.0b013e3181782c68.

Kao, C., Pang, R., Liu, S., & Bai, X. (2020). Most productive types of hospitals: An empirical analysis. Retrieved December 31, 2020, from https://www.sciencedirect.com/science/article/pii/S0305048320306642 https://doi.org/10.1016/j.omega.2020.102310

Kenton, W. (2020). How Multiple Linear Regression Works. Retrieved January 07, 2021, from https://www.investopedia.com/terms/m/mlr.asp

Letvak S & Buck R (26 June 2008). Factors influencing work productivity and intent to stay in nursing. Retrieved January 26, 2021, from https://pubmed.ncbi.nlm.nih.gov/18616053/

Lynch, M. G., Maa, A. C., Delaune, W. undefined, Chasan, J. undefined, & Cockerham, G. undefined. (2017). Eye Care Productivity and Access in the Veterans Affairs Health Care System. doi: doi.org/10.7205/MILMED-D-16-00103

A., L. R., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken: Wiley. doi:10.1002/9781119013563

Meghanathan, Natarajan. (2016). Assortativity Analysis of Real-World Network Graphs based on Centrality Metrics. Computer and Information Science. 9. 7. 10.5539/cis.v9n3p7.

Maesschalck, R. D., Estienne F., & Verdú-Andrés J. (1999). THE DEVELOPMENT OF CALIBRATION MODELS FOR SPECTROSCOPIC ... Retrieved from https://fabi.research.vub.be/sites/default/files/atoms/files/PCR.pdf

Medin, E., Häkkinen, U., Linna, M., Anthun, K. S., Kittelsen, S. A., & Rehnberg, C. (2013). International hospital productivity comparison: Experiences from the Nordic countries. Health Policy, 112(1-2), 80–87. doi: 10.1016/j.healthpol.2013.02.004

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2001). Introduction to linear regression analysis (3rd ed.). New York, NY: John Wiley & Sons.

Mosadeghrad, A. (2014). Factors Influencing Healthcare Service Quality. International Journal of Health Policy and Management, 3(2), 77-89. doi: 10.15171/ijhpm.2014.65

Myers, R. H. (1990). Regression analysis. In *Classical and modern regression with applications* (2nd ed.). Boston (Mass.). doi: https://lib.ugent.be/catalog/rug01:000851135

Nobakht, S., Jahangiri, K., Hajinabi, K. (2018). Correlation of Performance Indicators and Productivity of Emergency Departments. *Trauma Monthly*, 23(5), doi: 10.5812/traumamon.58686

Petersen, H. (2019). Veterans' eye care services – almost 2 million in 2018. Retrieved from https://www.blogs.va.gov/VAntage/55784/veterans-eye-care-services-almost-2-million-2018/ Accessed April 25, 2020.

Robinson, R. D., Dib, S., Mclarty, D., Shaikh, S., Cheeti, R., Zhou, Y. . . . Wang, H. (2020). Productivity, efficiency, and overall performance comparisons between attendings working solo versus attendings working with residents staffing models in an emergency department: A Large-Scale Retrospective Observational Study. Plos One, 15(2). doi: 10.1371/journal.pone.0228719

Smith, D. M., Martin, D. K., Langefeld, C. D., Miller, M. E., & Freedman, J. A. (1995). Primary Care Physician Productivity" The Physician Factor. doi: 10.1007/bf02602400

TP Usherwood. (1987). Clinical efficiency in general practice. Retrieved February 2021, from https://pubmed.ncbi.nlm.nih.gov/3609555/

Upreti, J. (2017, July 16). Principal Component Analysis - Case Study. Retrieved February 10, 2021, from https://rpubs.com/jainendraupreti/principal-component-analysis

Veterans' Health Administration DIRECTIVE. (2019, October 2). Retrieved April 05, 2021, from https://www.va.gov/OPTOMETRY/docs/1121_D_2019-10-02.pdf.

VHA Handbook 1605.01 Productivity and Staffing Guidance for Specialty Provider Group Practice. (2015, May 4). Retrieved April 05, 2021, from https://www.navao.org/wp-content/uploads/2016/02/VHA-Handbook-1065.01-Productivity-and-Staffing-Guidance-for-Specialty-Provider-Group-Practice.pdf.

VHA Handbook 1121.01. (2011, March 11). Retrieved July 15, 2020, from https://www.charlesmullen.com/wp/wp-content/uploads/publications/VHA-Handbook.pdf

Xu, S., Lu, B., Baldea, M., Edgar, T. F., Wojsznis, W., Blevins, T., & Nixon, M. (2015). Data cleaning is the process industry. Reviews in Chemical Engineering, 31(5). doi: 10.1515/revise-2015-0022

Williams, M. N., Grajales, C. A., & Kurkiewicz, D. (2013). Assumptions of Multiple Regression: Correcting Two Misconceptions. Retrieved from https://www.researchgate.net/publication/256980936_Assumptions_of_Multiple_Regression_Correcting_Two_Misconceptions https://doi.org/10.7275/55hn-wk47

Yauheniya, V., Carl Rudolf, B., Aleksandra, T., & Jonas, S. (2017). Comparing the Efficiency of Hospitals in Italy and Germany: Nonparametric Conditional Approach Based on Partial Frontier. doi: 10.1007/s10729-016-9359-1

Yaremko, R. M., Harari, H., Harrison, R. C., & Lynn, E. (1986). Handbook of research and quantitative methods in psychology: For students and professionals. Hillsdale, NJ: Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203767740

APPENDIX A: DATA COLLECTION SAMPLES

1. OPTOMETRY DATA SAMPLE

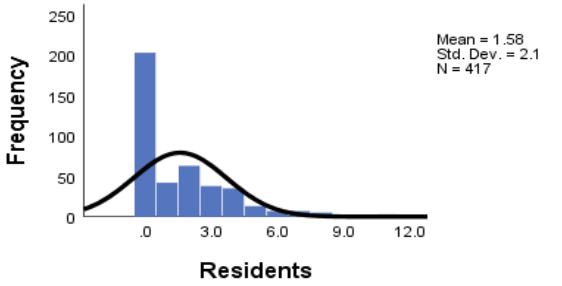
Year	Facility	MCG	Residents	Physician Clinical FTEE	Adj MDFTE	Specialty Unique Patients	Core Unique Patients	Physician Clinical FTEE per 10K Specialty Uniques	Physician Clinical FTEE per 100K Core Facility Uniques	Productivity Measure	Specialty Unique Encounters	Associate Providers per Physician Clinical FTEE	Admin Support staff per10k Physician Clinical FTEE	AssociateProviders per 100K Core Facility Uniques	Support Staff per Physician Clinical FTEE	Support Staff per Physician Clinical FTEE	Support Staff per Physician Clinical FTEE
2017	(1V01) (402) Togus, ME HCS	1c	2.00	5.31	5.20	11,991	42,120	4.43	12.61	5,504	20,324	0.00	0.01	0.02	0.17	1.50	1.68
2017	(1V01) (405) White River Junction, VT HCS	2	3.00	3.94	3.84	8,754	26,131	4.50	15.08	5,550	12,423				0.76	1.32	2.08
2017	(1V01) (518) Bedford, MA HCS	3	1.00	2.91	2.91	5,006	19,566	5.82	14.89	5,221	12,419				0.27		0.27
2017	(1V01) (523) Boston, MA HCS	1a	11.00	9.98	9.89	17,624	62,570	5.66	15.95	4,877	30,690				0.53	0.43	0.96
2017	(1V01) (608) Manchester, NH HCS	3	2.00	2.85	2.85	7,745	25,870	3.68	11.03	7,986	15,958	0.00	0.01	0.04	0.45	0.74	1.19
2017	(1V01) (631) Central Western Massachusetts HCS	3		4.94	4.90	8,349	26,550	5.92	18.60	4,861	14,851				0.24	1.09	1.33
2017	(1V01) (650) Providence, RI HCS	1c	3.00	4.67	4.67	11,950	36,207	3.90	12.89	7,484	19,854				0.46	0.69	1.15
2017	(1V01) (689) Connecticut HCS	1a	8.00	10.63	10.33	19,156	57,875	5.55	18.37	5,231	42,713				0.38	0.62	1.00
2017	(1V02) (526) Bronx, NY HCS	1b		1.92	1.92	4,074	26,030	4.71	7.36	4,589	5,117				0.27	1.23	1.50
2017	(1V02) (528) Western New York HCS	1b		2.41	2.41	7,758	44,755	3.11	5.39	4,827	9,898	0.00	0.01	0.02	0.47	1.65	2.11
2017	(1V02) (528A5) Canandaigua, NY HCS	3		1.82	1.82	5,083	20,121	3.58	9.06	4,833	6,573				0.15	1.06	1.21
2017	(1V02) (528A6) Bath, NY HCS	2		1.70	1.70	4,411	13,619	3.84	12.45	5,619	5,296	0.11	0.41	1.32	0.96	1.42	2.38
2017	(1V02) (528A7) Syracuse, NY HCS	1c		4.10	3.42	10,522	44,229	3.90	9.28	5,202	13,310	0.05	0.18	0.43	0.71	1.40	2.10
2017	(1V02) (528A8) Albany, NY HCS	1c		2.01	2.01	4,804	31,333	4.18	6.41	4,634	5,617	0.03	0.12	0.19	0.13	0.99	1.12
2017	(1V02) (561) New Jersey HCS	1c	3.00	8.89	8.51	16,879	56,101	5.27	15.84	4,417	24,115	0.09	0.45	1.35	0.98	0.68	1.66
2017	(1V02) (620) Hudson Valley, NY HCS	3	4.00	4.91	4.85	10,496	23,945	4.67	20.49	6,234	20,111				0.58	0.12	0.70
2017	(1V02) (630) New York Harbor HCS	1a	4.00	6.08	6.08	12,423	47,966	4.90	12.68	4,562	18,288				0.08	0.34	0.42
2017	(1V02) (632) Northport, NY HCS	1c	4.00	3.40	3.12	7,749	30,587	4.39	11.12	5,691	11,535	0.00	0.01	0.03	0.37	0.47	0.84
2017	(1V04) (460) Wilmington, DE HCS	2	3.00	5.46	5.42	7,676	30,449	7.11	17.92	4,757	10,994	0.01	0.10	0.26	0.23	0.67	0.91
2017	(1V04) (503) Altoona, PA HCS	3		1.94	1.89	7,791	26,230	2.48	7.38	5,290	12,025	0.28	0.71	2.10	1.45	2.56	4.01
2017	(1V04) (529) Butler, PA HCS	3		0.73	0.73	2,430	22,435	3.02	3.27	5,150	2,544	0.01	0.04	0.04	1.12	2.40	3.53
2017	(1V04) (542) Coatesville, PA HCS	3		2.15	2.15	5,342	19,427	4.02	11.06	5,555	6,775				0.00		0.00

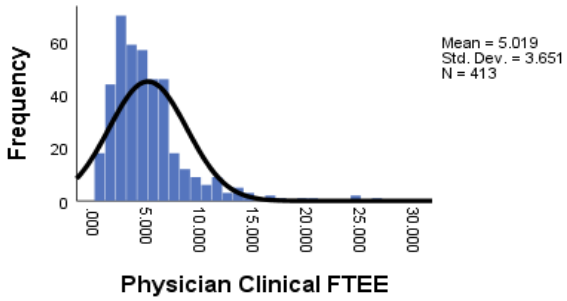
2. OPHTHALMOLOGY DATA SAMPLE

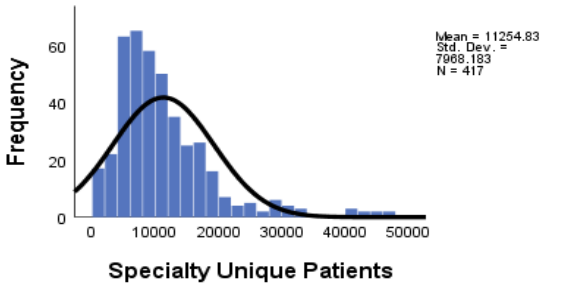
Year	Facility	MCG	Residents	Physician Clinical FTEE	Adj MDFTE	Specialty Unique Patients	Core Unique Patients	n Clinical FTEE per 10K Specialty Uniques	n Clinical FTEE per 100K Core Facility Uniques	Productivity Measure	Specialty Unique Encounters	Associate Providers per Physician Clinical FTEE	Admin Support staff per10k Physician Clinical FTEE	AssociateProviders per 100K Core Facility Uniques	Admin Support Staff per Physician Clinical FTEE	Clinical Support Staff per Physician Clinical FTEE	Total Support Staff per Physician Clinical FTEE
2017	(1V01) (402) Togus, ME HCS	1c	2.00	2.39	2.25	2,094	42,120	11.42	5.68	6,439	6,252				0.33	1.20	1.52
2017	(1V01) (405) White River Junction, VT HCS	2		1.03	0.91	586	26,131	17.59	3.95	5,809	2,186				0.09	1.16	1.25
2017	(1V01) (518) Bedford, MA HCS	3		0.12		437	19,566	2.75	0.61		735				0.18	0.47	0.65
2017	(1V01) (523) Boston, MA HCS	1a	8.00	5.11	4.14	4,312	62,570	11.85	8.17	8,667	18,305	0.55	6.49	4.47	1.52	1.87	3.39
2017	(1V01) (650) Providence, RI HCS	1c	2.75	2.43	2.41	2,197	36,207	11.07	6.72	7,747	9,285				0.89	1.76	2.65
2017	(1V01) (689) Connecticut HCS	1a	6.00	2.94	2.79	3,929	57,875	7.48	5.08	8,155	13,658				0.95	1.21	2.16
2017	(1V02) (526) Bronx, NY HCS	1b	5.00	2.27	2.20	4,032	26,030	5.62	8.71	10,524	12,097				1.07	2.16	3.23
2017	(1V02) (528) Western New York HCS	1b	3.50	2.33	2.32	4,736	44,755	4.92	5.20	9,169	11,086				0.94	2.37	3.31
2017	(1V02) (528A5) Canandaigua, NY HCS	3	0.50	0.32		1,174	20,121	2.70	1.58		2,104				0.30	4.76	5.06
2017	(1V02) (528A6) Bath, NY HCS	2		0.16		664	13,619	2.34	1.14		977				0.44	0.56	1.00
2017	(1V02) (528A7) Syracuse, NY HCS	1c	3.00	3.05	1.37	3,617	44,229	8.44	6.90	9,664	10,309				0.68	1.72	2.40
2017	(1V02) (528A8) Albany, NY HCS	1c	4.50	2.18	1.40	5,118	31,333	4.26	6.96	10,022	11,128				0.73	1.59	2.32
2017	(1V02) (561) New Jersey HCS	1c	4.00	1.54	0.59	2,389	56,101	6.45	2.75	4,594	5,863				1.50	1.00	2.50
2017	(1V02) (630) New York Harbor HCS	1a	10.00	4.72	4.25	7,611	47,966	6.20	9.84	5,415	17,001	0.05	0.33	0.52	0.38	0.73	1.11
2017	(1V02) (632) Northport, NY HCS	1c	3.00	2.12	2.09	2,683	30,587	7.90	6.93	7,109	8,686				0.62	0.94	1.56
2017	(1V04) (460) Wilmington, DE HCS	2	2.00	2.17	1.60	1,179	30,449	18.38	7.12	6,737	5,005				0.98	0.96	1.94
2017	(1V04) (523) Butler, PA HCS	3		0.83	0.83	2,017	22,435	4.13	3.71	6,176	3,181	0.01	0.05	0.04	0.68	1.26	1.95
2017	(1V04) (562) Erie, PA HCS	3		0.00		2	21,478	0.00	0.00		2						
2017	(1V04) (595) Lebanon, PA HCS	2	3.50	1.62		1,815	44,057	8.94	3.68		6,235				0.58	0.95	1.52
2017	(1V04) (642) Philadelphia, PA HCS	1b	4.00	4.01	3.82	4,602	56,547	8.71	7.09	6,110	13,671	0.00	0.02	0.02	1.00	0.63	1.62
2017	(1V04) (646) Pittsburgh, PA HCS	1a	3.00	3.32	2.95	6,473	71,403	5.13	4.65	7,891	14,296				1.23	1.37	2.60

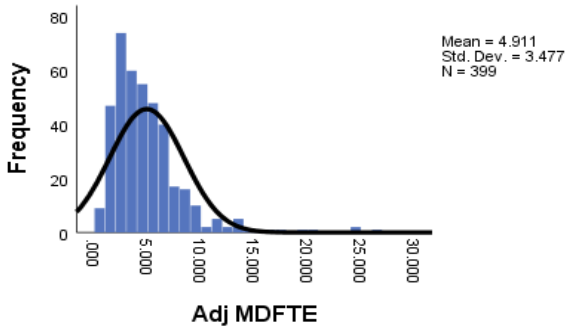
APPENDIX B: INDIVIDUAL GRAPHICAL SUMMARIES

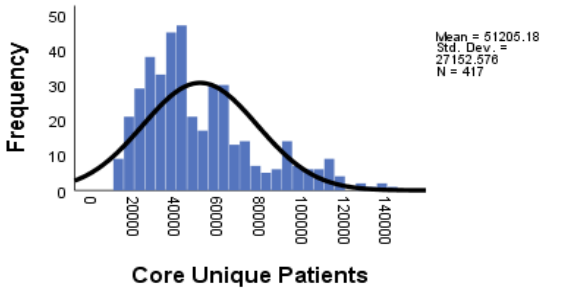
1. FOR OPTOMETRY DATA

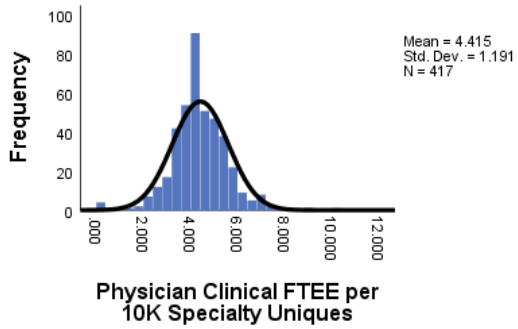


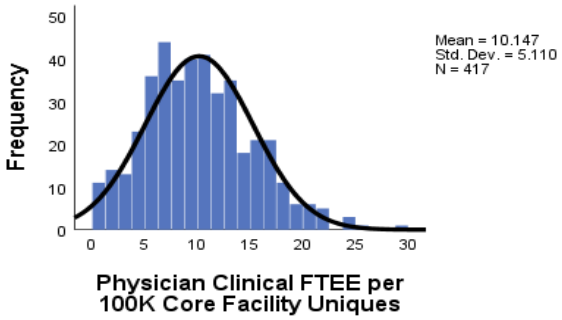


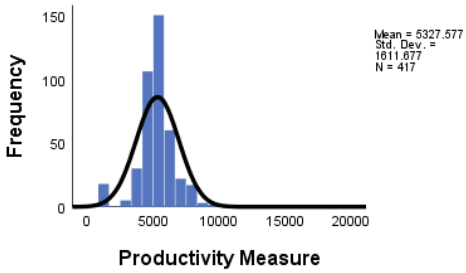


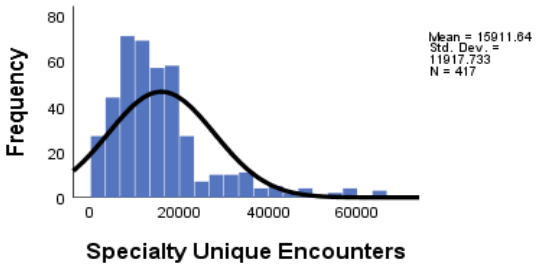


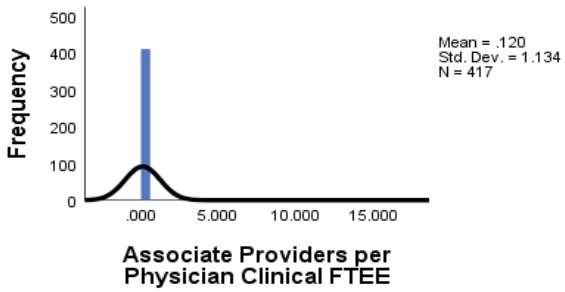


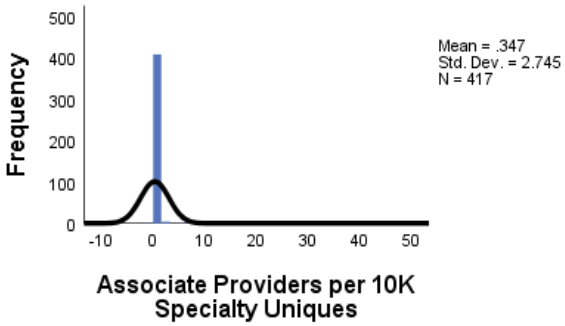


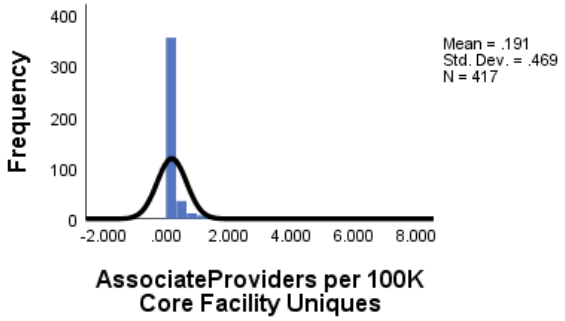


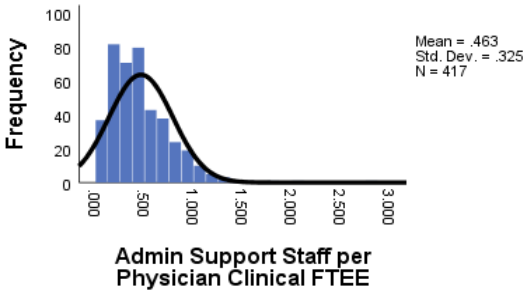


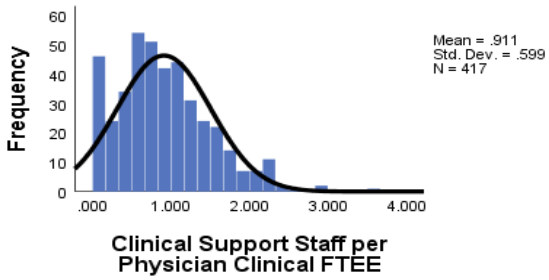

Frequency
Mean = 15911.64
Std. Dev. = 11917.733
N = 417
Specialty Unique Encounters

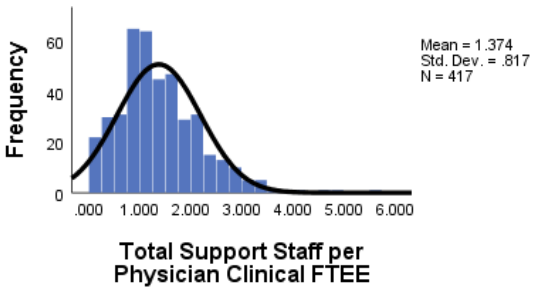

Frequency
Mean = .120
Std. Dev. = 1.134
N = 417
Associate Providers per Physician Clinical FTEE

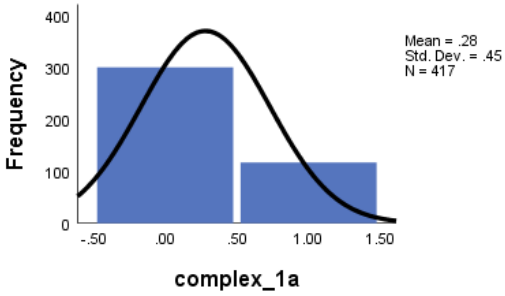

Frequency
Mean = .347
Std. Dev. = 2.745
N = 417
Associate Providers per 10K Specialty Uniques

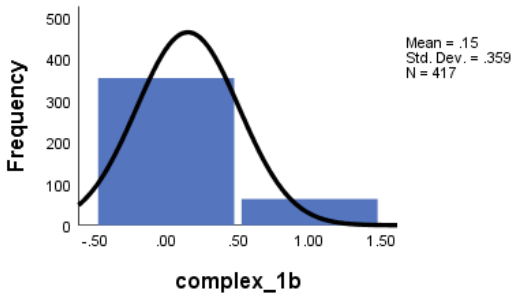

Frequency
Mean = .191
Std. Dev. = .469
N = 417
AssociateProviders per 100K Core Facility Uniques

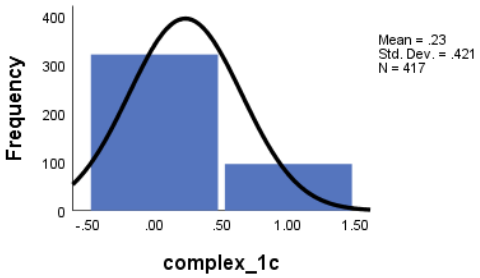

Frequency
Mean = .463
Std. Dev. = .325
N = 417
Admin Support Staff per Physician Clinical FTEE

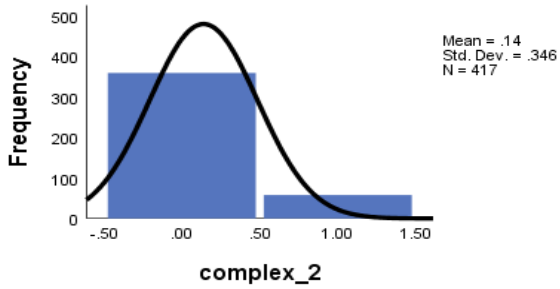

Frequency
Mean = .911
Std. Dev. = .599
N = 417
Clinical Support Staff per Physician Clinical FTEE

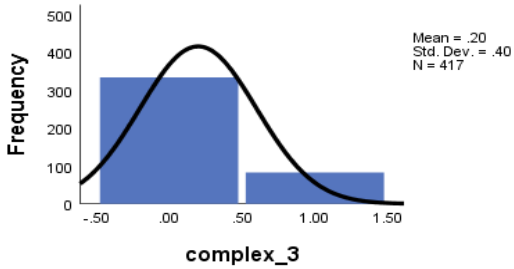

Frequency
Mean = 1.374
Std. Dev. = .817
N = 417
Total Support Staff per Physician Clinical FTEE


Frequency
Mean = .28
Std. Dev. = .45
N = 417
complex_1a

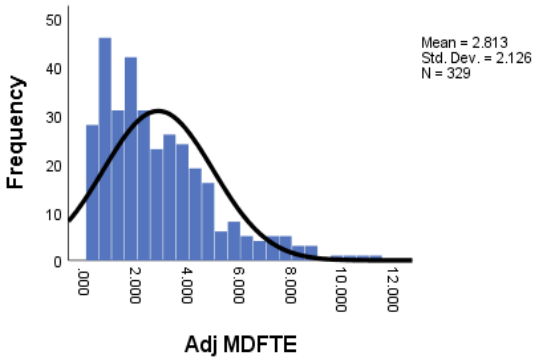


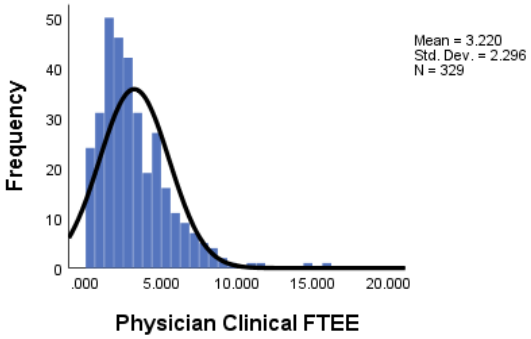


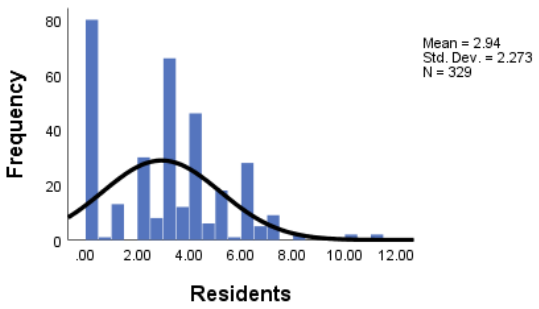


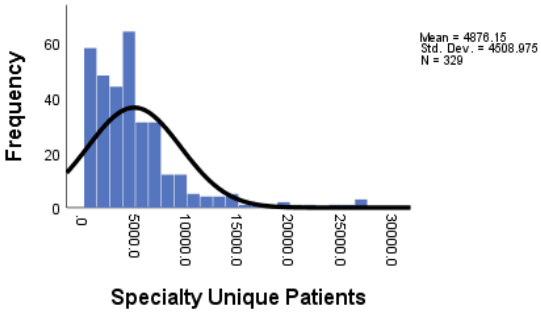


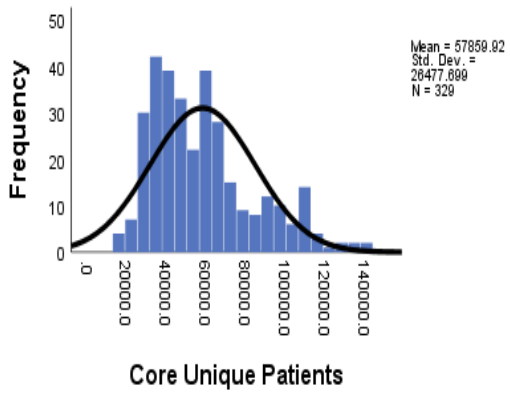
FOR OPHTHALMOLOGY DATA

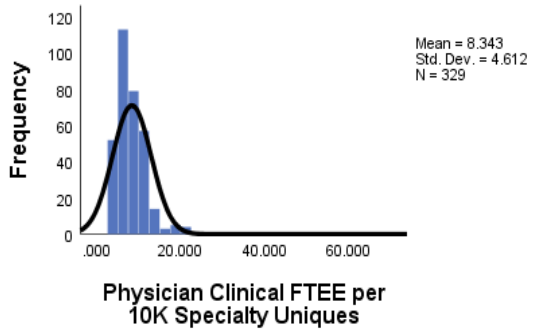


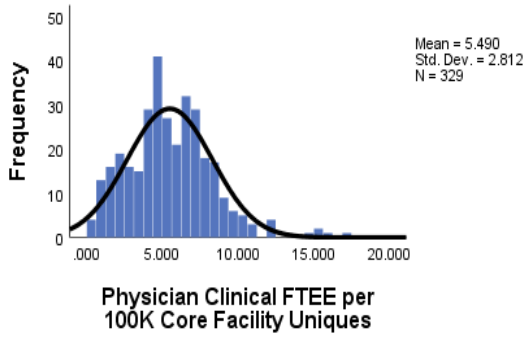


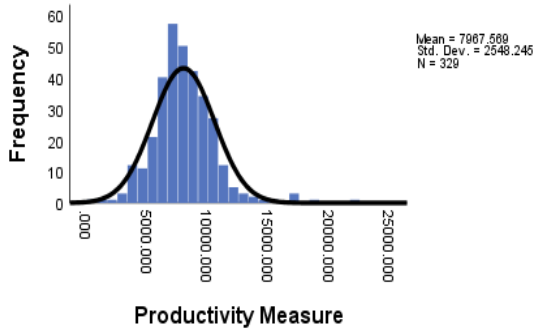


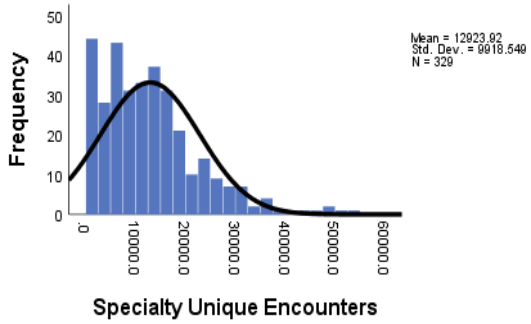


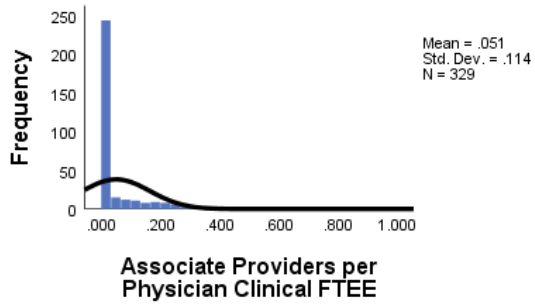

Mean = 57859.92
Std. Dev. = 26477.699
N = 329
Frequency
Core Unique Patients

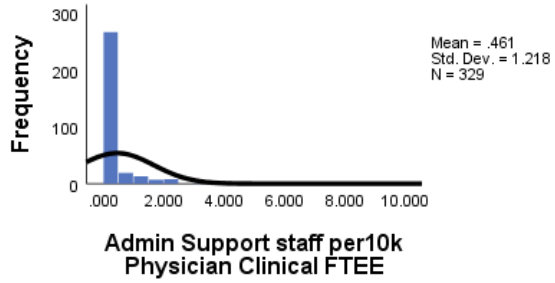

Mean = 8.343
Std. Dev. = 4.612
N = 329
Frequency
Physician Clinical FTEE per 10K Specialty Uniques

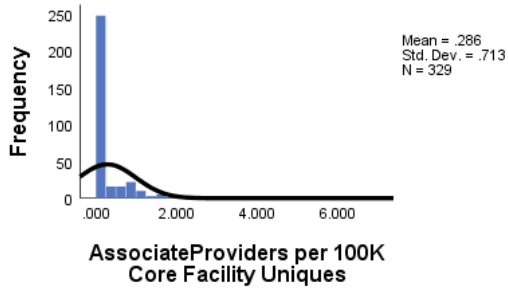

Mean = 5.490
Std. Dev. = 2.812
N = 329
Frequency
Physician Clinical FTEE per 100K Core Facility Uniques

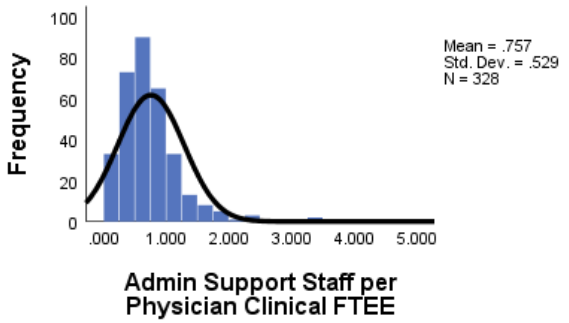

Mean = 7967.569
Std. Dev. = 2548.245
N = 329
Frequency
Productivity Measure

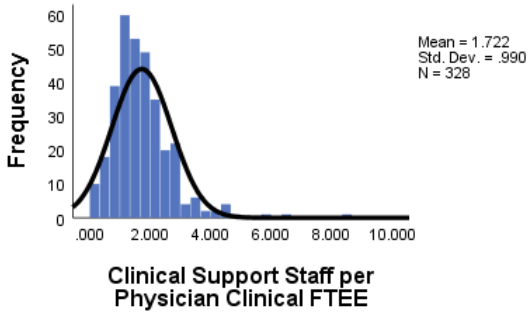

Mean = 12923.92
Std. Dev. = 9918.549
N = 329
Frequency
Specialty Unique Encounters

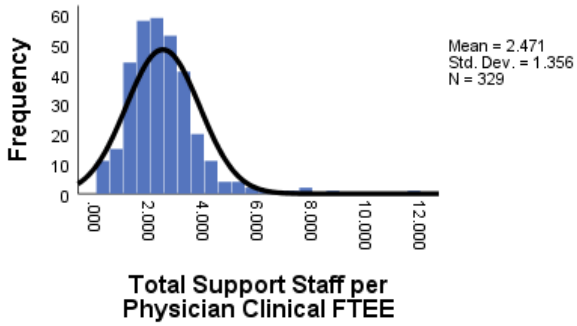

Mean = .051
Std. Dev. = .114
N = 329
Frequency
Associate Providers per Physician Clinical FTEE

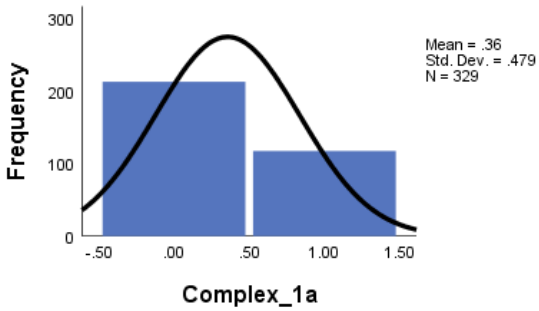

Mean = .461
Std. Dev. = 1.218
N = 329
Frequency
Admin Support staff per10k Physician Clinical FTEE

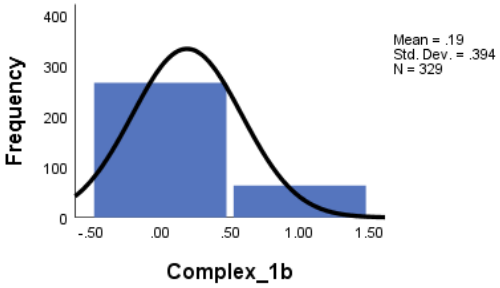

Mean = .286
Std. Dev. = .713
N = 329
Frequency
AssociateProviders per 100K Core Facility Uniques

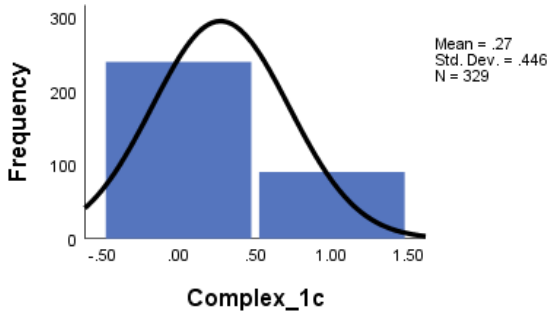

Frequency
Mean = .757
Std. Dev. = .529
N = 328
Admin Support Staff per Physician Clinical FTEE

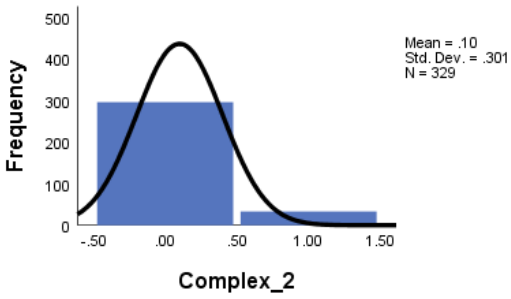

Frequency
Mean = 1.722
Std. Dev. = .990
N = 328
Clinical Support Staff per Physician Clinical FTEE

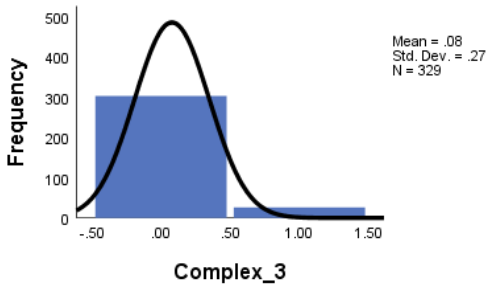

Frequency
Mean = 2.471
Std. Dev. = 1.356
N = 329
Total Support Staff per Physician Clinical FTEE


Frequency
Mean = .36
Std. Dev. = .479
N = 329
Complex_1a

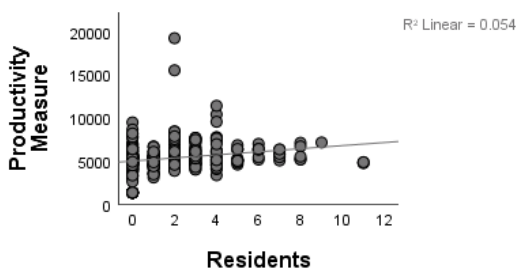

Frequency
Mean = .19
Std. Dev. = .394
N = 329
Complex_1b

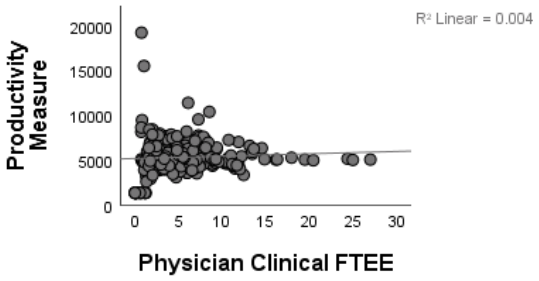

Frequency
Mean = .27
Std. Dev. = .446
N = 329
Complex_1c

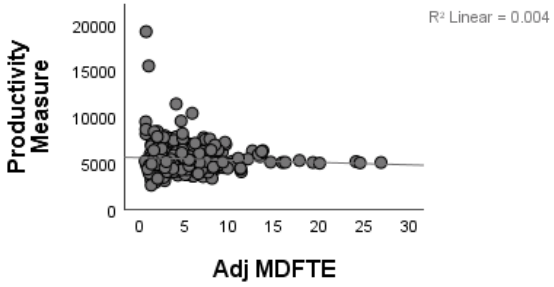

Frequency
Mean = .10
Std. Dev. = .301
N = 329
Complex_2

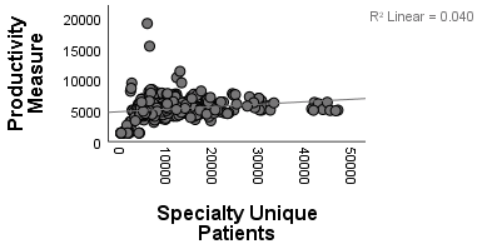

Frequency
Mean = .08
Std. Dev. = .27
N = 329
Complex_3

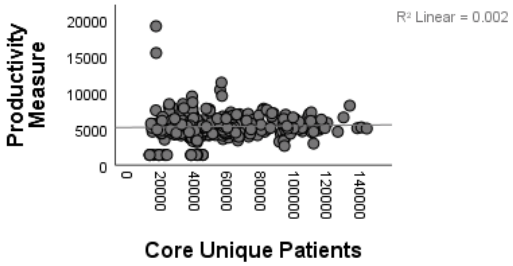
APPENDIX C: SCATTER PLOTS FOR OPTOMETRY DATA

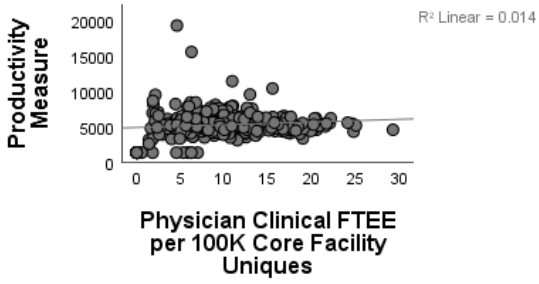


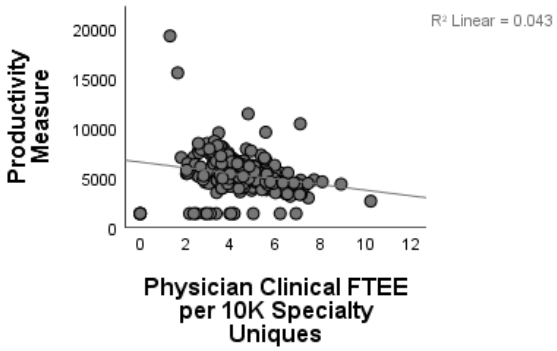


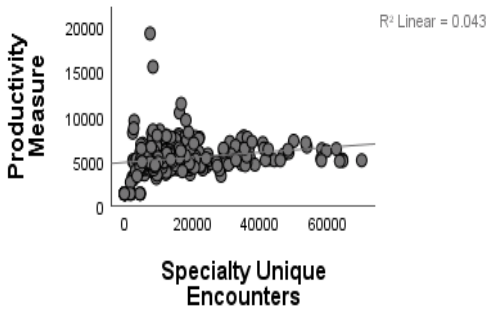


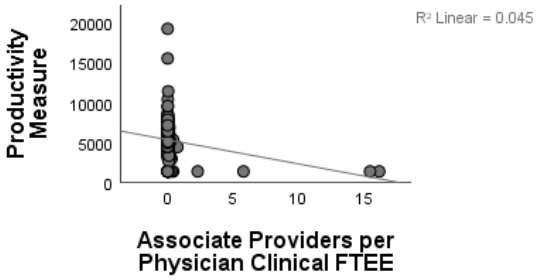


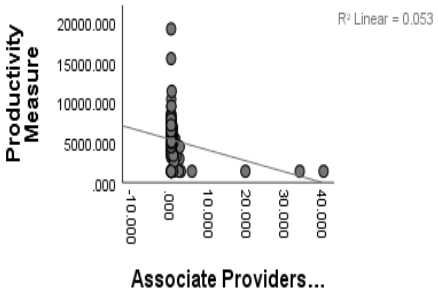


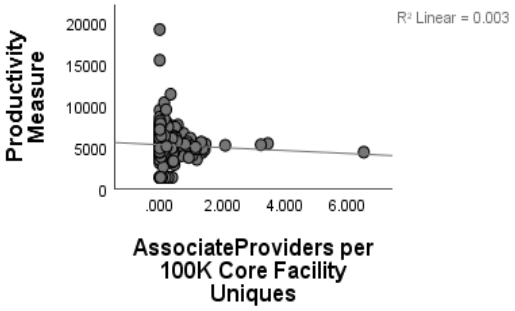


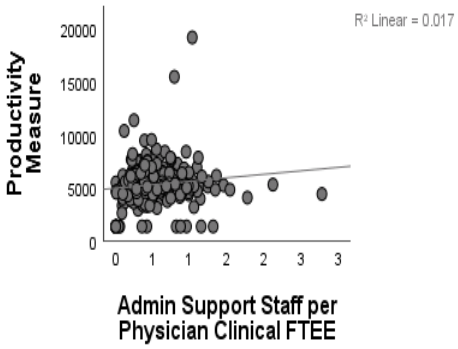

R² Linear = 0.043
Productivity Measure
20000
15000
10000
5000
0
0 2 4 6 8 10 12
Physician Clinical FTEE per 10K Specialty Uniques

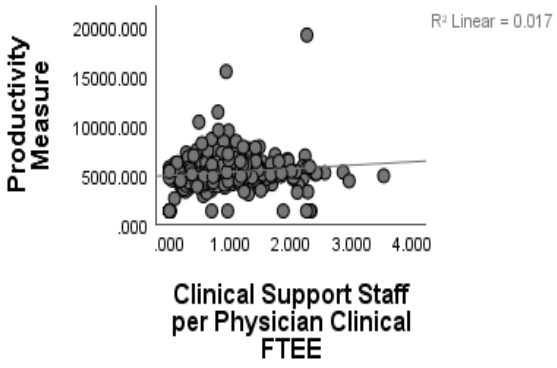

R² Linear = 0.043
Productivity Measure
20000
15000
10000
5000
0
0 20000 40000 60000
Specialty Unique Encounters

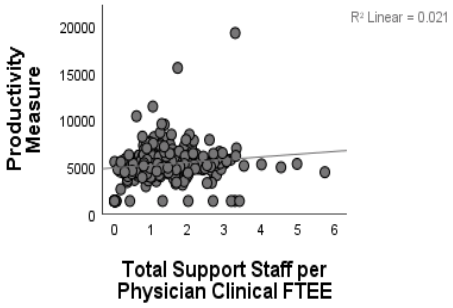

R² Linear = 0.045
Productivity Measure
20000
15000
10000
5000
0
0 5 10 15
Associate Providers per Physician Clinical FTEE


R² Linear = 0.053
Productivity Measure
20000.000
15000.000
10000.000
5000.000
.000
-10.000 .000 10.000 20.000 30.000 40.000
Associate Providers...

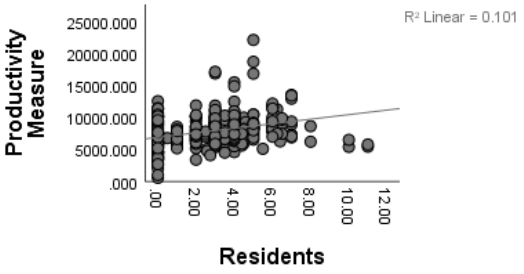

R² Linear = 0.003
Productivity Measure
20000
15000
10000
5000
0
.000 2.000 4.000 6.000
AssociateProviders per 100K Core Facility Uniques

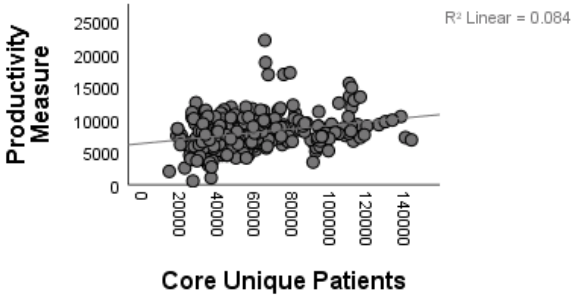

R² Linear = 0.017
Productivity Measure
20000
15000
10000
5000
0
0 1 1 2 2 3 3
Admin Support Staff per Physician Clinical FTEE

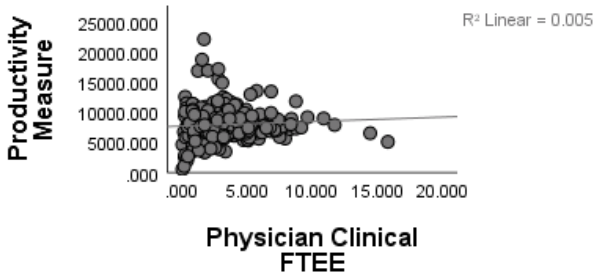

R² Linear = 0.017
Productivity Measure
20000.000
15000.000
10000.000
5000.000
.000
.000 1.000 2.000 3.000 4.000
Clinical Support Staff per Physician Clinical FTEE

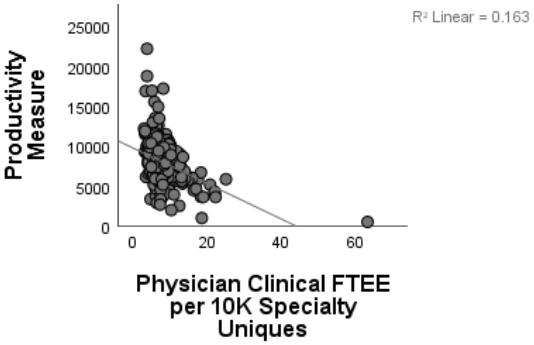

R² Linear = 0.021
Productivity Measure
20000
15000
10000
5000
0
0 1 2 3 4 5 6
Total Support Staff per Physician Clinical FTEE

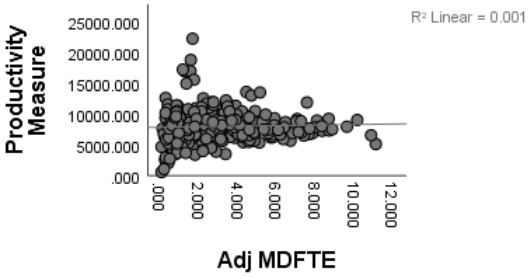
FOR OPHTHALMOLOGY DATA

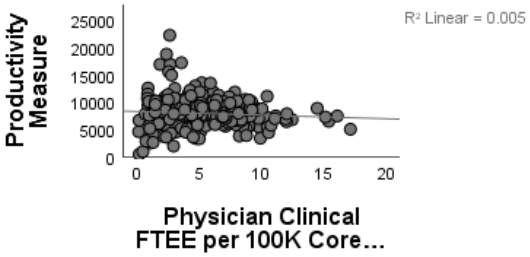


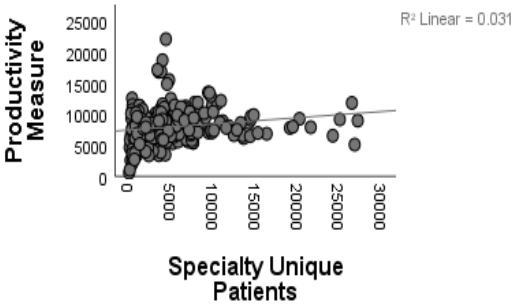


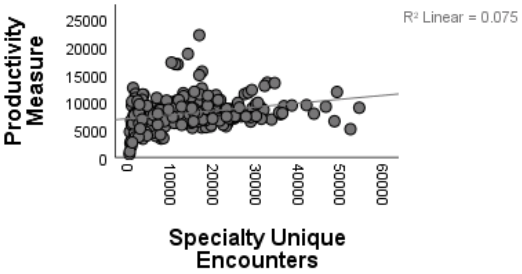


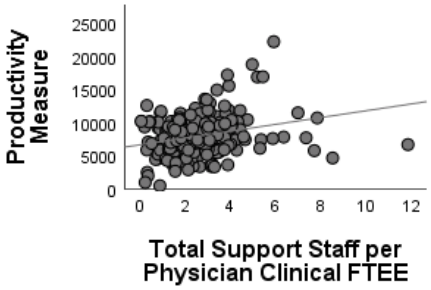


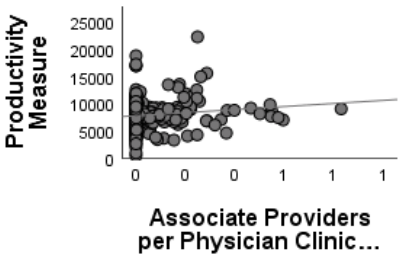


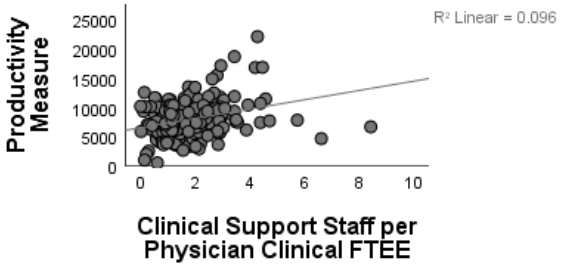


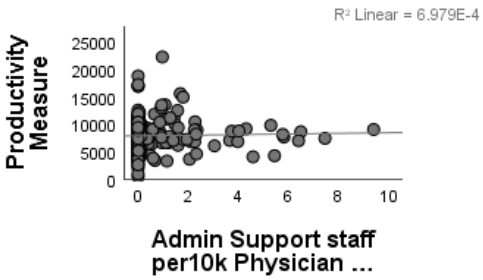





Productivity Measure
25000
20000
15000
10000
5000
0
0 2 4 6 8 10 12
Total Support Staff per Physician Clinical FTEE


Productivity Measure
25000
20000
15000
10000
5000
0
0 0 0 1 1 1
Associate Providers per Physician Clinic...


R² Linear = 0.096
Productivity Measure
25000
20000
15000
10000
5000
0
0 2 4 6 8 10
Clinical Support Staff per Physician Clinical FTEE


R² Linear = 6.979E-4
Productivity Measure
25000
20000
15000
10000
5000
0
0 2 4 6 8 10
Admin Support staff per10k Physician ...

APPENDIX D: OTHER MODELS

OPTOMETRY

1. Principal Component Analysis- MLR-PCA80 (M31)

- **Factor_2**: Admin Support Staff per Physician Clinical FTEE, Clinical Support Staff per Physician Clinical FTEE, Total Support Staff per Physician Clinical FTEE.
- **Factor_5**: Complex_1c, Complex_3
- **Factor_6**: Complex_1b
- **Factor_7**: Complex_2

Table 87: ANOVA Table for M31

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	42.066	7	6.009	6.573	0
Residual	373.934	409	0.914		
Total	416	416			

Table 88: Model Summary for M31

Regression Statistics	
R	0.318
R Squared	0.101
Adjusted R Squared	0.086
Standard Error	0.956
observations	416

Table 89: Regression Coefficients for M31

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-3.90E-16	0.047	0	1	
FACTOR_1	0.035	0.047	0.756	0.45	1
FACTOR_2	0.112	0.047	2.387	0.017	1
FACTOR_3	0.034	0.047	0.731	0.465	1
FACTOR_4	-0.058	0.047	-1.241	0.215	1
FACTOR_5	-0.196	0.047	-4.17	0	1
FACTOR_6	-0.096	0.047	-2.052	0.041	1
FACTOR_7	-0.188	0.047	-4.008	0	1

Table 90: Component Matrix for M31

Component Matrix				
	FACTOR_2	FACTOR_5	FACTOR_6	FACTOR_7
Residents	0.118	0.029	0.205	0.043
SpecialtyUniquePatients	0.231	0.072	0.081	0.013
CoreUniquePatients	0.1	0.007	0.167	0.015
PhysicianClinicalFTEEper10KSpecialtyUniques	0.017	0.182	0.262	0.072
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.282	0.056	0.11	0.05
SpecialtyUniqueEncounters	0.245	0.063	0.1	0.021
AssociateProvidersperPhysicianClinicalFTEE	0.445	0.043	0.029	0.163
AdminSupportstaffper10kPhysicianClinicalFTEE	0.444	0.04	0.031	0.157
AssociateProvidersper100KCoreFacilityUniques	0.387	0.167	0.003	0.101
AdminSupportStaffperPhysicianClinicalFTEE	0.741	0.061	0.02	0.008
ClinicalSupportStaffperPhysicianClinicalFTEE	0.734	0.111	0.031	0.05
TotalSupportStaffperPhysicianClinicalFTEE	0.833	0.057	0.031	0.034
PhysicianClinicalFTEE_1	0.184	0.01	0.05	0.036
AdjMDFTE_1	0.128	0.006	0.1	0.002
complex_1a	0.14	0.432	0.406	0.121
complex_1b	0.228	0.279	0.724	0.33
complex_1c	0.149	0.812	0.325	0.335
complex_2	0.045	0.11	0.262	0.773
complex_3	0.244	0.526	0.377	0.476

2. PCA-High-Level Performers - MLR-PCA80-HIGH (M32)

Factor_1: Residents, Specialty Unique Patients, Core Unique Patients, Specialty Unique Encounters, Admin Support Staff per Physician Clinical FTEE, Clinical Support Staff per Physician Clinical FTEE, Total Support Staff per Physician Clinical FTEE, Physician Clinical FTEE, Adj MDFTE, complex_1a

Table 91: ANOVA for M32

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	39.882	7	5.697	2.665	0.026
Residual	72.675	34	2.138		
Total	112.558	41			

Table 92: Model Summary for M32

Regression Statistics	
R	0.595
R Squared	0.354
Adjusted R Squared	0.221
Standard Error	1.460
observations	41

Table 93: Regression coefficient for M32

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	2.034	0.226	9.015	0	
Factor_1	-0.601	0.228	-2.632	0.013	1
Factor_2	-0.24	0.228	-1.051	0.301	1
Factor_3	0.429	0.228	1.88	0.069	1
Factor_4	0.079	0.228	0.345	0.732	1
Factor_5	-0.321	0.228	-1.405	0.169	1
Factor_6	-0.236	0.228	-1.034	0.308	1
Factor_7	-0.453	0.228	-1.982	0.056	1

Table 94: Component Matrix for M32

Component Matrix	
	FACTOR_1
Residents	0.67
SpecialtyUniquePatients	0.862
CoreUniquePatients	0.739
PhysicianClinicalFTEEper10KSpecialtyUniques	0.393
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.488
SpecialtyUniqueEncounters	0.847
AssociateProvidersperPhysicianClinicalFTEE	0.012
AdminSupportstaffper10kPhysicianClinicalFTEE	0.039
AssociateProvidersper100KCoreFacilityUniques	0.129
AdminSupportStaffperPhysicianClinicalFTEE	0.676
ClinicalSupportStaffperPhysicianClinicalFTEE	0.575
TotalSupportStaffperPhysicianClinicalFTEE	0.665
PhysicianClinicalFTEE_1	0.919
AdjMDFTE_1	0.89
complex_1a	0.669
complex_1b	0.07
complex_1c	0.199
complex_2	0.275
complex_3	0.426

3. PCA-Low-level performers- MLR-PCA80-LOW (M33)

- **Factor_1**: Residents, Specialty Unique Patients, Physician Clinical FTEE per 100K Core specialty Unique, Specialty Unique Encounters, Admin Support Staff per Physician Clinical FTEE, Clinical Support Staff per Physician Clinical FTEE, Total Support Staff per Physician Clinical FTEE, Physician Clinical FTEE, Adj MDFTE, complex_1a.
- **Factor_3**: Associate providers per 100K Core Facility unique, complex_2
- **Factor_4**: Physician Clinical FTEE per 10K specialty Unique, Clinical Support Staff per Physician Clinical FTEE, complex_1c.

Table 95: Anova for M33

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	1.315	6	0.219	5.283	0.001
Residual	0.996	24	0.041		
Total	2.31	30			

Table 96: Model Summary for M33

Regression Statistics	
R	0.754
R Squared	0.569
Adjusted R Squared	0.461
Standard Error	0.204
observations	30

Table 97: Regression coefficient for M33

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-1.298	0.037	-35.478	0	
Factor_1	0.095	0.037	2.558	0.017	1
Factor_2	-0.012	0.037	-0.317	0.754	1
Factor_3	-0.092	0.037	-2.464	0.021	1
Factor_4	-0.161	0.037	-4.327	0	1
Factor_5	0.019	0.037	0.505	0.619	1
Factor_6	0.003	0.037	0.089	0.93	1

Table 98: Component Matrix for M33

Component Matrix			
	FACTOR_1	FACTOR_3	FACTOR_4
Residents	0.859	0.029	0.11
SpecialtyUniquePatients	0.936	0.209	0.072
CoreUniquePatients	0.155	0.144	0.315
PhysicianClinicalFTEEper10KSpecialtyUniques	0.039	0.075	0.681
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.809	0.299	0.006
SpecialtyUniqueEncounters	0.95	0.191	0.024
AssociateProvidersperPhysicianClinicalFTEE	0.459	0.449	0.073
AdminSupportstaffper10kPhysicianClinicalFTEE	0.447	0.376	0.065
AssociateProvidersper100KCoreFacilityUniques	0.064	0.821	0.279
AdminSupportStaffperPhysicianClinicalFTEE	0.523	0.253	0.142
ClinicalSupportStaffperPhysicianClinicalFTEE	0.519	0.182	0.559
TotalSupportStaffperPhysicianClinicalFTEE	0.618	0.242	0.411
PhysicianClinicalFTEE_1	0.951	0.164	0.094
AdjMDFTE_1	0.928	0.224	0.001
complex_1a	0.64	0.144	0.133
complex_1b	0.393	0.271	0.499
complex_1c	0.117	0.212	0.548
complex_2	0.278	0.703	0.01
complex_3	0.138	0.151	0.425

4. PCA-Stepwise -Step-PCA80(M41)

- **Factor_2**: Admin Support Staff per Physician Clinical FTEE, Clinical Support Staff per Physician Clinical FTEE, Total Support Staff per Physician Clinical FTEE
- **Factor_5**: complex_1c, complex_3
- **Factor_6**: complex_1b
- **Factor_7**: complex_2

Table 99: Model Summary for M41

Regression Statistics	
R	0.309
R squared	0.095
Adjusted R squared	0.087
Standard Error	0.956
Observations	416

Table 100: ANOVA for M41

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	39.647	4	9.912	10.85	0
Residual	376.353	412	0.913		
Total	416	416			

Table 101: Regression Coefficients for M41

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-3.92E-16	0.047	0	1	
Factor_5	-0.196	0.047	-4.172	0	1
Factor_7	-0.188	0.047	-4.01	0	1
Factor_2	0.112	0.047	2.388	0.017	1
Factor_6	-0.096	0.047	-2.053	0.041	1

Table 102: Component Matrix for M41

Component Matrix				
	FACTOR_2	FACTOR_5	FACTOR_6	FACTOR_7
Residents	0.118	0.029	0.205	0.043
SpecialtyUniquePatients	0.231	0.072	0.081	0.013
CoreUniquePatients	0.1	0.007	0.167	0.015
PhysicianClinicalFTEEper10KSpecialtyUniques	0.017	0.182	0.262	0.072
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.282	0.056	0.11	0.05
SpecialtyUniqueEncounters	0.245	0.063	0.1	0.021
AssociateProvidersperPhysicianClinicalFTEE	0.445	0.043	0.029	0.163
AdminSupportstaffper10kPhysicianClinicalFTEE	0.444	0.04	0.031	0.157
AssociateProvidersper100KCoreFacilityUniques	0.387	0.167	0.003	0.101
AdminSupportStaffperPhysicianClinicalFTEE	0.741	0.061	0.02	0.008
ClinicalSupportStaffperPhysicianClinicalFTEE	0.734	0.111	0.031	0.05
TotalSupportStaffperPhysicianClinicalFTEE	0.833	0.057	0.031	0.034
PhysicianClinicalFTEE_1	0.184	0.01	0.05	0.036
AdjMDFTE_1	0.128	0.006	0.1	0.002
complex_1a	0.14	0.432	0.406	0.121
complex_1b	0.228	0.279	0.724	0.33
complex_1c	0.149	0.812	0.325	0.335
complex_2	0.045	0.11	0.262	0.773
complex_3	0.244	0.526	0.377	0.476

5. Step PCA100 (M42)

Table 103: ANOVA Table for Step PCA 100

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	110.098	5	22.02	29.585	0
Residual	305.902	411	0.744		
Total	416	416			

Table 104: Regression Statistics for Step PCA 100

Regression Statistics	
R	0.514
R squared	0.265
Adjusted R squared	0.256
Standard Error	0.862
Observations	416

Table 105: Regression Coefficients for Step PCA 100

Predictors	Coeffiicients	Std. Error	t	Sig.	VIF
(Constant)	-5.23E-16	0.042	0	1	
Factor_8	-0.414	0.042	-9.786	0	1
Factor_12	0.255	0.042	6.022	0	1
Factor_9	-0.106	0.042	-2.495	0.013	1
Factor_4	-0.097	0.042	-2.302	0.022	1
Factor_5	-0.088	0.042	-2.089	0.037	1

Table 106: Component Matrix for Step PCA 100

Predictors	Factor_8	Factor_12	Factor_9	Factor_4	Factor_5
Residents	0.062	0.209	0.014	0.263	0.029
SpecialtyUniquePatients	0.154	0.032	0.081	0.03	0.072
CoreUniquePatients	0.129	0.176	0.104	0.472	0.007
PhysicianClinicalFTEEper10KSpecialtyUniques	0.59	0.104	0.402	0.317	0.182
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.067	0.261	0.03	0.624	0.056
SpecialtyUniqueEncounters	0.138	0.011	0.059	0.047	0.063
AssociateProvidersperPhysicianClinicalFTEE	0.052	0.027	0.101	0.495	0.043
AdminSupportstaffper10kPhysicianClinicalFTEE	0.089	0.019	0.101	0.483	0.04
AssociateProvidersper100KCoreFacilityUniques	0.498	0.072	0.639	0.053	0.167
AdminSupportStaffperPhysicianClinicalFTEE	0.081	0.073	0.031	0.049	0.061
ClinicalSupportStaffperPhysicianClinicalFTEE	0.031	0.011	0.347	0.086	0.111
TotalSupportStaffperPhysicianClinicalFTEE	0.009	0.021	0.242	0.044	0.057
PhysicianClinicalFTEE_1	0.006	0.005	0.02	0.104	0.01
AdjMDFTE_1	0.003	0.032	0.049	0.165	0.006
complex_1a	0.055	0.194	0.036	0.372	0.432
complex_1b	0.31	0.077	0.134	0.331	0.279
complex_1c	0.054	0.047	0.132	0.18	0.812
complex_2	0.286	0.097	0.045	0.384	0.11
complex_3	0.149	0.154	0.017	0.193	0.526

- **Factor_8:** Physician Clinical FTEE per 10K specialty Unique

- **Factor_9:** Associate providers per 100K Core Facility unique
- **Factor_4:** Physician Clinical FTEE per 100K Core Facility Unique
- **Factor_5:** Complex_1c, complex_3

OPHTHALMOLOGY

1. MLR-PCA80 (M71)

- **Factor_1**: Residents, Specialty Unique patients, Core unique patients, Physician Clinical FTEE per 100K Specialty Unique, Specialty Unique Encounters, complex_1a, Physician Clinical FTEE, Adj MDFTE
- **Factor_2:** Associate Providers per Physician Clinical FTEE, Admin Support staff per Physician Clinical FTEE, Clinical Support staff per Physician Clinical FTEE, Total Support staff per Physician Clinical FTEE
- **Factor_4**: Complex_1c, Complex_2
- **Factor_6**: Complex_2, Complex_3

Table 107: ANOVA Table for M71

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	41.827	6	6.971	7.723	0.000
Residual	332.173	368	0.903		
Total	374	374			

Table 108: Model Summary for M71

Regression Statistics	
R	0.334
R Squared	11.2
Adjusted R Squared	0.097
Standard Error	0.950
observations	374

Table 109: Regression Coefficients for M71

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-1.16E-15	0.049	0	1	
FACTOR_1	0.18	0.049	3.655	0	1
FACTOR_2	0.218	0.049	4.429	0	1
FACTOR_3	0.059	0.049	1.193	0.234	1
FACTOR_4	0.133	0.049	2.717	0.007	1
FACTOR_5	0.021	0.049	0.434	0.664	1
FACTOR_6	0.103	0.049	2.09	0.037	1

Table 110: Component Matrix for M71

Component Matrix				
	FACTOR_1	FACTOR_2	FACTOR_4	FACTOR_6
Residents	0.765	0.076	0.103	0.027
SpecialtyUniquePatients	0.875	0.182	0.009	0.016
CoreUniquePatients	0.77	0.115	0.038	0.072
PhysicianClinicalFTEEper10KSpecialtyUniques	0.086	0.418	0.333	0.235
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.67	0.281	0.081	0.163
SpecialtyUniqueEncounters	0.946	0.146	0.01	0.006
AssociateProvidersperPhysicianClinicalFTEE	0.325	0.523	0.02	0.016
AdminSupportstaffper10kPhysicianClinicalFTEE	0.277	0.48	0.006	0.005
AssociateProvidersper100KCoreFacilityUniques	0.474	0.416	0.007	0.002
AdminSupportStaffperPhysicianClinicalFTEE	0.164	0.704	0.128	0.043
ClinicalSupportStaffperPhysicianClinicalFTEE	0.19	0.765	0.051	0.045
TotalSupportStaffperPhysicianClinicalFTEE	0.201	0.828	0.006	0.049
complex_1a	0.756	0.045	0.204	0.073
complex_1b	0.048	0.032	0.039	0.121
complex_1c	0.277	0.084	0.857	0.199
complex_2	0.344	0.047	0.538	0.675
complex_3	0.372	0.034	0.33	0.752
PhysicianClinicalFTEE	0.936	0.194	0.008	0.047
AdjMDFTE	0.852	0.225	0.141	0.023

2. MLR-PCA80-HIGH (M72)

- **Factor_3**: Physician Clinical FTEE per 10K Specialty Unique, Clinical Support Staff per Physician Clinical FTEE, and Total Support Staff per Physician Clinical FTEE

  Factor_4: complex_1c

Table 111: Model Summary for M72

Regression Statistics	
R	0.641
R Squared	0.411
Adj. R Squared	0.293
Standard Error	0.939
observations	36

Table 112: Anova for M72

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	18.462	6	3.077	3.488	0.010
Residual	26.465	30	0.882		
Total	44.927	36			

Table 113: Regression coefficient for M72

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	1.939	0.154	12.555	0	
Factor_1	-0.218	0.157	-1.39	0.175	1
Factor_2	0.195	0.157	1.245	0.223	1
Factor_3	0.423	0.157	2.704	0.011	1
Factor_4	0.453	0.157	2.892	0.007	1
Factor_5	0.019	0.157	0.123	0.903	1
Factor_6	-0.207	0.157	-1.325	0.195	1

Table 114: Component Matrix for M72

Component Matrix		
	Factor_3	Factor_4
Residents	0.388	0.186
SpecialtyUniquePatients	0.202	0.086
CoreUniquePatients	0.02	0.336
PhysicianClinicalFTEEper10KSpecialtyUniques	0.507	0.181
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.172	0.462
SpecialtyUniqueEncounters	0.122	0.035
AssociateProvidersperPhysicianClinicalFTEE	0.366	0.197
AdminSupportstaffper10kPhysicianClinicalFTEE	0.425	0.244
AssociateProvidersper100KCoreFacilityUniques	0.284	0.016
AdminSupportStaffperPhysicianClinicalFTEE	0.305	0.41
ClinicalSupportStaffperPhysicianClinicalFTEE	0.538	0.028
TotalSupportStaffperPhysicianClinicalFTEE	0.503	0.177
PhysicianClinicalFTEE_1	0.073	0.101
AdjMDFTE_1	0.035	0.143
complex_1a	0.129	0.034
complex_1b	0.028	0.404
complex_1c	0.398	0.628
complex_2	0.418	0.485
complex_3	0.288	0.428

3. MLR-PCA80-LOW (M73)

- **Factor_1**: Specialty Unique Patients, Core Unique Patients, Physician Clinical FTEE per 100K specialty unique, Specialty Unique Encounters, Associate Providers per 100K Core Facility Unique, Physician Clinical FTEE, Adj MDFTE

  Factor_4: Residents, complex_1a, complex_2

Table 115: Model Summary for M73

Regression Statistics	
R	0.584
R Squared	0.341
Adjusted R Squared	0.225
Standard Error	0.460
observations	40

Table 116: Anova table for M73

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	3.733	6	0.622	2.938	0.020
Residual	7.201	34	0.212		
Total	10.934	40			

Table 117: Regression coefficient for M73

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-1.613	0.072	-22.438	0	
Factor_1	0.186	0.073	2.55	0.015	1
Factor_2	0.126	0.073	1.726	0.093	1
Factor_3	-0.005	0.073	-0.071	0.944	1
Factor_4	0.196	0.073	2.7	0.011	1
Factor_5	0.026	0.073	0.363	0.719	1
Factor_6	-0.062	0.073	-0.848	0.402	1

Table 118: Component Matrix for M73

Component Matrix		
	Factor_1	Factor_4
Residents	0.375	0.755
Specialty Unique Patients	0.964	0.06
Core Unique Patients	0.527	0.401
Physician Clinical FTEE per 10K Specialty Uniques	0.333	0.191
Physician Clinical FTEE per 100K Core Facility Uniques	0.846	0.297
Specialty Unique Encounters	0.967	0.045
Associate Providers per Physician Clinical FTEE	0.031	0.109
Admin Support staff per10k Physician Clinical FTEE	0.024	0.021
AssociateProviders per 100K Core Facility Uniques	0.634	0.297
Admin Support Staff per Physician Clinical FTEE	0.019	0.112
Clinical Support Staff per Physician Clinical FTEE	0.045	0.036
Total Support Staff per Physician Clinical FTEE	0.026	0.073
PhysicianClinicalFTEE	0.972	0.119
AdjMDFTE	0.955	0.2
complex_1a	0.44	0.603
complex_1b	0.137	0.32
complex_1c	0.138	0.09
complex_2	0.031	0.538
complex_3	0.165	0.266

4. STEP-PCA80 (M81)

- **Factor_1**: residents, Specialty Unique Patients, Core Unique Patients, Physician Clinical FTEE per 100k Specialty Unique, Specialty Unique Encounters, complex_1a, Physician Clinical FTEE, Adj MDFTE.
- **Factor_2**: Associate Providers per Physician Clinical FTEE, Admin Support Staff per Physician Clinical FTEE, Clinical Support Staff per Physician Clinical FTEE, Total Support Staff per Physician Clinical FTEE.
- **Factor_3**: Associate Providers per Physician Clinical FTEE, Admin Support Staff per 10K Physician Clinical FTEE, Associate Providers per 100K Core Facility Unique, Clinical Support Staff per Physician Clinical FTE
- **Factor_4**: complex_1c, complex_2.

Models M22 and M24 have similar R-squared and adjusted R-squared values. Also, the significant variables for both models are almost the same.

Table 119: Model Summary for M81

Regression Statistics	
R	0.333
R Squared	0.111
Adjusted R Squared	0.101
Standard Error	0.948
observations	374

Table 120: ANOVA table for M81

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	41.376	4	10.344	11.506	0.000
Residual	332.624	370	0.899		
Total	374	374			

Table 121: Regression Coefficients for M81

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-1.16E-15	0.049	0	1	
FACTOR_1	0.226	0.049	4.614	0	1
FACTOR_2	0.164	0.049	3.353	0.001	1
FACTOR_3	-0.151	0.049	-3.088	0.002	1
FACTOR_4	-0.098	0.049	-1.99	0.047	1

Table 122: Component Matrix for M81

Component Matrix				
	FACTOR_1	FACTOR_2	FACTOR_3	FACTOR_4
Residents	0.765	0.076	0.071	0.103
SpecialtyUniquePatients	0.875	0.182	0.168	0.009
CoreUniquePatients	0.77	0.115	0.109	0.038
PhysicianClinicalFTEEper10KSpecialtyUniques	0.086	0.418	0.26	0.333
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.67	0.281	0.059	0.081
SpecialtyUniqueEncounters	0.946	0.146	0.132	0.01
AssociateProvidersperPhysicianClinicalFTEE	0.325	0.523	0.722	0.02
AdminSupportstaffper10kPhysicianClinicalFTEE	0.277	0.48	0.788	0.006
AssociateProvidersper100KCoreFacilityUniques	0.474	0.416	0.689	0.007
AdminSupportStaffperPhysicianClinicalFTEE	0.164	0.704	0.305	0.128
ClinicalSupportStaffperPhysicianClinicalFTEE	0.19	0.765	0.509	0.051
TotalSupportStaffperPhysicianClinicalFTEE	0.201	0.828	0.493	0.006
complex_1a	0.756	0.045	0.027	0.204
complex_1b	0.048	0.032	0.224	0.039
complex_1c	0.277	0.084	0.021	0.857
complex_2	0.344	0.047	0.087	0.538
complex_3	0.372	0.034	0.083	0.33
PhysicianClinicalFTEE	0.936	0.194	0.049	0.008
AdjMDFTE	0.852	0.225	0.005	0.141

5. STEP-PCA100 (M82)

- **Factor_3:** Associate Providers per Physician Clinical FTEE, Admin Support staff per 10k Physician Clinical FTEE, Associate Providers per 100K Core Facility Unique, Clinical Support Staff per Physician Clinical FTEE.
- **Factor_6:** complex_2, complex_3.
- **Factor_8:** Physician Clinical FTEE per 10K Specialty Unique.
- **Factor_1:** Residents, Physician Clinical FTEE, Specialty Unique Patients, Core Unique Patients, Physician Clinical FTEE per 100K Core Facility Unique, Specialty Unique Encounters, complex_1a, AdjMDFTE_1

ANOVA Table for Step PCA100

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	134.909	11	12.264	18.621	0
Residual	239.091	363	0.659		
Total	374	374			

Regression Statistics for Step PCA100

Regression Statistics	
R	0.601
R squared	0.361
Adjusted R squared	0.341
Standard Error	0.811
Observations	374

Regression Coefficients for Step PCA100

Predictors	Coefficients	Std. Error	t	Sig.	VIF
(Constant)	-1.288E-15	0.042	0	1	
Factor_9	0.242	0.042	5.76	0	1
Factor_3	0.24	0.045	5.716	0	1
Factor_16	0.228	0.042	5.434	0	1
Factor_11	-0.217	0.042	-5.171	0	1
Factor_7	0.198	0.044	4.719	0	1
Factor_17	-0.18	0.042	-4.3	0	1
Factor_6	0.167	0.043	3.974	0	1
Factor_13	-0.113	0.042	-2.695	0.007	1
Factor_15	0.109	0.042	2.591	0.01	1
Factor_8	-0.108	0.042	-2.58	0.01	1
Factor_1	0.098	0.042	2.346	0.02	1

Component Matrix for Step PCA100

Predictors	Factor_9	Factor_3	Factor_16	Factor_11	Factor_7	Factor_17	Factor_6	Factor_13	Factor_15	Factor_8	Factor_1
Residents	0.49	0.069	0.021	0.311	0.067	0.005	0.023	0.112	0.001	0.056	0.765
PhysicianClinicalFTEE	0.128	0.057	0.006	0.025	0.145	0.119	0.059	0.043	0.003	0.042	0.943
SpecialtyUniquePatients	0.211	0.165	0.104	0.122	0.103	0.014	0.015	0.079	0.036	0.071	0.875
CoreUniquePatients	0.236	0.107	0.023	0.094	0.402	0.043	0.075	0.12	0.063	0.259	0.771
PhysicianClinicalFTEEper10KSpecialtyUniques	0.056	0.262	0.003	0.049	0.456	0.004	0.237	0.018	0.009	0.565	0.085
PhysicianClinicalFTEEper100KCoreFacilityUniques	0.201	0.057	0.013	0.149	0.476	0.05	0.165	0.144	0.043	0.279	0.672
SpecialtyUniqueEncounters	0.084	0.129	0.16	0.091	0.032	0.014	0.004	0.062	0.02	0.008	0.946
AssociateProvidersperPhysicianClinicalFTEE	0.057	0.722	0.005	0.048	0.014	0.008	0.016	0.029	0.127	0	0.325
AdminSupportstaffper10kPhysicianClinicalFTEE	0.009	0.788	0.005	0.014	0.039	0.008	0.005	0.078	0.215	0.047	0.277
AssociateProvidersper100KCoreFacilityUniques	0.001	0.689	0.003	0.07	0.119	0.004	0.002	0.098	0.095	0.023	0.474
AdminSupportStaffperPhysicianClinicalFTEE	0.082	0.306	0.009	0.042	0.104	0.001	0.042	0.015	0.01	0.451	0.168
ClinicalSupportStaffperPhysicianClinicalFTEE	0.046	0.51	0.007	0.066	0.063	0.003	0.045	0.034	0.007	0.051	0.194
TotalSupportStaffperPhysicianClinicalFTEE	0.006	0.495	0.002	0.036	0.011	0.002	0.049	0.02	0.009	0.121	0.206
complex_1a	0.279	0.029	0.007	0.227	0.274	0.015	0.075	0.017	0.007	0.15	0.755
complex_1b	0.09	0.224	0.003	0.041	0.083	0.01	0.127	0.003	0.006	0.067	0.049
complex_1c	0.112	0.02	0.004	0.074	0.202	0.01	0.203	0.027	0.003	0.094	0.278
complex_2	0.102	0.085	0.005	0.144	0.094	0.006	0.676	0.011	0.011	0.277	0.342
complex_3	0.042	0.083	0.004	0.168	0.311	0.007	0.748	0.038	0.02	0.131	0.374
AdjMDFTE_1	0.252	0.002	0.001	0.051	0.136	0.033	0.024	0.294	0.039	0.069	0.849