STAT 7995 Picket Report

Dominic Matriccino

2022-09-27

1 Executive Summary

Our analysis of this data gave us a good picture of which variables are associated with shortages. We first subsetted our data to include only the 10 drug classifications that had the highest proportion of drugs that had ever been in a shortage. This information on its own is surely useful to know and it is something that could be studied further. As with our predictive abilities, all of our models were able to perform reasonably well considering our unbalanced data. Our logistic regression model, after having lowered the classification threshold to 0.325, received an accuracy rate of roughly 65%. Our two poisson regression models also performed well. With respect to predicting the total number of months that each drug has been in a shortage, our model was able to predict this number within 5 months of the true number, 77.9% of the time. With respect to predicting the number of shortages for each drug, our poisson model was able to predict this number within 1 shortage 68.87% of the time. Beyond just looking at the evaluative measures, it is important to look at the statistical significance of the predictors of our models because this will give Picket an idea of the attributes of the drugs that tend to go into shortage. Generally speaking, the variable Form appeared significant in all of our models. More specifically, the level of Form, SYRINGE and VIAL, were some of the most significant predictors across the entire analysis. More research should be done to determine if there is some causal relationship to why these levels of the variable Form are significant. There are many other variables that are close to being statistically significant that should be looked into further. Some of these variables that are on the border of being significant may, in actuality, be practically significant but statistically insignificant. Picket should determine whether these conclusions are consistent with their own knowledge of the nature of shortages, and if not, they should investigate further to determine if there are potentially areas of their business that they could improve on with respect to manufacturing drugs that are most likely to go into shortage.

2 Introduction

2.1 General Background

Drug patents are granted to companies that discover new methods to treat and prevent diseases. Patents protect drug companies by giving companies exclusive rights to sell the drug in the open market. The research and development stage of drug manufacturing can take upwards of 10-15 years. This process requires a significant amount of a company's resources, and thus companies require assurance that they will be able to profit off the drug once the development stage is completed. Patents, however, eventually expire, which allows outside companies to manufacture and sell the drug in the open market. Generic drugs are the name for drugs that are manufactured after a patent of a brand name drug has expired. When drug patents expire multiple manufacturers can enter the market, which usually results in a decrease in prices due to the competition. Lack of competition is highly correlated with drug shortages. When there are only a few manufacturers of a drug, the risk of having a significant shortage is higher. Generic drug manufacturers can play a role in solving this issue. More competitors must strategically enter certain markets to prevent drug shortages from impacting patient care across the U.S. Tragically, 90% of emergency room doctors report that they have experienced a shortage that impacted their ability to treat critically injured patients. The manufacturing of drugs is very complex, which is partly to blame for the frequency of shortages. Any little production flaw will cause a manufacturer to halt production and investigate the cause of the error. Drug manufacturing must be highly precise to avoid giving patients defective drugs. While there are many reasons why a drug may go into shortage, the most solvable of these issues are issues related to manufacturing. Companies that have a better understanding of the nature of shortages with respect to manufacturing will be at a significant advantage with regard to entering the market. While any company can make money selling drugs, the company that leverages knowledgeable insights and uses these insights to strategically enter certain markets will put themselves in a position to be extremely successful. Picket is a pharmaceutical manufacturing company that manufactures generic drugs. They hope to gain a better understanding of the nature of shortages so that they can successfully run their business while also preventing the consequences of drug shortages. A great solution to this problem will allow Picket to maximize their profits while also benefiting patients and their access to essential drugs.

2.2 Objectives

Our project revolves around understanding the nature of shortages. This is a rather broad question so we will break it down into four parts:

1. What type of drugs commonly experience shortages?
2. Predict which drugs have experienced shortages
3. Predict the frequency of shortages
4. Predict the length of a shortage

After determining this information, we can create a full-fledged recommendation regarding the attributes of drugs that Picket should look out for that may prove to be the most lucrative when deciding on what to manufacture and ultimately bring to market.

3 Approach to Project

To move forward with our analysis, we first need to create a few new variables that will act as the response variable for a few of our models. In the data we are given the start and end month of each shortage. We can use this information to create a new column in our dataset that imputes the number of shortages that have occurred for each unique drug within the data. This new column allows us to run a form of regression to predict the frequency of shortages for a drug. Another column we introduced into the data was a column that calculates the total number of months that each drug has been in a shortage. This column will also function as a response variable for one of our regression models as we try to predict the total number of shortage months for each unique drug in the data. The last column that we needed to create was a column that calculates when each drug was launched relative to the first month included in the data. This column will also serve as a predictor to help our models predict These columns will assist us as we create models that allow us to understand and predict the relationships between shortages and our covariates.

To answer our first question of interest, we can simply look at the given data to understand which types of drugs are most often in shortage. The drugs that we would expect to be most often in shortage are drugs where the number of patients that need these drugs is less consistent. Sudden variation in the demand for drugs is most certainly a cause of a shortage. Patients that have longstanding health issues likely require a steady amount of a drug per day/week, which would make it quite easy to predict how much of a drug that a hospital/pharmacy needs to order. Issues arise when emergency situations require medical professionals to use amounts of drugs that they usually would not need to administer in an average week. Having one week where there are an unusual amount of emergencies could potentially cause a hospital to use up a significant amount of certain medications which could leave more routine patients lacking medications that they require. While we do not have any information regarding whether a drug is in shortage within certain hospitals, it should be understood that certain drugs are more likely to go into shortage due to the reasons as to why the drug is administered. Is a drug absolutely and immediately needed for the survival of a patient? This would indicate that this drug is susceptible to sudden increases in the quantity demanded because patients that need these drugs are not able to wait for treatment. As we look through our dataset, we see that the column DrugClass describes the classification of the drug, or in other words, what the drug is used for. As previously described certain drugs may be more or less susceptible to shortages. To find this out we can simply create a table that lists each drug classification and the number of times each class appears in the data. From here we can break down the data even further to see how the count of each drug classification breaks down into the variable that indicates whether a drug has had at least one shortage, NDC_Shortage_FL. This will give us a better idea of which drugs tend to have more or less shortages.

Our second question of interest is if we are able to accurately predict whether or not a drug has had a shortage. Our dataset covers roughly 6 years, which should be plenty of data to allow our models to pick up on the nuances of the covariates. We will use a logistic regression model to predict whether a drug has experienced a least one shortage in the past 6 years. Our response variable that we will predict is NDC_Shortage_FL. This variable is binary categorical variable with level 1 indicating that a drug has been in shortage at least once and level 0 indicating that a drug has never been in shortage over the course of the 71 months contained in the data. While being able to predict whether a drug has been in shortage may be useful on its own, what may be more interesting are the coefficients that are determined from the models and how they relate to the response variable of interest. We can look at whether each coefficient is positive or negative as well as the magnitude. Before we conducted the analysis we needed to deal with the fact that there is a significant imbalance between levels of our response variable. Only 21% of the raw data contains a zero value for NDC_Shortage_FL. To make up for this we need to find a better subset of the data to allow for our model to make more valid predictions. To do this, we looked at the variable DrugClass and found the drugs that have a significant number of observations in both classes of our response variable while also having a decent sample size. We calculated the proportion of observations of each drug classification that have had at least one shortage. This full table is in the appendix.

Looking at this table and only using the drug classifications that have the largest proportions is certainly enticing for a logistic regression, but these proportions are misleading due to uneven observations within each drug classification. For example, there are only 5 drugs within the whole data with the drug classification of miotics antiglaucoma preps. As previously mentioned, we want to extract some of the drug classifications that are as balanced as possible with respect to our response variable while also considering the number of observations. Our final subsetted data that we will run all of our models on contains the top 10 drug classifications based on the proportion of observations that are classified as having a shortage, while also having at least 20 observations in the sample. Our final 10 drugs classifications are listed with the corresponding proportion of observations that have been in a shortage.

```
##
##
## ===========================  ====================================================
## \                            Proportion of Drugs that have Experienced a Shortage
## ===========================  ====================================================
## HYPOTHALAMIC HORMONES                                                   0.4000000
## MENTAL HEALTH                                                           0.3793103
## PAIN                                                                    0.3520000
## HOSPITAL SOLUTIONS                                                      0.3429952
## ANTI ULCERANTS                                                          0.3250000
## OTHER THERAPEUTICS                                                      0.3235294
## NERVOUS SYSTEM DISORDERS                                                0.2735043
## GASTRO PRODUCTS                                                         0.2545455
## CORTICOSTEROIDS PLAIN COMBO                                             0.2352941
```

```
## ANTITHROMBOTICS                                                     0.2247839
## ==========================  ==================================================
```

Our third and fourth question of interest can be answered using poisson regression. We know that both variables ShortageFreq and ShortageTotalMonths are count variables because they are positively skewed and also a majority of the responses for each variable are zeroes. Thus each variable can be counted and both variables approximately follow a poisson distribution. We can confirm that each variable approximately follows a poisson distribution by looking at the distribution of the two variables. The histograms for both variables are printed below.

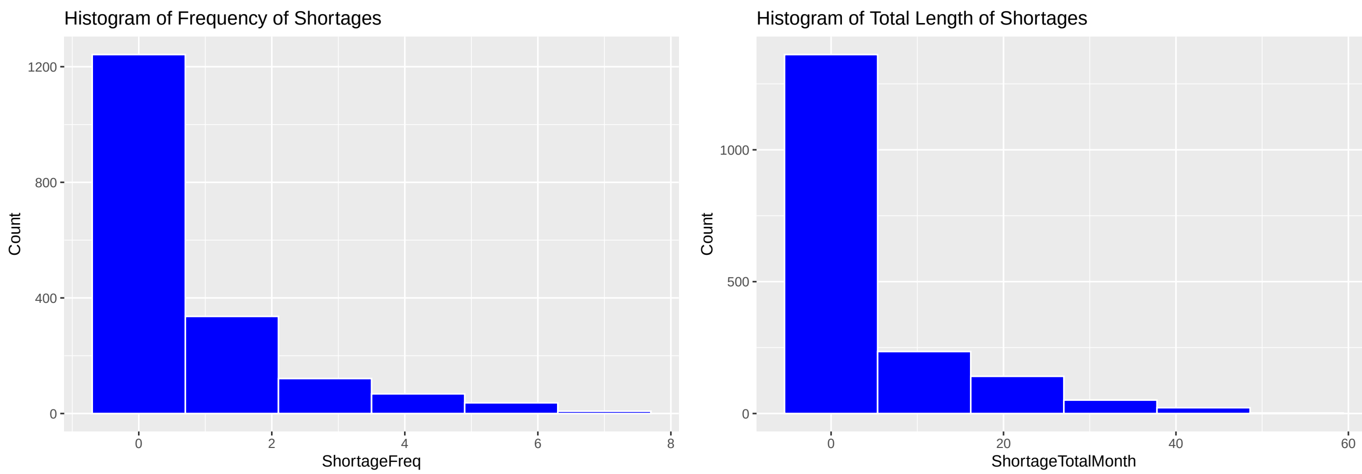


Figure 1: Histograms for ShortageFreq and ShortageTotalMonth

Going forward we can use poisson regression to predict the response of our two numeric response variables ShortageFreq and ShortageTotalMonths. This will help us understand the relationship that our predictor variables have with our response variables. This information will help Picket understand what to look out for when deciding which drugs it wants to manufacture. If there is any insight we can gain regarding which variables indicate that a certain drug is more likely to go into a longer shortage or will have more frequent shortages this information will be particularly useful because Picket's business model essentially relies on capitalizing on drug related supply issues. The more they know about drug attributes that tend to go into shortage, the better they will do as a business.

4 Results

The first of our questions has already been partially answered. We subset our data to only include the drug classes that had the highest proportion of drugs that had been in shortage over the 71 months included in the data. After making this subset we are ready to run our first model, logistic regression.

```
##
##     FALSE TRUE
##   0   316   10
##   1   105   22
```

Before we run the model we will split our 1812 observations into a training and testing set. Since we have sufficient number of observations we can do a 75/25 split which resulted in 1359 observations for training and 453 observations for testing. After running the logistic regression model and testing the model against our test data we received a confusion matrix as follows.

```
##
##      FALSE TRUE
##   0   316   10
##   1   105   22
```

We can see just from the confusion matrix just how unbalanced our predictions are. We have only 22 true positives and 316 true negatives. This result is not particularly surprising since the negative result (NDC_Shortage_FL=0) is the majority class, but our process of finding the drugs that had the highest proportion of shortages would have ideally given us a little more balance. The accuracy rate here is 74.61% which appears on the surface to be quite good, but this is of course heavily skewed by the true negatives. Looking at the false negative and false positive rate can give us a better picture of just how our model struggles to predict the minority class. Our false negative rate is a staggering 82.68%, while the false positive rate is only 3.07%. To make up for this difference we can change the classification threshold to make it easier to be classified as a true positive. To do this we would need to lower the threshold. This will, however, result in a lower accuracy rate and a higher false positive rate. We decided to change the classification threshold to 0.325 which gives the following confusion matrix.

```
##
##      FALSE TRUE
##   0   225  101
##   1    57   70
```

We can see here that the number of true positive predictions drops by 91 and the number of true negative predictions increases by 48. The false negative rate drops almost in half to 44.88% and the false positive rate increases to 30.98%. The accuracy rate also takes a hit and decreases to 65.12%. When creating a logistic regression model, it is tempting to maximize your accuracy rate and minimize your false negative and positive rates, and while it may be good to do so, the corresponding rates of a model are not necessarily how you gain the most information. Looking at the coefficients of the model is just as, if not more, insightful than its predictive ability.

```
##
##
## ====================================  ================================
## \                                     Logistic Regression Coefficients
## ====================================  ================================
## (Intercept)                                                 -1.6271730
## FormINJECTION                                                0.3848669
## FormPIGGYBACK                                                0.7788517
## FormSYRINGE                                                 -0.6273563
## FormVIAL                                                     0.4550131
## DrugClassANTITHROMBOTICS                                    -0.3656480
## DrugClassCORTICOSTEROIDS PLAIN COMBO                        -0.5659322
## DrugClassGASTRO PRODUCTS                                    -0.0666824
## DrugClassHOSPITAL SOLUTIONS                                  0.2083980
## DrugClassHYPOTHALAMIC HORMONES                               0.6711872
## DrugClassMENTAL HEALTH                                       0.3425159
## DrugClassNERVOUS SYSTEM DISORDERS                           -0.4326914
## DrugClassOTHER THERAPEUTICS                                 -0.0553536
## DrugClassPAIN                                                0.2650236
## LaunchMonth                                                 -0.0014079
## meanmanu                                                     0.0108209
## SpecPharmSPECIALTY                                          -0.0848960
## meansales                                                    0.0000027
## ====================================  ================================
```

We can look at the sign of each variable to better understand the relationship between the predictor variables and the response variable. For example, looking at the variables that our model deemed significant, we can see that when the variable Form equals to PIGGYBACK the estimated odds of a drug having experienced a shortage is multiplied by approximately 2.179, while all other variables are held constant. This means that drugs that are in the form of a piggyback are more likely to be classified as having had a shortage. On the other hand, when the variable Form equals to SYRINGE the estimated odds of a drug having experienced a shortage is multiplied by approximately 0.543, while all other variables are held constant. This means that drugs that are in the form of a syringe are less likely to be classified as having had a shortage and more likely to be classified as not having a shortage. Interestingly enough, most of the variables in our model are not statistically significant which makes the coefficients less reliable, but we did take a subset of the entire data to balance out the response variable, so it may be possible that some of these variables would be more significant while using all observations.

Moving on to our Poisson regression models we can similarly evaluate their ability to predict as well as discuss the corresponding coefficients of each model and how they relate to both response variables. First, for our model predicting the total number of months that each drug has been in shortage, our model was able correctly predict this number within 5 months of the true number 77.9% of the time. What is important in this model is that the coefficients can give us a better idea of how the predictor variables are associated with the response variable ShortageTotalMonth. These coefficients are listed below.

```
##
##
## =====================  ==============================================================
## \                      Poisson Regression Coefficients for Predicting TotalShortageMonth
## =====================  ==============================================================
## (Intercept)                                                                  1.2559123
## FormINJECTION                                                                0.0955796
## FormPIGGYBACK                                                                0.0553837
## FormSYRINGE                                                                 -0.1138232
## FormVIAL                                                                     0.4296361
## SpecPharmSPECIALTY                                                          -0.8410717
## CareClassCHRONIC CARE                                                       -0.4997274
## LaunchMonth                                                                 -0.0017400
## meanmanu                                                                    -0.1049102
## meansales                                                                    0.0000008
## =====================  ==============================================================
```

Looking at some of the significant variables, we can see that for the variable Form when this variable is equal to VIAL there is a 53% increase in the number of months that a drug will be in a shortage, while all other predictor variables are held constant. Another significant variable in the model is CareClass. We can see that when SpecPharm equals to CHRONIC CARE there is a 39% decrease in the number of months that a drug will be in a shortage, while all other predictor variables are held constant.

```
##
##
## =====================  ============================================================
## \                      Poisson Regression Coefficients for Predicting ShortageFreq
## =====================  ============================================================
## (Intercept)                                                               -0.7073057
## FormINJECTION                                                              0.1496209
## FormPIGGYBACK                                                              0.4788965
## FormSYRINGE                                                               -0.6150366
## FormVIAL                                                                   0.4215513
## SpecPharmSPECIALTY                                                        -0.1321563
```

```
## CareClassCHRONIC CARE                                                   -0.3752456
## LaunchMonth                                                             -0.0014105
## meanmanu                                                                -0.0259066
## meansales                                                                0.0000008
## =====================  ============================================================
```

Our last model is another poisson regression model to predict the number/frequency of shortages that a drug will experience. After creating our model and testing it on our test data, it was able to predict the number of shortages within 1 shortage of the true amount 68.87% of the time. Looking at some of the significant variables for predicting ShortageFreq, we can see that for the variable CareClass when this variable is equal to CHRONIC CARE there is a 31% decrease in the number of shortages that occur for a given drug, while all other predictor variables are held constant. Another variable that is significant in the model is Form when it equals to SYRINGE. We can see that when SpecPharm equals to CHRONIC CARE there is a 39% decrease in the number of shortages that a drug will experience, while all other predictor variables are held constant.

5 Conclusions and Appropriate Recommendations

Looking back at the objectives of our analysis, we were sufficiently able all of them. Our evaluative measures indicate that the predictor variables in the data provide at least some benefit to our predictions. While we may feel inclined to look only at these measures, what Picket should really grasp is the relationships that these variables have with shortages. False predictions will inevitably occur with any model, but these false predictions are using data from the past. Picket must learn about the relationships that were significant in the past to position themselves to be more successful in the future. One of these variables that was consistently significant across all models was Form. When Form equaled either SYRINGE or VIAL our models found that they were significant predictors. As a reminder, before conducting this analysis, we subsetted our data to include only the top 10 drugs classifications that had the highest proportion of drugs that experienced a shortage while also having a number of observations in the data greater than 20. Using this information, when deciding which drugs to manufacture, Picket should only look at these 10 drug classifications and look further into a drug if it is a Vial and avoid manufacturing a drug if it is a Syringe. The coefficients for the variable Form when it equals to VIAL is positive across all three models. This would seem to indicate that Vial based drugs are more likely to go into shortage. For syringes the opposite is true. The coefficients for all three models are negative for when FORM equals SYRINGE. This would seem to indicate that syringes are less likely to go into shortage and thus Picket should avoid manufacturing them. Further analysis could be conducted to confirm that these conclusions are valid. These conclusions are, of course, based only on the data we currently have. The relationships may very well have changed if the data is too old or if the data was collected during a particularly abnormal time (such as the during a pandemic). Ultimately there are several variables in this analysis that could be deemed to have a significant relationships with shortages, but it is up to Picket to determine if these variables are actually practically significant.

6 Appendix

Proportion of All Drugs that have experienced one shortage.

```
##
##
## ==============================================  =====================================================
## \                                               Proportion of Drugs that have Experienced a Shortage
## ==============================================  =====================================================
## MIOTICS ANTIGLAUCOMA PREPS                                                                 0.6000000
## ADHD                                                                                       0.5000000
## ANTITUBERCULARS                                                                            0.5000000
## SEX HORMONES                                                                               0.5000000
## HYPOTHALAMIC HORMONES                                                                      0.4000000
## MENTAL HEALTH                                                                              0.3793103
## DIAGNOSTIC EQUIP ACC ALLERGEN TESTS                                                        0.3750000
## PAIN                                                                                       0.3520000
## HOSPITAL SOLUTIONS                                                                         0.3429952
## ANTI PARASITICS ANTIMALARIALS INSECTICIDES                                                 0.3333333
## ANTI ULCERANTS                                                                             0.3250000
## OTHER THERAPEUTICS                                                                         0.3235294
## ALLERGY SYSTEMIC NASAL                                                                     0.2812500
## NERVOUS SYSTEM DISORDERS                                                                   0.2735043
## GASTRO PRODUCTS                                                                            0.2545455
## ANTIVIRALS HERPES                                                                          0.2400000
## CORTICOSTEROIDS PLAIN COMBO                                                                0.2352941
## ANTITHROMBOTICS                                                                            0.2247839
## CANCER DETOX AG ANTI NAUSEANTS                                                             0.2180851
## ANTIHYPERTENSIVES PLAIN COMBO                                                              0.2118959
## ANTIBACTERIALS                                                                             0.2076023
## LABOUR INDUCERS                                                                            0.1666667
## VITAMINS MINERALS                                                                          0.1525424
## ANTICOAGULANTS                                                                             0.1510791
## OTHER CARDIOVASCULARS                                                                      0.1463415
## BISPHOSPHONATES TUMOR RELATED BONY METASTASES                                              0.1176471
## SYSTEIMC ANTIFUNGALS                                                                       0.1153846
## BLOOD COAGULATION                                                                          0.1132075
## IMMUNOSUPPRESSANTS                                                                         0.1111111
## ONCOLOGICS                                                                                 0.1070932
## OTHER CNS                                                                                  0.0852018
## ANTI ANAEMICS IRON AND ALL COMBINATIONS                                                    0.0769231
## RESPIRATORY AGENTS                                                                         0.0769231
## THYROID ANTI THYROID AND IODINE PREPS                                                      0.0588235
## ANTIDIABETICS                                                                              0.0000000
## ANTIGOUT PREPS                                                                             0.0000000
## DIETETICS                                                                                  0.0000000
## HORMONAL CONTRACEPTION SYSTEMIC TOPICAL                                                    0.0000000
## IMAGING                                                                                    0.0000000
## MULTIPLE SCLEROSIS                                                                         0.0000000
## OSTEOPOROSIS                                                                               0.0000000
## OTHER ALIMENTARY TRACT AND METABOLISM PRODUCTS                                             0.0000000
## OTHER HAEMATOLOGICALS                                                                      0.0000000
## OTHER HORMONES                                                                             0.0000000
## OTHER RESPIRATORY                                                                          0.0000000
```

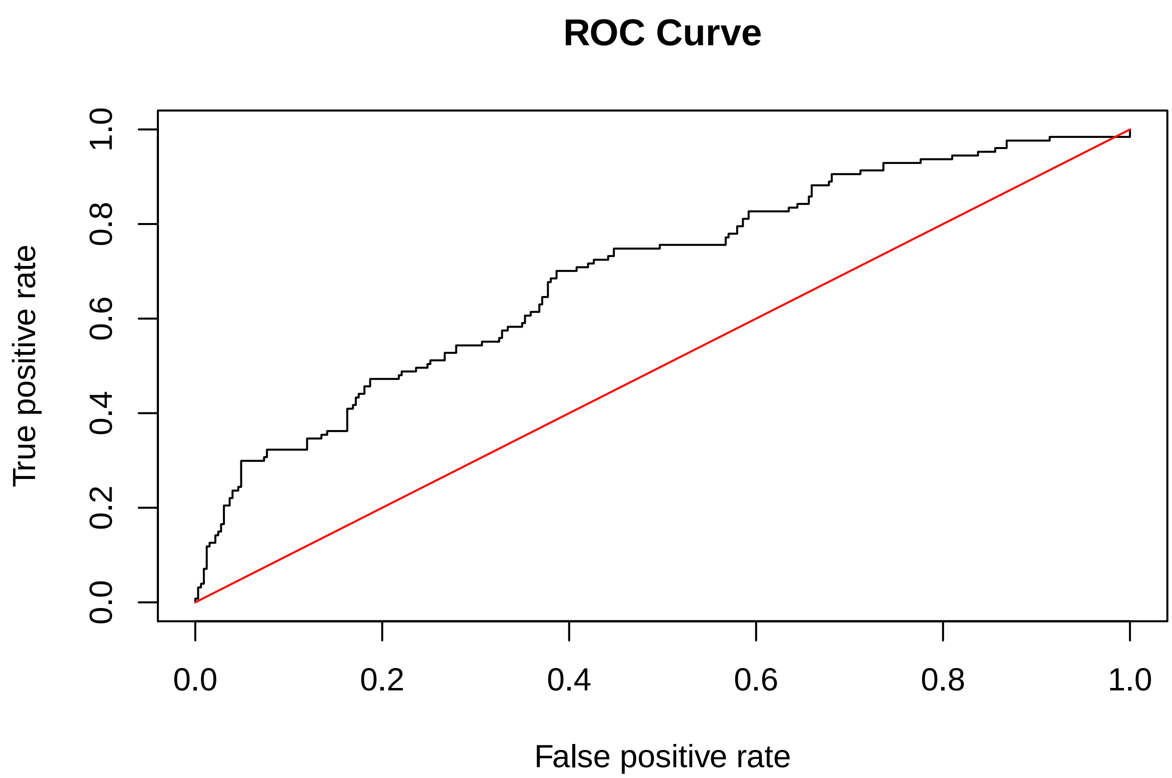
```
## POLYVAL IMMUNOGLOBLULINS IV IM                                                                                                                  0.0000000
## ==============================================  ======================================================
```

Logistic Regression Model Summary

```
##
## Call:
## glm(formula = NDC_Shortage_FL ~ Form + DrugClass + LaunchMonth +
##     meanmanu + SpecPharm + meansales, family = "binomial", data = topDC)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0661  -0.8531  -0.6587   1.1756   2.3955
##
## Coefficients:
##                                        Estimate Std. Error z value
## (Intercept)                          -1.627e+00  4.377e-01  -3.718
## FormINJECTION                         3.849e-01  3.502e-01   1.099
## FormPIGGYBACK                         7.789e-01  3.688e-01   2.112
## FormSYRINGE                          -6.274e-01  3.008e-01  -2.086
## FormVIAL                              4.550e-01  2.617e-01   1.739
## DrugClassANTITHROMBOTICS             -3.656e-01  4.033e-01  -0.907
## DrugClassCORTICOSTEROIDS PLAIN COMBO -5.659e-01  4.700e-01  -1.204
## DrugClassGASTRO PRODUCTS             -6.668e-02  4.838e-01  -0.138
## DrugClassHOSPITAL SOLUTIONS           2.084e-01  3.903e-01   0.534
## DrugClassHYPOTHALAMIC HORMONES        6.712e-01  8.251e-01   0.813
## DrugClassMENTAL HEALTH                3.425e-01  5.380e-01   0.637
## DrugClassNERVOUS SYSTEM DISORDERS    -4.327e-01  4.284e-01  -1.010
## DrugClassOTHER THERAPEUTICS          -5.535e-02  5.656e-01  -0.098
## DrugClassPAIN                         2.650e-01  3.659e-01   0.724
## LaunchMonth                          -1.408e-03  2.958e-04  -4.760
## meanmanu                              1.082e-02  3.818e-02   0.283
## SpecPharmSPECIALTY                   -8.490e-02  6.608e-01  -0.128
## meansales                             2.730e-06  3.216e-07   8.490
##                                      Pr(>|z|)
## (Intercept)                          0.000201 ***
## FormINJECTION                        0.271809
## FormPIGGYBACK                        0.034720 *
## FormSYRINGE                          0.037003 *
## FormVIAL                             0.082071 .
## DrugClassANTITHROMBOTICS             0.364570
## DrugClassCORTICOSTEROIDS PLAIN COMBO 0.228579
## DrugClassGASTRO PRODUCTS             0.890371
## DrugClassHOSPITAL SOLUTIONS          0.593405
## DrugClassHYPOTHALAMIC HORMONES       0.415981
## DrugClassMENTAL HEALTH               0.524358
## DrugClassNERVOUS SYSTEM DISORDERS    0.312458
## DrugClassOTHER THERAPEUTICS          0.922042
## DrugClassPAIN                        0.468819
## LaunchMonth                          1.94e-06 ***
## meanmanu                             0.776870
## SpecPharmSPECIALTY                   0.897771
## meansales                             < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2256.7  on 1811  degrees of freedom
## Residual deviance: 2049.5  on 1794  degrees of freedom
## AIC: 2085.5
##
## Number of Fisher Scoring iterations: 4
```

ROC Curve and AUC value




```
## [[1]]
## [1] 0.6952804
```

Poisson Regression Output for Predicting ShortageTotalMonth

```
##
## Call:
## glm(formula = ShortageTotalMonth ~ Form + SpecPharm + CareClass +
##     LaunchMonth + meanmanu + meansales, family = "poisson", data = train)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -7.598  -3.013  -2.421   0.486  13.075
```

```
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)           1.256e+00  5.650e-02  22.228  < 2e-16 ***
## FormINJECTION         9.558e-02  8.569e-02   1.115   0.2647
## FormPIGGYBACK         5.538e-02  8.334e-02   0.665   0.5063
## FormSYRINGE          -1.138e-01  6.513e-02  -1.748   0.0805 .
## FormVIAL              4.296e-01  5.767e-02   7.450 9.32e-14 ***
## SpecPharmSPECIALTY   -8.411e-01  1.109e-01  -7.584 3.36e-14 ***
## CareClassCHRONIC CARE -4.997e-01  3.996e-02 -12.506  < 2e-16 ***
## LaunchMonth          -1.740e-03  6.405e-05 -27.167  < 2e-16 ***
## meanmanu             -1.049e-01  7.307e-03 -14.358  < 2e-16 ***
## meansales             8.051e-07  2.942e-08  27.360  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 18281  on 1358  degrees of freedom
## Residual deviance: 16369  on 1349  degrees of freedom
## AIC: 18208
##
## Number of Fisher Scoring iterations: 7
```

Poisson Regression Output for Predicting ShortageFreq

```
##
## Call:
## glm(formula = ShortageFreq ~ Form + SpecPharm + CareClass + LaunchMonth +
##     meanmanu + meansales, family = "poisson", data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3709  -1.2134  -0.9998   0.4061   4.1450
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -7.073e-01  1.494e-01  -4.734 2.20e-06 ***
## FormINJECTION         1.496e-01  2.169e-01   0.690 0.490219
## FormPIGGYBACK         4.789e-01  1.953e-01   2.452 0.014208 *
## FormSYRINGE          -6.150e-01  1.855e-01  -3.316 0.000913 ***
## FormVIAL              4.216e-01  1.533e-01   2.750 0.005954 **
## SpecPharmSPECIALTY   -1.322e-01  2.001e-01  -0.660 0.508973
## CareClassCHRONIC CARE -3.752e-01  9.481e-02  -3.958 7.56e-05 ***
## LaunchMonth          -1.410e-03  1.580e-04  -8.927  < 2e-16 ***
## meanmanu             -2.591e-02  1.760e-02  -1.472 0.141120
## meansales             8.055e-07  7.166e-08  11.241  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2749.5  on 1358  degrees of freedom
## Residual deviance: 2448.5  on 1349  degrees of freedom
```

```
## AIC: 3646
##
## Number of Fisher Scoring iterations: 6
```

7 Bibliography

“Why Are Drug Patents Important: Everything You Need to Know.” UpCounsel, https://www.upcounsel.com/why-are-drug-patents-important.