To Determine the Effect of Variables on the Number of Complaints Received by the Doctors in an Emergency Department

By

Vinayak Kumar Gupta (24066272)

Student of Masters in Data Science, University of Western Australia

26th September 2023

ABSTRACT

This study is based on the analysis of complaints received by the 94 doctors working in the emergency department. Due to the front-line nature of the Emergency Departments and the high costs associated with this service, its performance is constantly being measured and reported. Complaints from ED patients may be used to better understand health service performance and utilised to improve patient care and satisfaction (Lawrence et al., 2018) [1]. Analysing the complaints doctors receive in the Emergency Department is like investigating why patients are unhappy with their care. We look at different variables and relate it to complaints received by doctors. Below is the short introduction of each variable and a subjective possible explanation of what role can this variable can play in complaints made by the patients. Detailed results of statistical analysis will be discussed later in the journal which will help us to identify the exact results and causes of the complaints. The variables being used are the Gender of doctor – this factor can play an important role in how patient perceives the treatment of the doctor, in my analysis I found out patients tend to prefer female doctors more. Another factor is the number of patient visits – high number of patients at a given time may increase waiting times for individual patients and also working hours of physicians which may effect patient satisfaction and doctor performance altogether. The residency training status is also an important factor to consider – doctors under training phase of their career are always under supervision, this may give different perspective to patients and some may not even want to be treated by newly made doctors. Also, those doctors have less field experience so chances of making the mistakes are not low even they are supervised. Then comes revenue earned by the doctor – we need analysis to check if doctors who make more revenue try to prioritise profits and earnings over patient health and satisfaction. For the total working hours of the doctor – longer working hours may lead to working stress and fatigue of the doctors which may impact performance, communication and analysis of doctors. By doing statistical analysis of these variables we can figure out what which factor plays a dominant role in making the patients upset and why. This helps hospitals make sure they have capable staff and doctors, provide better care, and reduce the chances of legal problems. It's about making sure patients get good care and feel satisfied when they come to the ER, especially when they're in urgent need of medical help.

In my analysis Residency, Gender and Revenue played an important role in the complaints received by the doctors. We will start by giving the context to this topic by analysing different research papers published in the past who have done similar research on complaints of doctors with different variables and study their results and comparing them with our own results which we got.

INTRODUCTION

The Emergency Department of the hospital is a place where people who need urgent medical help go and get help right away. This paper focuses on the number of complaints received for 94 doctors who worked in an emergency service at a hospital. There are many research papers published who analyzed the number of complaints received by the doctors with different variables. In this section we will explore different research papers and discuss their findings and methods used for analysis then compare with our results in the end.

Number Of Complaints Received

Complaint case rate was 1.17 per 1,000 visits *(Singapore Medical Journal – December 2007[2]).* For each complaint case, information required for analysis was entered into an Excel spreadsheet, coded and then analysed using the Statistical Package for Social Sciences version *12.0 (SPSS Inc, Chicago, IL, USA).*The complaints were organisation/logistics – waiting time, patient flow issues, lack of interim care (49.0 %), communication – rudeness/insensitive remarks, poor communication and attitude (26.0 %), standard of care – misdiagnosis, inappropriate treatment/examination (22.9 %) and other issues (1.3 %). *(Table-1)* Most standard of care (76.0 %) and half of organisation/logistics complaints (46.8 percent) were not valid. Most communication complaints were valid (73.7 percent). Most complaints (82.8 percent) were resolved with an explanation/apology. Age group specific and triage-specific complaint rates were highest among adult patients and among priority 3 patients, respectively; ethnic group and gender-specific complaint rates were highest among Chinese patients and among female patients, respectively.[2]

Number of Patient Visits

Although frequent ED users represent only a very small percentage of visits, they consume health care costs disproportionate to their numbers. This study explored characteristics of frequent ED users at a large Midwestern urban hospital and factors predictive of high ED utilization. The sample included adult patients with at least 6 visits in 2005-2006 (N = 201). For each, 6 visits were randomly chosen for chart review (N = 1200 visits) of demographic, health history, and clinical factors such as chief complaints. Frequent users were commonly female, 35 years old, white, single, unemployed, living alone, with private insurance/ Medicaid and a primary care physician. Top chief complaints were abdominal pain, headache, chest pain, low back pain, and lower extremity pain. However, a Poisson regression found that the following characteristics were associated with a higher number of ED visits: male, non-Black race, part-time employment, retired/ unemployed, having Medicare, and having a chief complaint of upper respiratory infection. Headache approached significance as an independent predictor of more visits.[3]

Residency Training Status

The research aimed to assess resident doctors' performance using two outcome variables: number of attempts at Part One examination before success and delay from the commencement of residency to passing Part One examination in the respective postgraduate medical colleges. Electronic questionnaires was distributed to respondents through their contacts or emails. Data was analysed using mean, standard deviation, simple tables as well as *t*-test and Chi-square test.

The Many works of literature have found stress and mental burnout as common problems affecting doctors in residency training due to the demanding nature of their program. These have negative effects on their health, well-being, interactions, learning as well as performance in residency training however, only a few studies have tried to access the performance of resident doctors in training. Although not significantly associated with success in Part One examinations, the study found that most resident doctors were unable to read consistently, do experience mental exhaustion after work, do not have relevant literature in their departmental library, and do not receive prompt and adequate sponsorship in residency training. These factors do have negative effects on various aspects of their performance in residency training, as highlighted by previous studies. [4]

Gender of the Doctor

A survey was conducted with ordered logistic regressions for women and men to determine the unique association of physician gender with patient ratings of 5 interpersonal aspects of care, their trust of the physician, and their overall ratings of the physician, controlling for patient age, health status, language and interpreter status, literacy level, and expected satisfaction. Female patients trusted female physicians more ($P = .003$) than male physicians and rated female physicians more positively on the amount of time spent ($P = .01$), on concern shown ($P = .04$), and overall ($P = .03$). Differences in ratings by female patients of male and female physicians in terms of friendliness ($P = .13$), respect shown ($P = .74$), and the extent to which the physician made them feel comfortable ($P = .10$) did not differ significantly. Male patients rated male and female physicians similarly on all dimensions of care (overall, $P = .74$; friendliness, $P = .75$; time spent, $P = .30$; concern shown, $P = .62$; making them feel comfortable, $P = .75$; respect shown, $P = .13$; trust, $P = .92$). Having a female physician was positively associated with women's satisfaction, but physician gender was not associated with men's satisfaction [5].

Revenue Earned by the Doctor (dollars)

It was difficult to find papers who analysed directly the complaints of the doctor with their income level, hence I divided my analysis for this variable into two parts. Firstly, I researched on how increase in revenue effects the doctor's job satisfaction level and then how job satisfaction level effects doctor's performance level. And naturally, the better the performance of the doctor, the more satisfied are the patients and lesser the complaints.

Findings have shown that pay, recognition, promotion opportunities, and meaningful work are factors of compensation management which have direct effect on job satisfaction on doctors. The Pearson value of 0.529 shows the strong positive co-relation between pays and job satisfaction. Employees with above 60000 incomes reflect the high level of job satisfaction with a mean of (2.60) and standard deviation of (.879) than the people with 10,000-20,000 income and 21000-40000 income with mean value of 2.96 and 3.05 respectively.[6]

Another study measuring job satisfaction level with their performance was done. The result were as follows - Data were collected using a self-administered, anonymous questionnaire containing 35 questions about job satisfaction, personal and professional variables, and three open questions. the average score of job satisfaction to be 42.10 and that for performance to be 87.52. The overall satisfaction was 32.4%. The respondents' were less satisfied in domains of survival or personal maintenance, security, and recognition but more satisfied in domains of status, companionship, and quality or style of supervision. The relationship between satisfaction and performance, though not significant, but satisfied persons had about 3 times better performance than the unsatisfied persons had.[7]

Working Hours of a Doctor

The average number of consecutive hours worked was 32.75, with a maximum of 39 h. Total number of hours slept and longest period of consecutive sleep ranged from 0 to 4 h. The average change in attention scores from rested to post call. was 25.03, indicating an average decrease of 4.8%. Information processing speed showed the greatest average change from rested to post call, representing a reduction in information processing speed of 7.2%. Scores in motor skills showed an average change of 25.67 from the rested to the post-call state, indicating a 5.5% decrease. Clinical decision-making scores were, on average, 15.2% lower post call than when rested, while time taken to complete clinical decision-making problems was greater by 10% post call than when rested. Thus a doctor who is working an on-call shift and whose attentional resources are decreasing may find it increasingly difficult to focus and hence may miss critical information.**[8]**

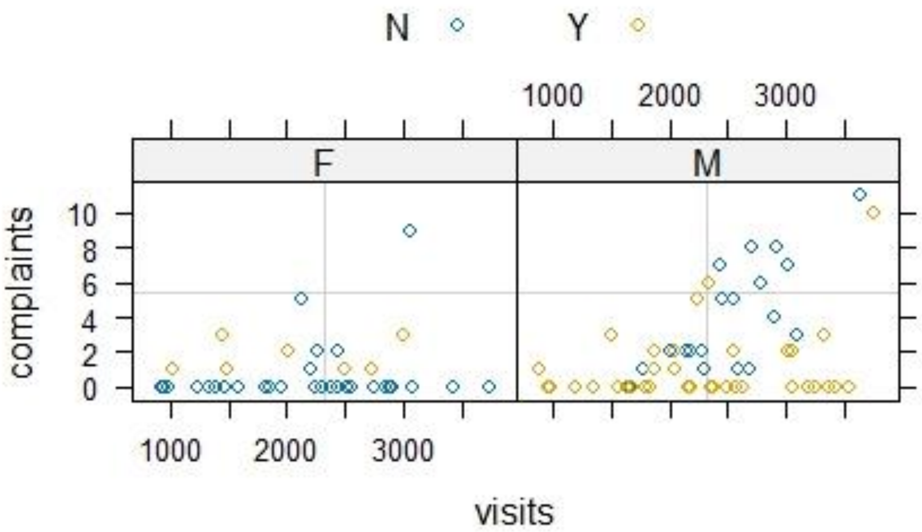
Here we saw how different variables that we are going to analyse effects a doctor in Emergency Department or in general. In my model I have used Visits, Residency, Gender, Revenue as individual variables along with some interaction terms such as residency : gender, residency : revenue, residency : hours, gender : revenue. In the upcoming section we will check the methodology used for my analysis followed by the results which I got. Then we will compare those results with the results of other research papers discussed here.

Methodology

I first started with looking out at the data. Here our initial dataset is 94 observation of 6 variables like visits, complaints, residency, gender, revenue, hours. I converted gender and residency variables to factors and visits , complaints to numericals so system can analyse each of them as categories and numbers respectively. I then looked at the overall summary which gave me a rough of distribution of data and potential outliers. Relation between different variables was also checked using Corelation and Boxplot functions. Now to check the normal distribution of complaints received I made three different histograms based on complaints received, gender complaints, residency training. Different types multi- variable plots were also made . First plot included complaints, visits, gender, residency status. Second plot is a 3D plot which included Complaints, Hours and Revenue. Through histograms I found out zero inflation in all categorical variables. To check the dispersion values an initial Poisson model was made with all the interaction terms included. 15 terms were there in total with only 3 significant variables and an AIC of 317.8. After doing step AIC and removal of non-significant variables through multiple trial and errors, a model with AIC = 316.76 with 9 variables and 4 significant variables was finalised. This was named *finalmodel.pois.* Then it's Pearson's residual v/s fitted graphs were checked though two different plots. And finally a rootogram was made to analyse zero inflation problems for this model. To make improvements a quasi-poisson model was made to deal with dispersion problems. Here only values of standard errors and dispersion values changed and rest model remained same. Rootogram was not applicable in this method. To improve my model more I made a negative binomial model. It was a good model with a step AIC of 288.84. Dispersion Parameter improved drastically. And residual v/s fitted plots more or less remained the same in all models. Rootogram improved a bit with lesser zero's in underfitting section.
Once I fixed the issue with data dispersion, I faced another challenge: too many zero values. To solve this, I created two zero inflation models – one using a Poisson distribution and the other with a negative binomial distribution. However, during the model creation, I faced computational errors. Some sections of both models displayed NaN values, but this didn't have a significant

impact because these sections had very small coefficients. In the big picture, these NaNs didn't disrupt my analysis.
I also used a rootogram to visualize the excessive zeros, and the models showed improvement. The extra zeros became closer to the zero line. To choose the best model, I compared the five models based on several factors, including their AIC values, standard deviations, residuals, p-values, and dispersion parameters. Ultimately, I finalised a zero-inflated Poisson model named *finalmodel.zerinf*, which had a step AIC value of 262.56. This model was the most suitable for my analysis, taking into account all the considerations and improvements made throughout the process. Then final model equation was made and interpreted. I then compared my results with results in other research papers as discussed in the introduction section.

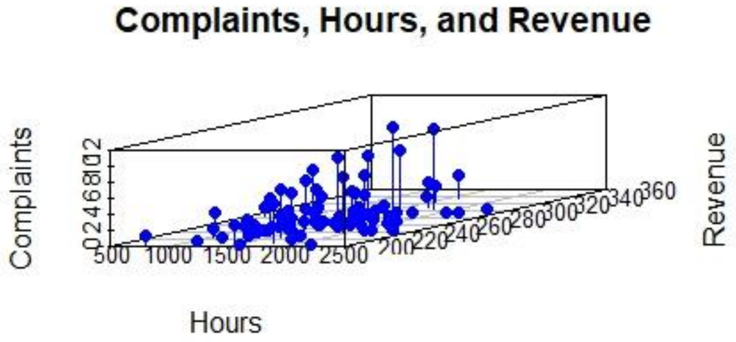
Results



Comparison chart show relation between Number of Complaints received by the doctor with the number of visits by patients differentiated by the gender and residency status of the doctor.

The above graph shows the relationship between four variables out of which two are categorical. that majority of the complaints are received for the male doctors (showed under M box). And those doctors are not in residency training. Also, as the number of visits increase, complaints increases, especially for the male doctors.

The Pearson Residuals, which are a measure of the difference between observed and expected values in a statistical model, provide valuable insights into the goodness of fit of a model. In this case, the minimum Pearson Residual is -1.3723, the median is -0.2624, and the maximum is 3.8521. These values indicate how well the model aligns with the observed data. The AIC (Akaike Information Criterion) is a measure of how well the model explains the data while considering model complexity. In this instance, the AIC is 262.561, and it is typically used for model comparison. A lower AIC indicates a better-fitting model among competing models.The DOF (Degrees of Freedom) in the model is 20, which is the number of parameters estimated by the model. It helps determine the model's complexity and flexibility. The Dispersion value of 2.681 indicates how much the observed data varies from the model's predictions. The Log-likelihood is a measure of how well the model fits the data. In this case, it is -111.3. A higher log-likelihood suggests a better fit.



This figure shows the relationship between three continuous variables, Complaints, Hours and Revenue

The figure shows that as the working hours increase, revenue earned by the doctors also increase but shows no significant effect of complaints on hours. The complaints almost remain same when hours increase apart from some few outliers. Interesting thing is that when the revenue increase, number of complaints increase, majority of them spike in the range of 240 -300 dollars.

Summary of Dataset

Visits – number of patient visits	Min: 879 , Median : 2299 , Mean : 2271 , Max. : 3763
Complaints – no. of complaints in previous year	Min: 00 , Median : 00 , Mean : 1.564 , Max. : 11
Residency – doctor in residency training (Y=yes or N = no)	N = 49 , Y =45
Gender – M= male, F = female	F = 37 , M = 57
Revenue – Doctor's hourly income (dollars)	Min: 203.9 , Median : 263.7 , Mean : 263.8 , Max. : 342.9
Hours – number of hours a doctor worked in a year	Min: 589 , Median : 1494 , Mean : 1469 , Max. : 2269

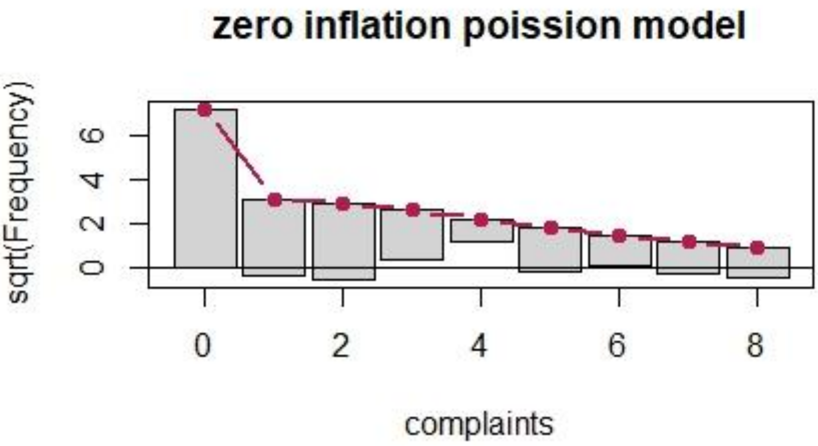
Modelling Results of FinalModel – zeroinflation

When I finalised my model, I got results in terms of poison with log link and Binomial with logit link. I will explain the result in both the terms.

Zero inflation model for Poisson with log link and Binomial with logit link :

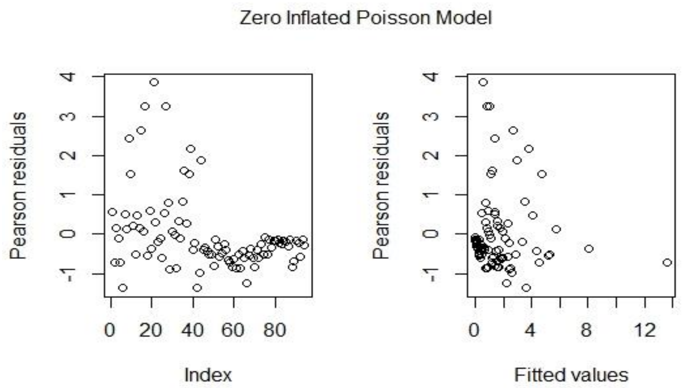
Variables	Coefficient - log link	P-value,(loglink)	Coefficient - binomial logit link	P-value binomial-logit link
Visits	0.0014395	NaN	4.61e -04	0.726
Residency Y	3.9213305	0.000127	-1.60e +01	0.989
Gender M	1.3448850	0.622472	4.910e + 00	0.996
Revenue	-0.0007202	0.948470	4.980e -02	0.168

ResidencyY:GenderM	0.4154154	0.445589	2.689e +01	0.906
ResidencyY:revenue	-0.0066686	0.296982	-6.521e -03	0.999
ResidencyN:hours	0.0003002	NaN	-8.432e -04	0.714
ResidencyY:hours	-0.0014475	0.003214	1.939e -03	0.409
GenderM:revenue	-0.0054050	0.632940	-7.520e -02	0.985
Pearson Residuals	Min (–1.3723) Med. (–0.2624) Max (3.8521)	AIC – 262.561 DOF - 20	Dispersion – 2.681	Log-likelihood -111.3



Rootogram for Zero-inflation Poisson Model

In this plot, each bar represents a data category, and the height of the bar corresponds to the difference between observed and expected frequencies. This rootogram is more overly fit than underly fit as almost all the bars are above zero line. In the first bar, there are many zero's, the condition is called Zero Inflation and zero-inflation models are used to fit overly inflated graphs. But this is the best rootogram which I could get after fitting my dispersions and zero's. When I compared to initial rootograms, I can say that there is an improvement and this rootogram is fairly fit.



The above figure shows Residuals v/s Index and Residuals v/s Fitted Values plots.

Here in the Residual v/s Index chart we can observe potential outliers in the range 0 - 4 in the in y-axis. Apart from that it does not show heteroscedasticity and looks fine. In the Residual v/s

Fitted values chart there are some outliers in the range of 1 -4 in y-axis and 6 -14 in x -axis. Rest looks evenly distributed. Hence it suggests that the assumptions of constant variance and linearity are met.

1) **Zero Inflation Model – Poison with log link**

*λ = exp(-2.712 + 0.0014 * visits + 3.921 *residencyY + 1.344 * genderM - 0.0007 * revenue + 0.4154 * residencyY:genderM - 0.0066 * residencyY:revenue + 0.0003 * residencyN:hours - 0.00144 * residencyY:hours - 0.0054 * genderM:revenue)*

In this equation, the λ represents the expected count of complaints, and each coefficient is associated with a specific predictor variable or interaction term, with a plus sign (+) indicating their contribution to the expected count.

2) **Zero Inflation Model – Binomial with logit link**

logit(p) = -10.79 + 0.0004 * visits - 16.04* residencyY + 4.91 * genderM + 0.049 * revenue + 26.89* residencyY:genderM – 0.0065 * residencyY:revenue – 0.00084 * residencyN:hours – 0.00013* residencyY:hours - 0.075* genderM:revenue

Where "logit(p)" represents the log-odds of excess zeros (i.e., the likelihood of having zero counts) for the binomial distribution.

For my case study analysis I will use only Equation(1) for interpreting the results and then compare those results with other research papers in the upcoming section.

Discussion

From the equation (1) I got the following findings -

1. Intercept (-2.712): This is the expected count of complaints when all predictor variables are zero. It's negative, showin a lower expected count when all other factors are at their reference levels.

2. Visits (0.0014): For every one-unit increase in the number of visits, the expected count of complaints increases by a factor of exp(0.0014) i.e 1 times. This suggests a small positive effect of the number of visits on the count of complaints.

3. residencyY (3.921): If the doctor is in residency training , the expected count of complaints increases by a factor of exp(3.921) compared to when the doctor is not in residency i.e it's chances increases by 50.45 times. This indicates a substantial positive effect of being in residency.

4. genderM (1.344): If the doctor is male the expected count of complaints increases by a factor of exp(1.344) i.e 3.83 times , compared to being female.

5. revenue (-0.0007): For every one-unit increase in revenue, the expected count of complaints decreases by a factor of exp(-0.0007) i.e 0.999 times. This indicates a small negative effect of revenue on the count of complaints.

Interaction Terms

6. residencyY:genderM (0.4154): This interaction term indicates the combined effect of being in residency and being male. A positive coefficient suggests that, when a doctor is both in residency and male, the expected count of complaints increases by a factor of exp(0.4154) i.e 1.51times compared to doctors who are not in residency and female.

7. residencyY:revenue (-0.0066): This interaction term represents the combined effect of being in residency and the doctor's revenue. The negative coefficient suggests that when a doctor is in residency and their revenue increases by one unit, the expected count of complaints decreases by a factor of exp(-0.0066) i.e 1 times.

8. residencyN:hours (0.0003): This interaction term involves doctors who are not in residency and their working hours ("hours"). The positive coefficient indicates that, for doctors not in residency, for every one-unit increase in working hours, the expected count of complaints increases by a factor of exp(0.0003) i.e 1 times.

9. residencyY:hours (-0.00144): This interaction term represents the combined effect of being in residency and working hours. The negative coefficient implies that, for doctors in residency, for every one-unit increase in working hours, the expected count of complaints decreases by a factor of exp(-0.00144) i.e 0.99 times.

10. genderM:revenue (-0.0054): This interaction term combines the effects of being male and the doctor's revenue. The negative coefficient indicates that, for male doctors, for every one-unit increase in revenue, the expected count of complaints decreases by a factor of exp(-0.0054) i.e 0.99 times.

So in overall, only residencyY , gender M , residenyY:genderM have significant impact on complaints received by the doctors.

Comparison

In our discussion of different research papers, variable residency Y matched with research paper[4] but with a different interpretation. It was concluded there, that doctors in Residency training have more stress and work which results in decrease in performance. Hence decrease in performance will increase chances of making more mistakes and hence more complaints.

With gender M , our result is similar with research paper [5] where it was concluded that female patients have preferences for female doctors but male do not have any specific preferences. Although males scored more to female doctors in terms of behaviour, friendliness etc. Hence it can be concluded that gender of a doctor somehow effects on complaints of a patient.

Apart from that , one individual variable Revenue shows a negative relationship with complaints, which matches with the result in research papers in [6] and [7]. The conclusion was that revenue increased the job satisfaction level of doctors which improved performance of doctors by 3 times despite they being not significantly related to each other. Same is in our case, the more the revenue, the better the performance and lesser the chances of having complaints. And in our case also revenue is not significantly related to complaints.

References

[1] Lawrence, P. *et al.* (2018) 'Wait times are not the problem! detailed analysis of unsolicited patient complaints from a metropolitan Australian Emergency Department', *Emergency Medicine Australasia*, 30(5), pp. 672–677. doi:10.1111/1742-6723.12975.

[2] Wong, L & Ooi, Shirley & Goh, Lee Gan. (2007). Patients' complaints in a hospital emergency department in Singapore. Singapore medical journal. 48. 990-5.

[3] Lawrence, P. *et al.* (2018) 'Wait times are not the problem! detailed analysis of unsolicited patient complaints from a metropolitan Australian Emergency Department', *Emergency Medicine Australasia*, 30(5), pp. 672–677. doi:10.1111/1742-6723.12975.

[4] Umar SS, Babandi ZS, Suleiman AG, Umar UM, Olorukooba AA, Ahmad AJ, *et al.* Assessing performance of resident doctors in training in Northwestern Nigeria. Niger J Med 2021;30:264-70.

[5] (2001) *Does physician gender affect satisfaction of men and women visiting the emergency department?* [Preprint]. doi:10.7249/rp1189.

[6] Yaseen, A. (2013) 'Effect of compensation factors on employee satisfaction- A study of doctor's dissatisfaction in Punjab', *International Journal of Human Resource Studies*, 3(1), p. 142. doi:10.5296/ijhrs.v3i1.3351.

[7] Jabbari H, Pezeshki MZ, Naghavi-Behzad M, Asghari M, Bakhshian F. Relationship between job satisfaction and performance of primary care physicians after the family physician reform of east Azerbaijan province in Northwest Iran. Indian J Public Health 2014;58:256-60.

[8] Fiona Flinn, Claire Armstrong, Junior doctors' extended work hours and the effects on their performance: the Irish case, *International Journal for Quality in Health Care*, Volume 23, Issue 2, April 2011, Pages 210–217, https://doi.org/10.1093/intqhc/mzq088

Appendix

The dataset was named complaints.

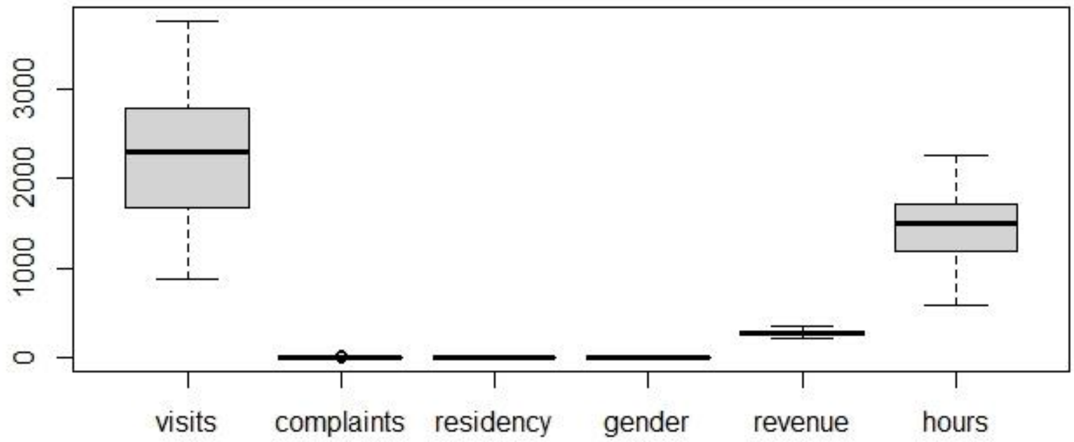
```{r}

# Variables such as gender , residency , visits , complaints were converted as factors and numerics

complaints$gender <-  factor(complaints$gender)
complaints$residency <-  factor(complaints$residency)
complaints$visits <-  as.numeric(complaints$visits)
complaints$complaints <-  as.numeric(complaints$complaints)

summary(complaints)

```

The results were –

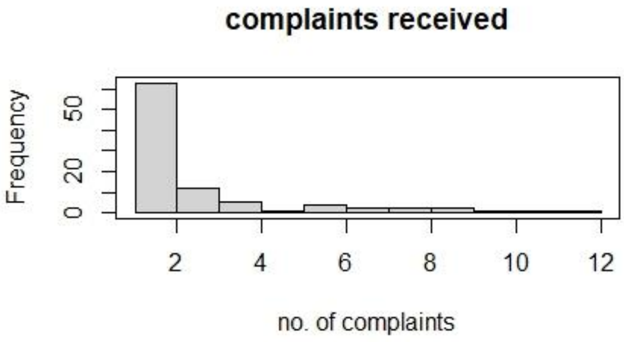
```
     visits        complaints      residency gender    revenue          hours
 Min.   : 879   Min.   : 1.000   N:49      F:37    Min.   :203.9   Min.   : 589
 1st Qu.:1698   1st Qu.: 1.000   Y:45      M:57    1st Qu.:243.8   1st Qu.:1201
 Median :2299   Median : 1.000                     Median :263.7   Median :1494
 Mean   :2271   Mean   : 2.564                     Mean   :263.8   Mean   :1469
 3rd Qu.:2776   3rd Qu.: 3.000                     3rd Qu.:288.0   3rd Qu.:1700
 Max.   :3763   Max.   :12.000                     Max.   :342.9   Max.   :2269
```

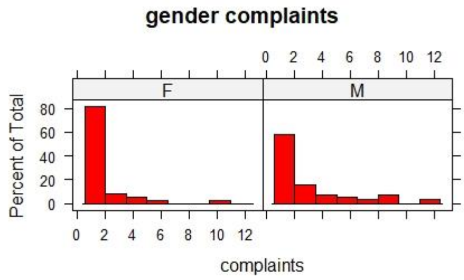
To check some corelation between variables boxplot was made. We got a rough idea how they are related to each other.

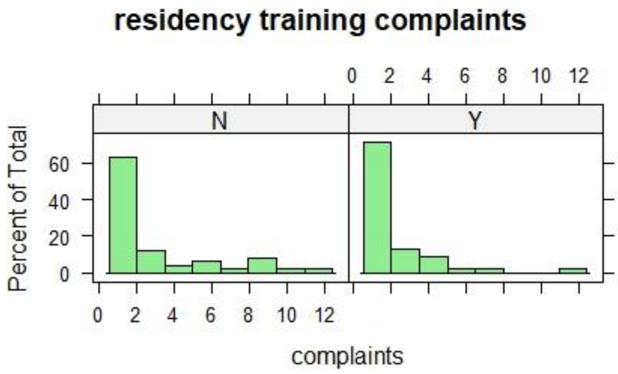
```{r}

boxplot(complaints )

```

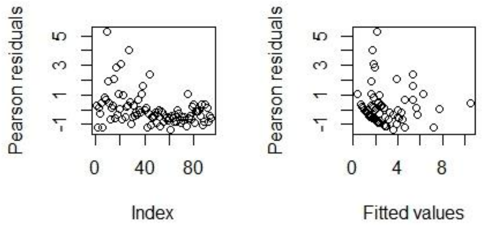



```{r}

# Checking for zero - inflation  model between the variables. From this code I came to know we have zero inflation problem

hist(complaints$complaints , main = "complaints received" , xlab = "no. of complaints" )
box()
histogram(~complaints | gender , complaints, main = "gender complaints" , col = "red" )
histogram(~complaints | residency , complaints, main = "residency training complaints" , col = "lightgreen" )

```









Then a model with all the interaction terms was made and residual v/s fitted plots were checked. Th residual v/s fitted and index charts were same for all the models hence showing only once.

```{r}

library(MASS)

complaints.pois <- glm(complaints ~ .^2, data = complaints, family = poisson)
summary(complaints.pois)
```

```
Call:
glm(formula = complaints ~ .^2, family = poisson, data = complaints)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)         6.158e+00  3.690e+00   1.669  0.09515 .
visits             -1.492e-03  1.086e-03  -1.375  0.16925
residencyY          4.680e-01  1.579e+00   0.297  0.76684
genderM            -2.144e+00  1.821e+00  -1.178  0.23893
revenue            -2.630e-02  1.400e-02  -1.879  0.06030 .
hours               1.439e-04  2.349e-03   0.061  0.95115
visits:residencyY  -2.029e-04  4.348e-04  -0.467  0.64074
visits:genderM      1.852e-04  4.481e-04   0.413  0.67948
visits:revenue      6.959e-06  3.953e-06   1.760  0.07837 .
visits:hours        1.093e-08  3.105e-07   0.035  0.97191
residencyY:genderM -1.143e+00  3.872e-01  -2.953  0.00315 **
residencyY:revenue  8.042e-03  5.902e-03   1.363  0.17303
residencyY:hours   -1.152e-03  8.716e-04  -1.322  0.18617
genderM:revenue     6.470e-03  6.432e-03   1.006  0.31440
genderM:hours       6.424e-04  8.622e-04   0.745  0.45626
revenue:hours       2.677e-07  8.967e-06   0.030  0.97619
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 174.689  on 93  degrees of freedom
Residual deviance:  86.699  on 78  degrees of freedom
AIC: 358.75

Number of Fisher Scoring iterations: 5
```

After checking the following plots – it was decided that this model and upcoming models meet all the assumptions , i.e. The observations are independent of each other , The relationship between the predictors and the expected value of the dependent variable is linear, Cinstant Variance (Homoscedasticity) is present except from some outliers in the initial part of the plot. Overdispersion was still there.

After stepAIC we ade a poisson model and got the details of deviation and dispersion.

```{r}

library(lattice)
library(AER)

finalmodel.pois <- glm(formula = complaints ~ visits + residency + gender + revenue
      + residency:gender + residency:revenue +
   residency:hours + gender:revenue, family = poisson, data = complaints)

summary(finalmodel.pois)

dispersiontest(finalmodel.pois)

```

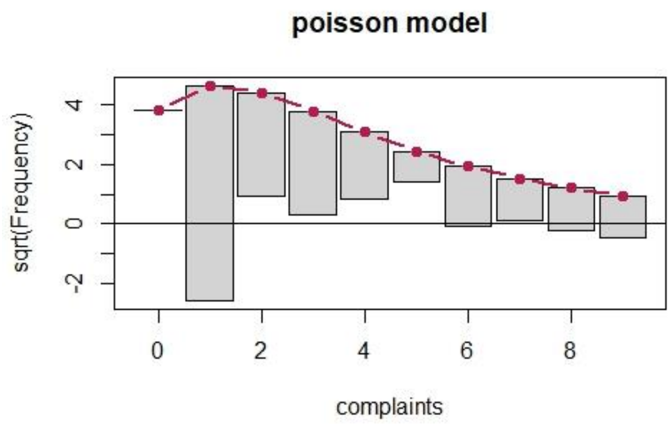
```
Call:
glm(formula = complaints ~ visits + residency + gender + revenue +
    residency:gender + residency:revenue + residency:hours +
    gender:revenue, family = poisson, data = complaints)

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.8871215  1.3001842   0.682 0.495046
visits              0.0004308  0.0001247   3.454 0.000551 ***
residencyY          0.9474937  1.3167461   0.720 0.471789
genderM            -1.3830140  1.3547521  -1.021 0.307320
revenue            -0.0085489  0.0047891  -1.785 0.074248 .
residencyY:genderM -1.5244736  0.3267959  -4.665 3.09e-06 ***
residencyY:revenue  0.0042003  0.0044342   0.947 0.343514
residencyN:hours    0.0004679  0.0003375   1.386 0.165634
residencyY:hours   -0.0004728  0.0003418  -1.383 0.166545
genderM:revenue     0.0095871  0.0053101   1.805 0.071007 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 174.689  on 93  degrees of freedom
Residual deviance:  94.735  on 84  degrees of freedom
AIC: 354.78

Number of Fisher Scoring iterations: 5
```

Dispersion and deviance values

```
    Null deviance: 174.689  on 93  degrees of freedom
Residual deviance:  94.735  on 84  degrees of freedom
AIC: 354.78

Number of Fisher Scoring iterations: 5


        Overdispersion test

data:  finalmodel.pois
z = 0.79199, p-value = 0.2142
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
   1.23369
```

Rootogram of Poisson Model

```{r}

library(countreg)
rootogram(finalmodel.pois , main = " poisson model ")
box()

```



Then quasi-poisson model was made to deal with dispersion. Everything was same except changes in standard errors.

```{r}

finalmodel.qpois <- glm(formula = complaints ~ visits + residency + gender + revenue
    + residency:gender + residency:revenue +
   residency:hours + gender:revenue, family = quasipoisson, data = complaints)
summary(finalmodel.qpois)

```

```
(Dispersion parameter for quasipoisson family taken to be 1.40943)

    Null deviance: 174.689  on 93  degrees of freedom
Residual deviance:  94.735  on 84  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

Similiarly a negative binomial model.

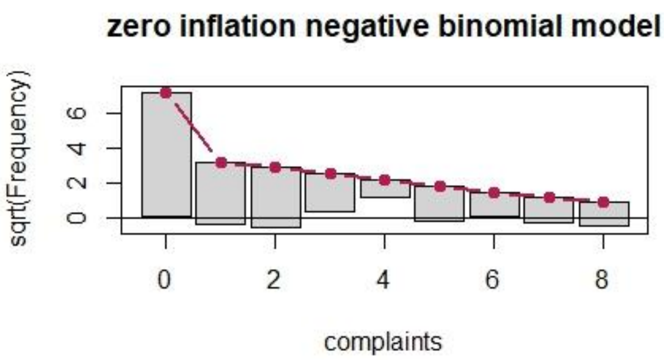
```{r}
finalmodel.nb <- glm.nb(formula = complaints ~ visits + residency + gender + revenue
    + residency:gender + residency:revenue +
   residency:hours + gender:revenue, data = complaints)
summary(finalmodel.nb)

```

```
(Dispersion parameter for Negative Binomial(11.5344) family taken to be 1)

    Null deviance: 136.545  on 93  degrees of freedom
Residual deviance:  73.848  on 84  degrees of freedom
AIC: 353.74

Number of Fisher Scoring iterations: 1


              Theta:  11.53
          Std. Err.:  7.96

 2 x log-likelihood:  -331.737
```

Then I made zero-inflation model which is described in the above report.

 # After that zero-inflation negative binomial model was made which was almost same in every aspect with zero inflation with a little difference in rootogram.

```{r}

library(countreg)
rootogram(finalmodel.zerinfnegbin , main = " zero inflation negative binomial model ")
box()

```



 # For comparing the best model I compared stepAIC of those models which was the final step for my model selection of zero inflation model.

```{r}

AIC(finalmodel.nb,finalmodel.pois,finalmodel.qpois,finalmodel.zerinf,finalmodel.zerinfnegbin)

```

Description: df [5 x 2]

	df <dbl>	AIC <dbl>
finalmodel.nb	11	353.7367
finalmodel.pois	10	354.7842
finalmodel.qpois	11	NA
finalmodel.zerinf	20	262.5614
finalmodel.zerinfnegbin	21	264.5185

5 rows