Zero-Inflated Negative Binomial Regression Project

Abstract

Patients' complaints serve as critical indicators of healthcare quality, revealing systemic and individual factors that impact patient satisfaction. Past research has identified variables such as communication skills, physician workload, and demographic factors like gender and experience as predictors of complaint rates. Using a dataset containing demographic information on 94 doctors, this study examines the relationship between complaints and variables including patient visits, residency status, gender, revenue, and hours worked. Initial analysis indicated overdispersion in the complaint data, prompting the use of a Zero-Inflated Negative Binomial (ZINB) model to account for the high frequency of zero complaints and variation in complaint counts. Our model revealed that patient visit volume, revenue, and hours worked significantly influence complaint rates. Specifically, increased patient visits were associated with a higher likelihood of complaints, while higher revenue and more hours worked correlated with fewer complaints. Additionally, male physicians not in residency training were found to have higher complaint rates compared to their female or resident counterparts.

Section 1: Introduction

An emergency department (ED), also known as an accident and emergency department (A&E), is a medical centre specializing in the treatment of immediate acute illnesses and injuries. (Cooke et al., 2020) It emerged from the need for rapid, specialized care for critical conditions such as heart attacks, strokes, and severe infections. With the rise in time-sensitive treatments for conditions listed above, emergency care has become essential in major hospitals, where EDs often operate 24 hours a day to meet patient needs. Emergency departments are high-pressure environments with large patient volumes, limited resources, and a fast-paced setting (Cooke et al., 2020). This atmosphere can lead to longer wait times, rushed interactions, and inattentiveness, all of which may contribute to patient dissatisfaction with the hospital or its doctors.

A major area of concern for healthcare quality management is complaints against physicians in hospital emergency department, which have an impact on patient outcomes and the hospital's image. Patient complaints are frequently regarded as important markers of the quality of healthcare since they provide insight into organisational flaws, systemic problems, and the competence of individual doctors. Communication problems, perceived mistakes, service delays, and even the interpersonal relationships between patients and healthcare

professionals can all be the source of complaints. (Wolfe, Taylor, and Cameron, 2002) Besides, identifying the factors associated with receiving complaints is crucial, as complaints have significant negative impact a doctor's mental health and well-being. According to Haysom (2016), 8,000 doctors in the UK was examined and found that those undergoes professional complaints have higher risk of anxiety, depression, and even suicidal thoughts. Moreover, doctors with recent or ongoing complaints were 3.78 times more likely to experience suicidal ideation compared to those without complaints. (Haysom, 2016) Depending on the type of complaint, the degree of psychological distress varied; doctors who were subject to formal disciplinary procedures reported the highest levels of anxiety, depression, and thoughts of self-harm.

Past research indicates that various demographic and professional factors, such as gender, years of experience, work hours, and patient volume, are linked to differing rates of complaints among healthcare professionals. For instance, a study conducted by Daniel et al. in 1999 found that over half of the incidents occurred in doctors' consulting rooms, with 87% of the involved doctors being men and more than half general practitioners. The researchers distributed a 32-item questionnaire to 500 complainants through the New South Wales Health Care Complaints Commission (HCCC). Out of 314 responses, 290 were analysed, revealing that 64% of complaints related to clinical care, while 22% involved rudeness or poor communication, and 14% concerned unethical behaviour. (Daniel et al., 1999) Next, a Medical Council of Canada cohort research shown that lowering complaints and malpractice claims requires efficient patient-physician communication. The study tracked the complaint records of 3,424 doctors in Ontario and Quebec who took the clinical skills exam between 1993 and 1996 until 2005. According to the study, 17.1% of doctors still had problems, and 81.9% of them had to do with poor communication. (Tamblyn et al., 2007) Retained complaints increased by 38% when communication scores decreased by 2 standard deviations (RR = 1.38). (Tamblyn et al., 2007) Communication scores were found to be a significant predictor of complaints, even after controlling for clinical decision-making scores. The lowest quartile was responsible for an additional 9.2% of complaints. (Tamblyn et al., 2007) This demonstrates that in order to improve patient care, medical licensure exams require more communication training. Additionally, emergency physicians encounter additional difficulties that may affect their interactions with patients, especially those undergoing residency training. High patient loads, extended work hours, and a lack of experience all lead to higher complaint rates. A retrospective longitudinal cohort study by Hickson et al. (2002) examined 645 general and specialist physicians in a large U.S. medical group from January 1992 to March 1998, accounting for 2,546 physician-years of care. The study found that while experience might help mitigate complaint rates, certain physician demographics, such as gender, are correlated with complaint frequency. (Hickson et al., 2002)

In this study, we analyse a dataset containing information on the number of complaints received by doctors working in an emergency service, alongside demographic information for each doctor. Our objective is to determine which factors are significantly associated with the number of complaints a doctor receives, focusing on variables such as visits, residency status, gender, revenue, and hours worked. This paper is structured into 5

main parts. Section 2 presents the methodology, providing an statistical overview of the our approach to achieve our current model and the interactions selected for analysis. Section 3 reports the results of our modelling, highlighting significant predictors and interpreting the implications of each. Section 4 discusses our findings in the context of existing research on complaints against doctors, particularly in emergency settings. Lastly, Section 5 and 6 provides references used throughout the paper, while an appendix with annotated Rmarkdown code and model outputs is included to support our analysis.

Section 2: Methodology

To gain a better understanding of our data, we began by examining it through both numerical and graphical summaries. We first used the 'summary()' and 'str()' command to provide an overview of the data types, which clarified the dataset's structure. Next, we converted variables into their appropriate data types to ensure accurate analysis. We then created boxplots to explore relationships between variables and identify potential interactions. A histogram of complaint frequencies revealed evidence of zero-inflation, highlighting the need for further investigation into the distribution of our response variable.

Our analysis began with fitting a Poisson regression model to the dataset, followed by a dispersion test that indicated the Poisson model exhibited overdispersion. A rootogram analysis confirmed that the zero counts were underfitted, while most other counts were overfitted. To address these issues, we subsequently fitted a quasi-Poisson model. However, this also showed a dispersion parameter exceeding 2, indicating it was not suitable as well. To further tackle the overdispersion, we applied a negative binomial model, which provided a more flexible variance structure and demonstrated a better fit compared to the previous models. Given the excess zeros in the data, we then employed Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) models, incorporating separate equations for the count and zero-inflation components. By applying this model, we aim to obtain insights into the specific factors influencing complaint rates and to identify whether certain variables contribute to it.

Model evaluation involved residual analysis, AIC scores, and a dispersion test to determine overall goodness of fit. Rootogram were also use to determine goodness of model fit through graphical aspect. Interactions between variables were explored to improve model accuracy. In the count model, the interaction between residency and gender was included to assess how the effect of gender on complaint rates varies by residency status while in the negative binomial distribution part, the interaction between visits and revenue was examined to determine whether the relationship between the number of visits and complaints changes based on the revenue generated. Statistical significance was set at $\alpha = 0.05$ (5%), and all statistical analyses were conducted in the R statistical environment. Statistical significance was set at $\alpha = 0.05$ (5%), and all statistical analyses were conducted in the R statistical environment.

Section 3: Results

"Compdat" includes information on the amount of complaints received by 94 doctors who worked in a hospital emergency department, as well as some demographic data. The data for a single doctor is shown in each row and each row have 6 variables. The summary of data can be found below:

Variable	Description	Summary
visits	Number of patient visits in a year (count)	Min.: 879.000 1st Qu.: 1698.000 Median: 2299.000 Mean: 2271.000 3rd Qu.: 2776.000 Max.: 3763.000
complaints	Number of complaints against the doctor in a year (count)	Min.: 0.000 1st Qu.: 0.000 Median: 0.000: Mean: 1.564 3rd Qu.: 2.000 Max.: 11.000
residency	Doctor's status of residency training (Y = Yes, N = No)	Yes: 45.000 No: 49.000
gender	Gender of the doctor (M = male, F = female)	Male: 57.000 Female: 37.000
revenue	Doctor's hourly income (dollars)	Min.: 203.900 1st Qu.: 243.800 Median: 263.700 Mean: 263.800 3rd Qu.: 288.000 Max.: 342.900
hours	total number of hours the doctor worked in a year (count)	Min.: 589.000 1st Qu.: 1201.000 Median: 1494.000 Mean: 1469.000 3rd Qu.: 1700.000 Max.: 2269.000

Table 1

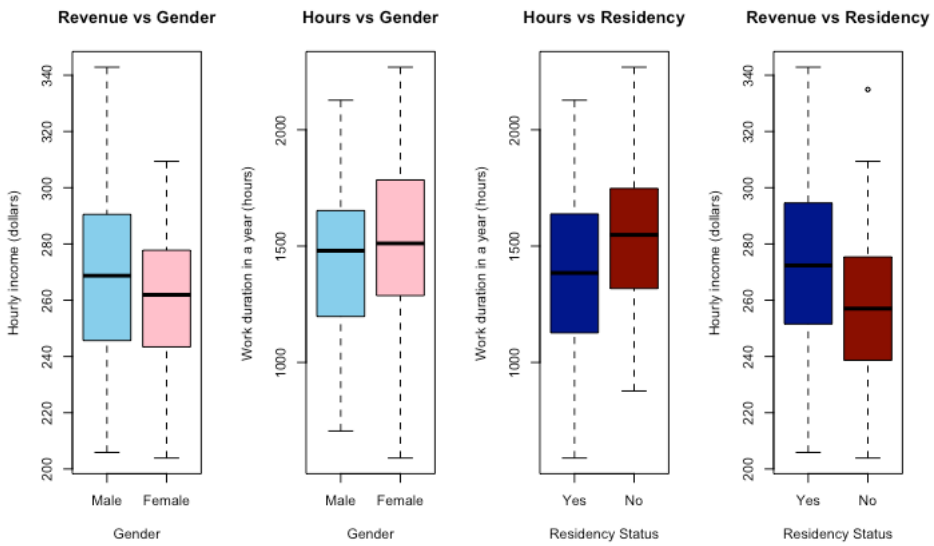

Figure 1: Boxplot on Doctor's Gender and Residency Training Status

In the "Revenue vs Gender" boxplot, it suggest that male doctors tend to have slightly higher hourly income than female doctors, with male doctors having a much wider income range. Next, "Hours vs Gender" boxplot shows that female doctors generally work more hours per year than male doctors. However, there isn't much different between the median of both male and female doctors. The "Hours vs Residency" boxplot shows that residents generally work less hours per year than non-residents, with residents showing a more variation in hours worked in a year. In the "Revenue vs Residency" boxplot, non-resident doctors have significant lower hourly income than resident doctors, with the median of residents at around 275 dollars per hour and median of non-residents at around 260 dollars per hour. Non-residents also displayed less variation in income.

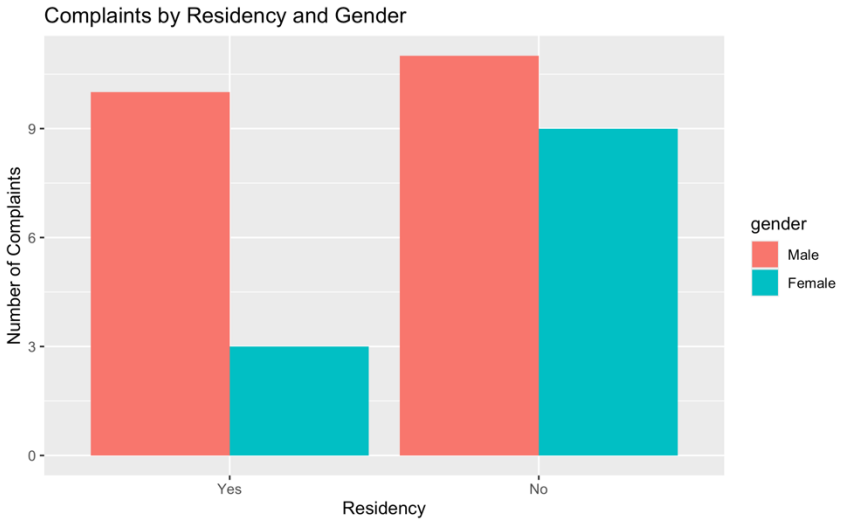

Figure 2: Complaints by Residency and Gender

As we observe form the bar graph, male physicians typically have high complaint counts regardless of residency. On the other hand, female physicians appear to have less complaints when they are during residency training. These trends may be influenced by

variables such as patient load, experience, and perhaps even the working environment in residency programs. This pattern raises the possibility that complaints to male and female doctors may differ depending on their residency status.

Model Selection:

Our analysis began by fitting a Poisson regression model, but a dispersion test revealed overdispersion. A rootogram confirmed that zero counts were underfitted and most other counts were overfitted. We then fitted a quasi-Poisson model, which also indicated a dispersion parameter exceeding 2. To address overdispersion, we applied a negative binomial model, offering a more flexible variance structure and a better fit. Due to the excess zeros in the data, we further utilized Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) models, which included separate equations for the count and zero-inflation components. This approach lead us to our final model that enable us to study on the factors influencing complaint rates and identify contributing variables. The 2 equations of Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) models can be found below:

Count Model:

$$\log(\lambda) = 1.2434 + 0.0017\cdot\text{visits} + 0.3298\cdot\text{gender(Female)} - 0.0100\cdot\text{revenue} - 0.0016\cdot\text{hours} + 0.6332\cdot(\text{gender(Male)} : \text{residency(No)}) - 0.5002\cdot(\text{gender(Female)} : \text{residency(No)})$$

Zero-inflation model:

$$\log\left(\frac{p}{1-p}\right) = -0.0545 + 1.6000\cdot\text{visits} + 1.6330\cdot\text{revenue} - 2.4640\cdot\text{hours} - 0.0058\cdot(\text{visits*revenue}) + 0.0091\cdot(\text{revenue*hours})$$

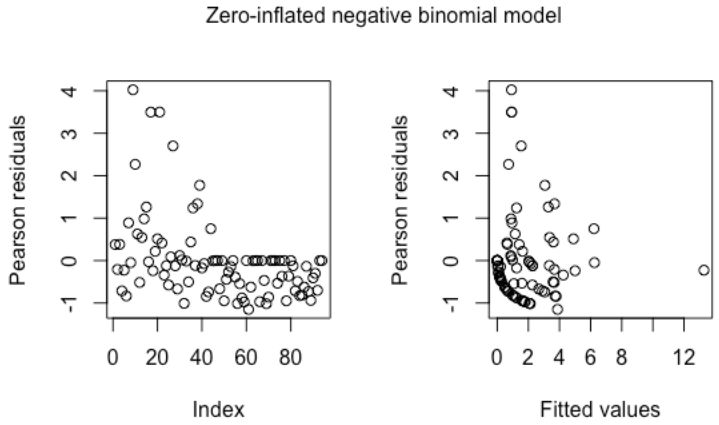

Figure 3: Pearson Residual Diagnostics for Zero-Inflated Negative Binomial Model

1. Index vs Pearson Residuanls
   This plot examines the Pearson residuals across observations to assess the model's fit consistency. The random scatter around zero suggests that the model does not display a systematic bias across the dataset.

2. Fitted values vs Pearson Residuanls
   This plot illustrates the relationship between Pearson residuals and fitted values, helping to detect any patterns in the residuals relative to the predicted values. The concentration of residuals around zero suggests that the model's predictions align closely with observed values, particularly at mid-range fitted values.

Section 4: Discussion

Based on the model equations derived in the previous section, all variables are significant predictors of the number of complaints received by doctors in the emergency service except female and female that is not in residency training. The effects of the variables on the number of complaints are as follows:

- **Number of patient visits (visits)**: For every additional patient visit, the expected number of complaints increases by a factor of 1.0017, indicating a strong positive correlation between the volume of visits and the likelihood of receiving complaints.
- **Revenue (revenue)**: A unit increase in hourly income results in a decrease in the number of complaints by a factor of 0.9901, suggesting that higher earnings are associated with fewer complaints, possibly due to perceived quality of care.
- **Total hours worked (hours)**: For each additional hour worked per year, the expected number of complaints decreases by a factor of 0.9984, indicating that more hours on the job may correlate with better management of patient interactions as physicians have ample of experience.
- **Gender interaction with residency status (genderMale)**: This interaction approaches significance with a coefficient of 0.6332, suggesting that male doctors who are not in residency training may face a higher risk of complaints compared to their female counterparts or those in residency.
- **Zero-inflation model for visits**: The coefficient of 1.600 indicates that as the number of visits increases, the log odds of having zero complaints also increases significantly.
- **Zero-inflation model for hours**: The coefficient of -2.464 suggests that doctors who work fewer hours are more likely to have zero complaints, highlighting a potential relationship between workload and patient satisfaction.
- **Interaction between visits and revenue (visits)**: A coefficient of -0.00576 suggests that the effect of patient visits on complaints is moderated by the doctor's revenue.

- **Interaction between revenue and hours (revenue)**: The positive coefficient of 0.00905 indicates that as both revenue and hours increase, the likelihood of having zero complaints also increases.

In light of the findings from our analysis, it is evident that the demographic and professional factors influencing complaints against emergency department physicians align with previous research. For instance, the significant role of communication highlighted in the studies by Tamblyn et al. (2007) and Hickson et al. (2002) reinforces the need for targeted training in this area to mitigate complaints. Our results also indicate that high patient volumes and extended working hours are critical factors contributing to complaint rates, echoing concerns raised in the literature about the pressures faced by emergency physicians. Moreover, the interplay between gender and residency status in relation to complaint frequency provides further insight into the complexities identified in previous studies. Overall, our findings contribute to a deeper understanding of the dynamics at play in emergency care settings and underscore the necessity for continuous quality improvement in physician-patient interactions.

Section 5: References

Cooke, M., Mann, C., Edwards, N., & North, J. (2020). Emergency medicine. In M. McKee, S. Merkur, N. Edwards, & E. Nolte (Eds.), *The Changing Role of the Hospital in European Health Systems* (pp. 181–200). chapter, Cambridge: Cambridge University Press. https://www.cambridge.org/core/books/changing-role-of-the-hospital-in-european-health-systems/emergency-medicine/9CE4BF36BF84A1EBC9472CD0DD7ADD49

Daniel, A. E., Burn, R. J., & Horarlk, S. (1999). Patients' complaints about medical practice. *Medical Journal of Australia, 170*(12), 598-602. **https://doi.org/10.5694/j.1326-5377.1999.tb127910.x**

Haysom, G. (2016). The impact of complaints on doctors. *Australian family physician*, *45*(4), 242–244. https://pubmed.ncbi.nlm.nih.gov/27052144/

Hickson, G. B., Federspiel, C. F., Pichert, J. W., Miller, C. S., Gauld-Jaeger, J., & Bost, P. (2002). Patient complaints and malpractice risk. *JAMA*, *287*(22), 2951–2957. https://doi.org/10.1001/jama.287.22.2951

Tamblyn, R., Abrahamowicz, M., Dauphinee, D., Wenghofer, E., Jacques, A., Klass, D., Smee, S., Blackmore, D., Winslade, N., Girard, N., Du Berger, R., Bartman, I., Buckeridge, D. L., & Hanley, J. A. (2007). Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA*, *298*(9), 993–1001. https://doi.org/10.1001/jama.298.9.993

Taylor, D. M., Wolfe, R., & Cameron, P. A. (2002). Complaints from emergency department patients largely result from treatment and communication problems. *Emergency Medicine Au stralasia, 14*(1), 43-49. https://doi.org/10.1046/j.1442-2026.2002.00284.x

Section 6: Appendix

1. Read in data and display the first 6 rows
2. Display the structure of the dataset, including data types and variable names.
3. Check for duplicate for data cleaning
4. Factoring residency and gender into categorical variable
5. Generate summary statistics for each variable in the dataset.

```{r}
CD <-read.table("compdat.txt", header = TRUE)
head(CD)
str(CD)

sum(duplicated(CD))

CD$residency <- factor(CD$residency, levels = c("Y", "N"), labels = c("Yes", "No"))
CD$gender <- factor(CD$gender, levels = c("M", "F"), labels = c("Male", "Female"))
summary(CD)
```

Output:

Description: df [6 × 6]

	visits <int>	complaints <int>	residency <chr>	gender <chr>	revenue <dbl>	hours <dbl>
1	2014	2	Y	F	263.03	1287.25
2	3091	3	N	M	334.94	1588.00
3	879	1	Y	M	206.42	705.25
4	1780	1	N	M	226.32	1005.50
5	3646	11	N	M	288.91	1667.25
6	2690	1	N	M	275.94	1517.75

6 rows

```
'data.frame':    94 obs. of  6 variables:
 $ visits    : int  2014 3091 879 1780 3646 2690 1864 2782 3071 1502 ...
 $ complaints: int  2 3 1 1 11 1 2 6 9 3 ...
 $ residency : chr  "Y" "N" "Y" "N" ...
 $ gender    : chr  "F" "M" "M" "M" ...
 $ revenue   : num  263 335 206 226 289 ...
 $ hours     : num  1287 1588 705 1006 1667 ...
[1] 0
     visits       complaints     residency    gender      revenue          hours
 Min.   : 879   Min.   : 0.000   Yes:45    Male  :57   Min.   :203.9   Min.   : 589
 1st Qu.:1698   1st Qu.: 0.000   No :49    Female:37   1st Qu.:243.8   1st Qu.:1201
 Median :2299   Median : 0.000                         Median :263.7   Median :1494
 Mean   :2271   Mean   : 1.564                         Mean   :263.8   Mean   :1469
 3rd Qu.:2776   3rd Qu.: 2.000                         3rd Qu.:288.0   3rd Qu.:1700
 Max.   :3763   Max.   :11.000                         Max.   :342.9   Max.   :2269
```

We first use Poisson distribution as the response variable is set to be the number of complaints doctor received in the previous year. Which specially specify a fixed interval. We also carried out dispersion test to further test if this data set is zero-inflated as the test's z-score and highly significant p-value indicate over dispersion—where the observed variance is much greater than what a standard Poisson model would predict.

```{r}
#install.packages("AER")
library(AER)

pm.model <- glm(complaints ~ visits + residency + gender + revenue + hours, family =
poisson , data = CD )
summary(model)
dispersiontest(model)
```

Output:

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

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.2313546  0.8103654  -1.520    0.129
visits        0.0008537  0.0001597   5.346 8.98e-08 ***
residencyNo   0.8648113  0.1879513   4.601 4.20e-06 ***
genderFemale -1.1202803  0.2142123  -5.230 1.70e-07 ***
revenue      -0.0010474  0.0028260  -0.371    0.711
hours        -0.0002270  0.0003289  -0.690    0.490
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 312.62  on 93  degrees of freedom
Residual deviance: 221.66  on 88  degrees of freedom
AIC: 356.44

Number of Fisher Scoring iterations: 6


        Overdispersion test
data:  model
z = 3.5054, p-value = 0.000228
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
  2.651855
```

- We then fit it into a quasi-poisson model as previous model is over disperse

```{r}
qp.model <- glm(complaints ~ visits + residency + gender + revenue + hours, data = CD,
family = quasipoisson)
summary(qp.model)
```

Output:

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

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.2313546  1.4084010  -0.874  0.38434
visits        0.0008537  0.0002775   3.076  0.00279 **
residencyNo   0.8648113  0.3266561   2.647  0.00961 **
genderFemale -1.1202803  0.3722973  -3.009  0.00342 **
revenue      -0.0010474  0.0049115  -0.213  0.83163
hours        -0.0002270  0.0005717  -0.397  0.69228
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be 3.020584)

    Null deviance: 312.62  on 93  degrees of freedom
Residual deviance: 221.66  on 88  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6
```

- **Now we fit it into a Negative Binomial model**

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

```
Call:
glm.nb(formula = complaints ~ visits + residency + gender + revenue +
    hours, data = CD, init.theta = 0.614428857, link = log)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.4788997  1.6486836   0.897   0.3697
visits        0.0013164  0.0002997   4.392 1.12e-05 ***
residencyNo   0.4204865  0.3867910   1.087   0.2770
genderFemale -0.7071510  0.3894724  -1.816   0.0694 .
```

```
revenue      -0.0087937  0.0057936  -1.518   0.1291
hours        -0.0013480  0.0005991  -2.250   0.0245 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(0.6144) family taken to be 1)

    Null deviance: 108.406  on 93  degrees of freedom
Residual deviance:  83.535  on 88  degrees of freedom
AIC: 301.04

Number of Fisher Scoring iterations: 1


              Theta:  0.614
          Std. Err.:  0.167

 2 x log-likelihood:  -287.040
```

- Fit into zero inflated poisson model

```
```{r}
library(pscl)
zip.model <- zeroinfl(complaints ~ visits + residency + gender + revenue + hours | residency
+ gender + revenue + hours, dist = "poisson", data = CD)
summary(zip.model)
```
```

```
Call:
zeroinfl(formula = complaints ~ visits + residency + gender + revenue + hours | residency +
gender + revenue + hours,
    data = CD, dist = "poisson")

Pearson residuals:
    Min      1Q  Median      3Q     Max
-1.5151 -0.6762 -0.4027  0.3066  3.5687

Count model coefficients (poisson with log link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.005e+00  1.776e+00   0.566  0.57169
visits        1.537e-03  5.939e-05  25.887  < 2e-16 ***
residencyNo   3.187e-01  2.614e-01   1.219  0.22275
genderFemale -7.151e-02  2.745e-01  -0.261  0.79448
revenue      -8.621e-03  3.330e-03  -2.589  0.00963 **
hours        -1.146e-03  7.809e-04  -1.467  0.14241

Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3227353  3.2012073  -0.726   0.4681
residencyNo  -0.4596960  0.7608640  -0.604   0.5457
genderFemale  1.2595161  0.6115812   2.059   0.0395 *
revenue       0.0025529  0.0096367   0.265   0.7911
```

```
hours         0.0007135  0.0017551   0.407   0.6844
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 19
Log-likelihood: -135.4 on 11 Df
```

- Lastly, we fit it into a zero inflated negative binomial model

```
```{r}
library(pscl)
zipNB_model <- zeroinfl(complaints ~ visits + residency + gender + revenue + hours | visits
+ residency + gender + revenue + hours, data = CD, dist = "negbin", EM = TRUE)
summary(zipNB_model)
```
```

```
Warning: NaNs produced
Call:
zeroinfl(formula = complaints ~ visits + residency + gender + revenue + hours | visits +
residency + gender + revenue +
    hours, data = CD, dist = "negbin", EM = TRUE)

Pearson residuals:
     Min       1Q   Median       3Q      Max
-1.03032 -0.69009 -0.08503  0.27626  3.14892

Count model coefficients (negbin with log link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.5342434  1.4969256  -3.029 0.002453 **
visits       -0.0005980        NaN     NaN      NaN
residencyNo   0.4486384  0.3607169   1.244 0.213595
genderFemale -0.4192660  0.3523527  -1.190 0.234084
revenue       0.0114601  0.0049428   2.319 0.020421 *
hours         0.0025506  0.0007219   3.533 0.000410 ***
Log(theta)    0.5230834  0.1580763   3.309 0.000936 ***

Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -33.119221        NaN     NaN      NaN
visits        -0.007596   0.001773  -4.283 1.84e-05 ***
residencyNo   -0.754756   1.085047  -0.696    0.487
genderFemale   1.644742   1.537453   1.070    0.285
revenue        0.087029   0.020243   4.299 1.71e-05 ***
hours          0.017281        NaN     NaN      NaN
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 1.6872
Number of iterations in BFGS optimization: 1
Log-likelihood: -128.6 on 13 Df
```

- After studying the interactions, we come to our final model

```{r}
zipNB_model2 <- zeroinfl(complaints ~ visits + gender + revenue + hours  +
residency:gender| visits*revenue+hours*revenue  , data = CD, dist = "negbin", EM =
TRUE)
summary(zipNB_model2)
```

```
Call:
zeroinfl(formula = complaints ~ visits + gender + revenue + hours + residency:gender |
visits * revenue + hours *
    revenue, data = CD, dist = "negbin", EM = TRUE)

Pearson residuals:
      Min        1Q    Median        3Q       Max
-1.150e+00 -6.126e-01 -1.220e-01 -4.333e-09  4.024e+00

Count model coefficients (negbin with log link):
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)            1.2433897  1.5186682   0.819   0.4129
visits                 0.0017189  0.0001091  15.753   <2e-16 ***
genderFemale           0.3298146  0.4825020   0.684   0.4943
revenue               -0.0099623  0.0045867  -2.172   0.0299 *
hours                 -0.0016069  0.0007143  -2.250   0.0245 *
genderMale:residencyNo  0.6331766  0.3244766   1.951   0.0510 .
genderFemale:residencyNo -0.5002029  0.5486383  -0.912   0.3619
Log(theta)             0.6969628  0.0531361  13.117   <2e-16 ***

Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.448e+02  2.084e+01  -26.15   <2e-16 ***
visits         1.600e+00  1.220e-03 1311.70   <2e-16 ***
revenue        1.633e+00  7.401e-02   22.07   <2e-16 ***
hours         -2.464e+00  7.735e-03 -318.60   <2e-16 ***
visits:revenue -5.761e-03  2.037e-05 -282.74   <2e-16 ***
revenue:hours   9.046e-03  4.292e-05  210.76   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 2.0076
Number of iterations in BFGS optimization: 1
Log-likelihood: -120.2 on 14 Df
```