TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

Study the Environmental Factors that affect Children for Asthma

Azad Adil Shareef[1,2], Assistant Lecturer, [1]IS Directorate, Presidency of Duhok University, Duhok University
[2] Department of Statistics, College of Administration and Economics, Duhok University.

Article Info
Volume 83
Page Number: 2904 – 2917
Publication Issue:
May - June 2020

Article History
Article Received*: 11August 2019*
Revised*: 18November 2019*
Accepted*: 23January 2020*
Publication*: 10May2020*

Abstract:

Asthma has played an important role in statistics, particularly in terms of environmental factors in most developed countries in the world over the past fifty years. This research examines some significant aspects of correlation and fitting the best models for selecting and evaluates the different kinds of hypothesis testing. It is concluded that the most efficient method to use the graphical chain modelling to depict the association structure between the background variables and how they are associated with the chid admission in Mexico city. There are two under sections, the first is Poisson distribution and Negative Binomial distribution are used in order to choose models that are both simple and fit the data well. Various environmental factors have been considered for the cause of asthma among children in Mexico and it can be seen that, amount of Rain and Tree pollen arise as major factors. According to the analysis, month of September recorded the highest number of child admission from asthma.

Keywords*: Environmental Factors, Asthma, Poisson distribution*

I. INTRODUCTION

Asthma is one of the respiratory diseases that causes the patient to be unable to breathe properly. The main cause of the disease has not been discovered, but the common belief is that the cause is a combination of genetic factors, environmental factors. Asthma is becoming increasingly common in the developed world and is now the most common chronic condition in the west. More than 5.2 million people in the UK are being treated for asthma and about 1.1 million of these are children. Asthma a ects approximately one in 12 adults and one in eight children in the UK. This means there is a person with asthma in one in five households in the UK. It can a ect almost anyone, at any age, anywhere although it tends to be worse in children and young adults (Agertoft and Pedersen ; 1994). Asthma occurs as a result of a combination of complex environmental and genetic reactions that are not fully understood, affecting asthma severity and response. Treatment The recent increase in asthma rates is thought to be due to changes in sequencing factors (genetic factors differ from those associated with the hypoxic) And in the living environment. Pre-eclampsia is usually associated with genetic factors, but its onset after 12 years is due to environmental factors. During the latter part of the last century there has been a steady increase in countries that depend on Western lifestyle as well as in developing countries, as current estimates indicate that 300 million people worldwide suffer from the Asthma.

Objective of the study

The main objective of the study is to understand the relationship between asthma among children and environmental factors. It has been suggested that asthma can be triggered and exacerbated by exposure to many environmental factors. The American Academy of Paediatrics has recently published a book about childhood environmental health problems, which states: "Avoiding environmental

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

allergens and irritants is one of the primary goals of good asthma management" .
Environmental factors that increase the risk of developing asthma include:
• Exposure to allergens.
• Cold air, wind, rain and sudden changes in the weather can sometimes bring on an asthma attack.
• Time of year when the pollen count is high.
• Air pollution.

Importance of the Study
Reach the stage of disease (Asthma) stabilization.

Reduce the number of acute asthma attacks and use as few bronchodilators as possible.

The Children continues to practice his normal life without any obstacles.

Problem of the study
Asthma is caused by a combination of factors including genetic predisposition and its interaction with environmental factors. The difficulty of breathing as mentioned, accompanied by cough and asthma symptoms, especially when the person is exposed to any additional infections as a result of some diseases such as colds and flu, and these symptoms very much at night.

What are the causes?
Over the years, intensive researches have been carried out in order to understand the in uential factors cause for asthma and mainly it can be narrowed down to two main categories: (Holgate; 1999).
• Genetics and Asthma.
• Environmental factors and Asthma.

1-Material and methods
The study based on the records of child admissions to the Mexico City hospital (Busse; 1996).. Daily number of patients diagnosed as su ering from "asthma" were extracted by a trained health-care professional between the month July and October. In addition to that, several other environmental data has been collected between the period July and October. Variables in this study can be divided into three aspects:

• Bio particles: Grass Pollen (Grass), Tree Pollen (Tree), Weed pollen (Weed), Basidiomycete spores (Basidiom), Ascomycete spores (ASC), Deuteromycete spores (Deuterom);
• Air pollutants: Max. Hourly Ozone (O3), Number of hours Ozone (O3hours), Max. Hourly nitrogen dioxide (NO2), Max.hourly sulphur dioxide (SO2);
• Weather: Daily rainfall (Rain), Maximum daily temp. (MaxTemp), Minimum daily temp. (MinTemp), Daily average relative humidity (Rain).

1.1 Missing Values
The most common issue which arises in almost surveys is missing data. There are various reasons due to happening this, losing or corrupting samples is one possible cause, in addition to it, there might be some questions which partic ipants may not want to answer the questions. In fact, it is always essential to handle the nature of the distribution of data. Having too much missing data can a ect the results of any statistical analysis that are performed. Therefore, in order to prepare the data for further analysis, one solution that is regularly applied is imputation, that is, finding appropriate replacements for the missing data. However, many methods can be used to evaluate them.

1.1.1 How to deal with?
With reference to Little & Rubin (1987) missing values are classified into three different classes of missingness regardless of their dependence structure. Missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), which are explained as following:
• The term MCAR refers to data where there is no relationship between variables, in addition, it means both missing at random and observed at random (The random collected data do es not depend on any other in the data set). In order for case deletion to be valid this condition is required. (Rubin, 1976).

TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

• With regard to MAR, a more general assumption, missing at random, is the probability a variable is missing relates only on existing information.
• The MNAR refers to Not Miss at Random, (or informatively missing, as it is often known) shows when the missingness mechanism depends on the real value of the missing data. This is the most hard condition to model for.

We have defined missing data in detail and mentioned their types as well. In our data, several variables has missing values and ignoring them may lead to substantial biases in the analyses. The statistical package SPSS can be used to either showing the percentage and computing the replacements of missing data. First, let's consider the Frequency analysis of missing values. We found that the highest percentage of missing value is for RH which has about 10.57%. Similarly, NO2 and SO2 come after it in the second and third place with having 7.32% and 6.50% respectively, which means these missing values should be taken into account. However, for O3, we can ignore it and there may not be exist a big change in the consequence due to having only two missing values. In total there are 32 missing values, it does in uence the result if they are not replaced. There are many methods to evaluate missing values, here we use EM algorithm.

1.1.2 The EM algorithm

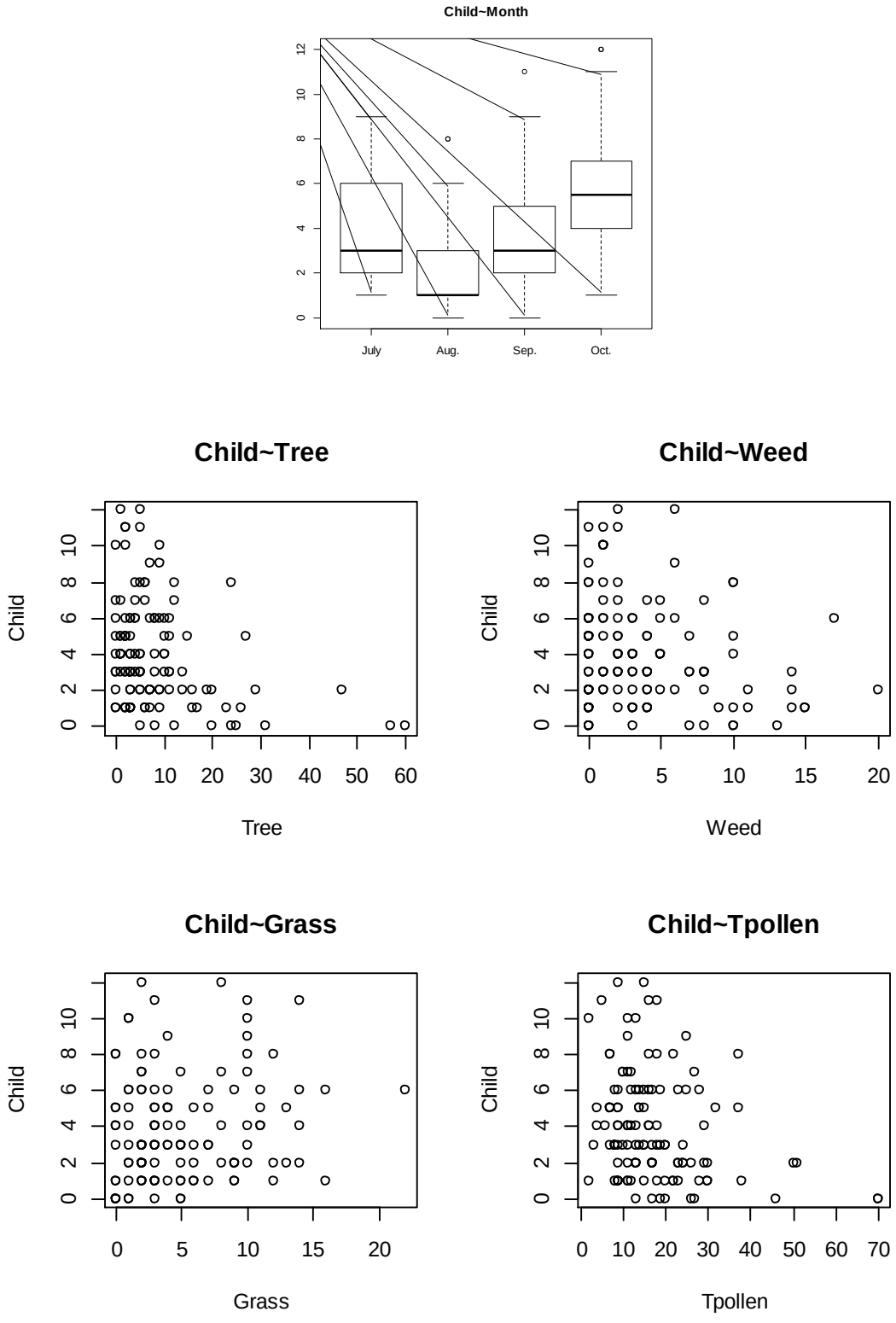
The EM algorithm is based on a two step process to compute for estimating model parameters. It integrates missing data in the estimation process, thus bypassing the need to impute. The basic algorithm consists of two steps: expectation (E step) and maximization (M step). Initially the data is parti tioned into missing and non-missing, and then begin with starting values for the parameters. Firstly, using the parameters, calculate the predicted scores for the missing data (the expectation). Secondly, using the found scores for the missing data, maximize the likelihood function to attain new parameter estimates. Reiterate the process until convergence is occurring. Nevertheless, we are not doing it by theoretically, it can be computed by SPSS.

2- Exploratory Data Analysis

Exploratory data analysis is used initially to learn about data set. From histograms of all variables we found that most of them are positively skewed (e.g. Child, Deuterom, etc.), and histograms of MaxTemp, MinTemp and RHare negatively skewed. Only O3 seems to have Normal distribution. From boxplots we can read o the minimum lower quartile, median, upper quartile and maximum. We can also find outliers according to the boxplots of all variables, but we may need further modelling to explain for these outliers.By using QQ plot we can confirm whether a variable is normally distributed and it seems that only O3 is normally distributed variable. Let's look at some summary statistics for Child by other variables. For Child and Month, Figure 2.1 shows that September has the highest mean and median Child, while July has the lowest Child. With this graph we consider that there exist some kinds of relationships between Child and Month, so we should try to include Month as a covariate when predicting Child. Then we consider other potential covariate. From the scatterplots of other variables we found obvious negative relationships between Child and Tree, Tpollen, Rain. This means that if Tree or Rain increases, Child decreases. It could be that more trees and rainfall bring fresher air, and then less children will have asthma. We can calculate the correlations (cor(Child,Tree)) and (cor(Child,Rain)) as -0.298 and -0.217, respectively. Without further modelling we will not know whether they are significant relationship.

Chi ld ~Mon t h , Child~Tree 'Aug ' 'Jul' ' Oct' 'Sep ' and the figures below show the relationships between some variables in this studies such as Child by Month, Child by Tree and Child by Tpollen

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917



Fiqure 2.1 Relationship between child with grass,T pollen and tree

3- Modeling

The section above has explored the data and the present section fit the regression model and use the graphical chain modelling to depict the association structure between the background variables and how they are associated with the chid admission in Mexico city. There are two under sections, the first

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

is Poisson distribution and the second is Negative Binomial distribution are used in order to choose models that are both simple and fit the data well.

3.1 Model choice

Choosing models are the most important part in obtaining the appropriate result to make implication for the future (Gaur; 2006). There are some assumptions that we may take into account to select the model which fits the data well, such as whether the response variable plays as counted observations, or by looking its distribution and its association with predictor variables and so on. Now, here the response variable is a binary variable (yes or no) which means that "Yes" stands for the child who has asthma and "No" has not. So, it would seem that the child admission variable is more likely to be approximately.(Mann; 2007*)* Poisson distribution due to it implies to count the children who are a ected by asthma. Due to the fact that the response variable is both positive and integer, then it is obvious a generalised linear model with having Poisson family is appropriate here. However, we cannot say that the response is exactly distributed as a Poisson distribution. As long as it has integer values, we can assume Child admission follows a Poisson:

Child $\sim$ Poisson(μ) $\log(\mu_i) = X_i^T \beta$

Where Child is the child admission of asthma, X is a vector of possible exi planatory variables and β is the corresponding parameter vector.

Table 3.1: Correlation between Response Variable (Child) and the Exploratory Variables

cor(Deuterom,Child)	0.1346457
cor(ASC,Child)	-0.06412334
cor(Basidiom,Child)	-0.1878432
cor(Tree,Child)	-0.2977368
cor(Weed,Child)	-0.2004219
cor(Grass,Child)	0.1541982
cor(Tpollen,Child)	-0.2771709
cor(MaxTemp,Child)	0.03046704
cor(MinTemp,Child)	0.04442969
cor(RH,Child)	-0.03452375
cor(Rain,Child)	-0.2166712
cor(O3,Child)	-0.05699631
cor(NO2,Child)	-0.03363497
cor(SO2,Child)	-0.1392338
cor(O3hours,Child)	-0.0278378

3.2 Model Selection (Poisson Distribution)

Now, let's begin fitting models to the data set. It is important to take one of the strategy of testing models to achieve the best one although there cannot be the best one. Here, our aim is to decide which of a range of competing models fits the data best (Howell; 2010). For doing so we will be using the Likelihood Ratio Test. We build the model sequentially (forward selection) starting from the null model (intercept only) and moving to more complicated models considering the significance level 5%. By using R, we obtain Model 0:(Child) 1.36990 with deviance 275.04 on 122 degrees of

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

freedom. As we can see that the residual deviance is much larger than the degrees of freedom. Thus, we need to add a term into our model to see whether wecan get a better one or not. Now, we add Month as an explanatory variable, the reason to addit at the first time is we have already checked the model with every single predictor variable and as shown on the table above, correlation between the response and predicted variables gives good indication for the selection of variables. We have to be very sensible with choosing model, and then we compared them with the null model's residual deviance, and we took a difference between them. So, the following table contains all the mo dels with their di erence in residual deviance. Therefore, we found that Month should be added as the first one. It would seem that there are several independent variables which are significant and suggest to start with. We should, however, choose a model which has the largest di erence in deviance. This means Month is added firstly with 58.324 as a factor to the null model.

Table 3.2: Different plausible log-linear model selected (family=poisson)

Models	Differences in df	Differenc in deviance	P-value
Child ~ 1	-	-	-
Child ~ Month	3	3 58.324	1.34e-12
Child ~ Tree	1	30.235	3.828e-08
Child ~ Tpollen	1	23.895	1.017e-06
Child ~ Rain	1	14.361	0.0001509
Child ~ Weed	1	11.429	0.000723
Child ~ Basidiom	1	10.36	0.001288
Child ~ Grass	1	6.0486	0.01392
Child ~ SO2	1	5.2249	0.02227
Child ~ Deuterom	1	4.5704	0.03253
Child ~ ASC	1	1.1287	0.2881
Child ~ O3	1	0.85978	0.3538
Child ~ MinTemp	1	0.52668	0.468
Child ~ RH	1	0.31324	0.5757
Child ~ NO2	1	0.30482	0.5809
Child ~ MaxTemp	1	0.24676	0.6194
Child ~ O3hours	1	0.20429	0.6513

Models Difference in df, Difference in deviance and P-value

Child ~ 1

After adding Month we obtain Model 1:

(Child) ~ 1.37820 - 0.62253 * (Jul) - 0.05859 * (Oct) + 0.40799 * (Sep)

with deviance 216.71 on 119 degrees of freedom. Compared with Model 0: "(Child)~ 1.36990", we find that the change in deviance is very big and p value = 1.34e - 12 < 0.05. This means that, the null hypothesis is rejected and we conclude that the more complex model that includes Month fits the data better than the null (intercept only) model. Having decided on the usefulness of Month, we will then add an additional explanatory variable. We decide

TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

toadd Tree, because it comes as the second most significant shown in the Table 3.2. Further, according to the exploratory analysis and Table 4.1 there is a strong negative relationship between number of Child admission and Tree pollen. Hence repeat the process of fitting a binary log-linear model but adding Tree now. Then we can obtain Model 2:

(Child) ~ 1.481150 - 0.511179 * (Jul) - 0.066457* (Oct) + 0.348016 * (Sep) -0.014921 * (Tree) with deviance 212.16 on 118 degrees of freedom. Compared with Model 1, although the deviance is quite small, it is still significant at 5% level. So we will keep Tree in the model. The further model should be fitted in order to improve the mo del, we then add Tpollen as an additional explanatory variable and we obtain Model 3: (Child) ~ 1.489559 - 0.511442 * (Jul) - 0.071038 *(Oct) + 0.345907 * (Sep) -0.014265 * (Tree) - 0.000701 * (Tpollen) with deviance 212.16 on 117 degrees of freedom. Compared with Model 2, the change of deviance is quite small and the p value is greater than 0.05. Consequently, the Tpollen cannot a ect the variability of child admission when the factors of the Month and Tree are available. Therefore, it cannot be necessary to consider on the covariate Tpollen. In the next model Rain will be added as an additional independent variable due to coming as the third most significant in the Table 3.2. Then we obtain Model 4: (Child) ~ 1.543903 - 0.382769 * (Jul) - 0.024854 * (Oct) + 0.347701 * (Sep)- 0.018256 * (Tree) -0.019882 * (Rain) with deviance 205.85 on 117 degrees of freedom. Compared with Model 2, it is significant and we conclude that the addition of Rain improves the fit of the model significantly.

It will not be necessary to add any more additional explanatory variables because with regard to the Table 3.1 and Table 3.2 after Rain except Weed although they are all significant but their p-value are not smaller than 0.05 compared to the others, So we do not consume our time. Hence, we will start adding their interaction between these variables that have improved our models. Firstly we consider the interaction between Month and Tree, then we obtain Model 5:

(Child) ~ 1.268358 + 0.127037 * (Jul) + 0.342099 * (Oct) + 0.493096 * (Sep) + 0.021149 * (Tree) - 0.023145 * (Rain) - 0.057569 * (Jul * T ree) - 0.052384 * (O ct * T ree) + 0.007428 * (Sep * Tree) with deviance 194.38 on 114 degrees of freedom. Compared with Model 4 we find it is significant at 5% level. That is, the interaction between Month and Tree and is significant in explaining the variability of accruing asthma of Child. Next we consider adding the interaction between Month and Rain. We obtain Model 6: (Child) ~ 1.2591792 + 0.0805258 * (Jul) + 0.3184436 *(Oct) + 0.5700490 * (Sep) + 0.0210671 * (Tree) - 0.0179584 * (Rain) - 0.0562102 * (Jul * Tree) - 0.0504026 * (Oct * T ree) + 0.0122800 * (Sep * Tree) + 0.0008231 * (Jul * Rain)

+ 0.0005131 * (Oct * Rain) - 0.0409650 * (Sep * Rain) with deviance 191.49 on 111 degrees of freedom. Compared with Model 5, p value = 0.4079 > 0.05, and we conclude that the interaction among Month and Rain is not significant. So, we will remove it for the next additional.

Then we consider the interaction between Tree and Rain. We obtain Model 7: (C hild) ~ 1.226901 + 0.137201 * (Jul) + 0.294390 * (Oct) + 0.500822 * (Sep) + 0.029658 * (Tree) - 0.007964 *(Rain) - 0.056741 * (Jul * Tree) - 0.045911 * (Oct * Tree) + 0.008590 * (Sep * Tree) - 0.004013 * (Tree * Rain) with deviance 189.95 on 113 degrees of freedom. Compared with Model 5, p value = 0.03536 < 0.05, so we accept the more complicated model with having two interaction terms. Finally we consider all interactions, and we obtain Model 8: (Child) ~ 1.124767 + 0.177594 * (J ul) + 0.443686 * (Oct) + 0.742303 * (Sep) + 0.038021 * (Tree) + 0.036443 * (Rain) - 0.061324 * (Jul *

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

Tree) - 0.065586 * (Oct * Tree) - 0.022015 * (Sep * Tree) - 0.033707 * (Jul * Rain) - 0.052441* (Oct * Rain) - 0.114978 * (Sep * Rain) - 0.007916 * (Tree * Rain) + 0.002509 * (Jul * T ree * Rain) + 0.007455 * (Oct * Tree * Rain) + 0.015217 * (Sep * T ree * Rain) with deviance 183.06 on 107 degrees of freedom. Compared with Model 7, p value = 0.3307 > 0.05. Therefore, we ended up with the model with having two interaction terms between Month and Tree as well as Tree and Rain and we take this model as our chosen model. It can be stated that these variables lead to o ccur asthma on Child. The final mo del is Model 7: Child ~ Month + T ree + Rain + Month * T ree + T ree * Rain

3.3 Model Checking (Poisson Distribution)

As soon as a model is selected, there needs to check its residual plots by using the standardised residuals. The graph displays that there cannot be seen an exact relationship between residuals and the fitted values. In case of Poisson mo del the raw residuals have a variance proportional to the mean. Nevertheless, this plot illustrates that it does not re ect such properties. As shown in Figure 3.2, the standard residual is not likely to have a normal distribution. It does not only have a bell-shaped distribution, but also its variance is relatively too large. The sample variance of that is 1.554. Thus, for these two reasons and including the overdispersion, we can conclude that this mo del still does not fit the data well. There are various explanation for this, it is possible to say that there may be other explanatory variable(s) which we have not measured into the mo del in addition to Month, Tree and Rain. Similarly, changing the link function can be another possible reason which should be taking into account. Moreover, transforming one or more covariate variables play an essential role, such as taking logs, power and so on.

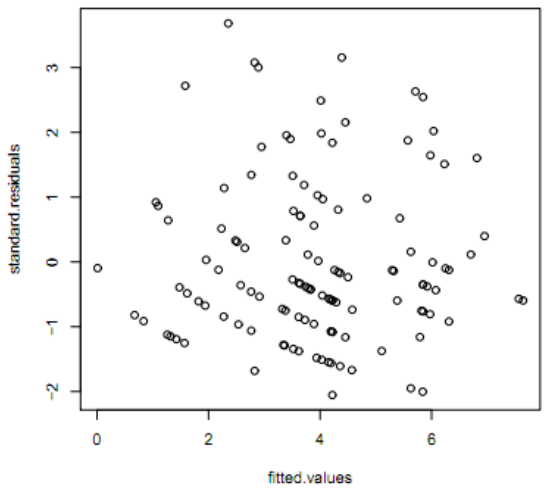


Figure 3.1: Residual plot of Model 7

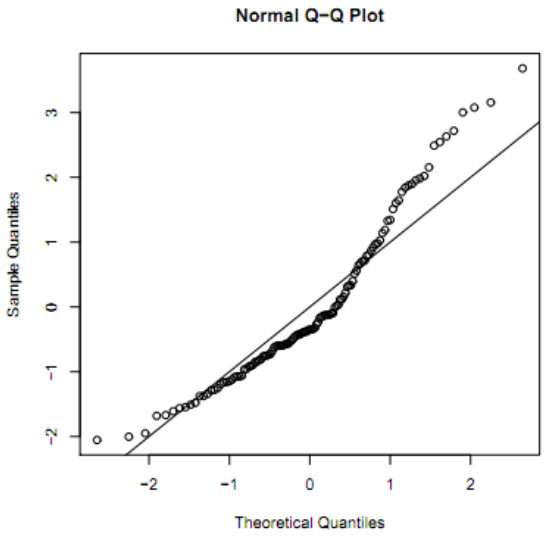


Figure 3.2: Q Q plot of Model 7

3.4 Negative Binomial Distribution

We have completed the task of fitting a Poisson model to the child admission data. We have been successful in choosing the best model, analyzing the residuals and using the model for prediction. The only problem we have encountered is over-dispersion. One possible explanation is that we do not have the in uential explanatory variables such as genic conditions of the child and family background etc to fit the model accurately. Orales, the model is incorrect. For count data, the Poisson distribution is a good starting point but there are alternatives, such

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

as the negative binomial distribution (Spiegel; 1992). It is more complicated to fit the negative binomial distribution because it has an additional parameter. We can either specify or we can estimate using maximum likelihood. we wish to estimate as well then we use a slightly diverent function. In the previous poisson model , we have identified the important explanatory variables which are Month, Tree and Rain as shown in Table1. We shall now repeat the mo del selection procedure using the negative binomial distribution. Lets start with the null model. Null model Model 9: (Child) ~ 1.36990and twice log-likelihood is -582.833 on 122 degrees of freedom. In addition to the usual GLM output, R has estimated ˆ= 3.183. It has also stated the log-likelihood which we shall need for comparing models. Now we shall try adding Month. Model 10: (Child) ~ 1.37820 - 0.62253 * (Jul) - 0.05859 * (Oct) + 0.40799 * (Sep) and twice log-likelihood is -552.494 on 119 degrees of freedom. We compare Model 9 and Model 10 using twice the difference in the log-likelihood, and we obtain p value < 0.05. So we should keep Month in the model. Then lets add Tree into the model. Model 11: (Child) ~ 1.506873 - 0.510300 * (Jul) - 0.076088 * (Oct) + 0.332850 *(Sep) - 0.017654 * (Tree) and twice log-likelihood is -548.3317 on 118 degrees of freedom. After compare the twice log-likelihood of Model 10 and Model 11 we find that p value = 0.04133351 < 0.05. Therefore, Model 11 should be accepted. Then we add Rain into the model. Model 12: (Child) ~ 1.572649 - 0.381160 * (Jul) - 0.035897 * (Oct) + 0.326558 * (Sep) - 0.021254 * (Tree) - 0.019650 * (Rain) and twice log-likelihood is -544.3666 on 117 degrees of freedom. After compare the twice log-likelihood of Model 11 and Model 12 we find that p value = 0.04645403 <0.05, the increase of the log-likelihood is significant with 1 degrees of freedom and hence Model 12 should be accepted. Model 13: (Child) ~ 1.282943 + 0.121769 * (J ul) + 0.347593 * (Oct) + 0.472424 * (Sep) + 0.019140 * (Tree) - 0.022832 * (Rain) - 0.056215 * (Jul * T ree) - 0.053416 * (Oct * Tree) + 0.009852 * (Sep * Tree) with twice log-likelihood -536.8628 on 114 degrees of freedom. Compared with Model 12 using twice the dierence in the log-likelihoo d we found that p value = 0.0574601 > 0.05. It can be seen that, with contrast to the poisson mo del, adding interacion term between Month and Tree do es not produce a significant result under negative binomial family. This is an unexpected outcome. Hence we will go back to Model 12. Lets consider the interaction between Month and Rain. Model 14: (Child) ~1.560644 - 0.439498 * (Jul) - 0.046377 * (Oct) + 0.429011 * (Sep) -0.019761 * (Tree) - 0.019117 * (Rain) + 0.006564 * (Jul * Rain) + 0.002673 * (Oct *Rain) - 0.039490 * (Sep * Rain) with twice log-likeliho od -542.4905 on 114 degrees of freedom. Compared with Model 12 using twice the di erence in the log-likelihood we found that p v alue = 0.5985152 > 0.05. It can bee seen that adding interaction term between Month and Rain do es not make significant either. Hence we root back to Model 12. Lets add interaction between Tree and Rain to Mo del 12. Model 15: (Child) ~ 1.515749 - 0.370363 * (Jul) - 0.040241 * (Oct) + 0.347589 * (Sep) - 0.011021 * (Tree) - 0.003638 * (Rain) - 0.004069 * (Tree * Rain) with twice log-likelihood -540.3500 on 116 degrees of freedom. Compared with Model 12 using twice the difference in the log-likelihood we found that p value = 0.0450525 < 0.05. Therefore, addition interaction between Tree and Rain became significant. Hence we accept more complicated model, Model 15. We have considered all the possible interaction between predictor variables that we found important and it is clear that Model 15 is less complicated and more realistic than poisson model, Model 7.

3.5 Model Checking(Negative Binomial Distribution)

We now need to check the residual plots using the standard residuals. However, variance of the new model (1.04451) is much smaller than the best

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

chosen model under poisson family. Additionally, overdisperson is reduced under new model hence overall best chosen model will be Model 15.

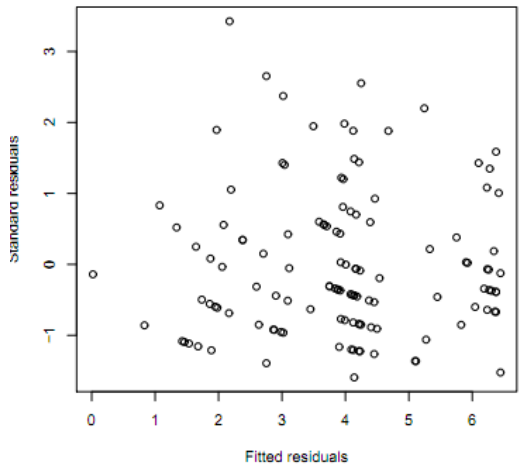


Figure 3.3: Residual plot of Model 15

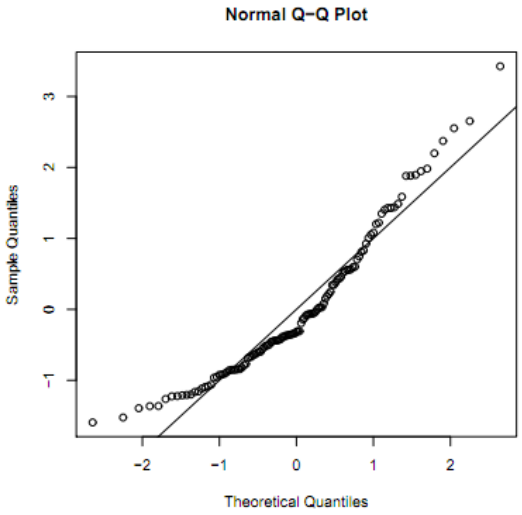


Figure 3.4: QQ plot of Model 15

Conclusion

As above, various environmental factors have been considered for the cause of asthma among children in Mexico and it can be seen that, amount of Rain and Tree pollen arise as major factors. According to the analysis, month of September recorded the highest number of child admission from asthma. This does not contradict with our finding as in September dry season is about to start in Mexico and Rainfall and Tree-pollen concentration drop. As we know, Asthma is common condition that a ects the air ways in the lungs. Possible explanation could be when the Rainfall and tree pollen drops, there is a lack of fresh air and children become vulnerable for asthma. On the other hand, several epidemiologic studies have investigated the statistical relationships between asthma and air pollution. However, in our studies, air pollution did not contribute significantly to child admission from asthma. It could be the fact that adults are more vulnerable to air pollution than children. Finally, above analysis have been carried out based on the environmental factors and we believe that there are other important factors such as genetic conditions and family backgrounds of the children need to be accounted along with environmental factors . According to a survey carried out by Asthma and Allergy Association in America, young people with asthma (those aged 15-24 years) showed that more blacks than whites died of the disease from 1980 to 1993. Among children aged 0-4 years in 1993, blacks were six times more likely to die from asthma than whites. Among children aged 5-14, blacks were four times more likely than whites to die of the illness. Hence it shows that genetic condition may play an important role in this contest. However, cost of involving extracting those details from each individual could be higher and time consuming. Nevertheless, based on the resources we have been provided, it is apparent that Tree, Rain and child asthma are correlated each other.

Recommendation

Asthma plays an important role in controlling and co-existing with asthma if the following guidelines are followed:

Prevention of internal and external allergies

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

The children must have a medical file in the hospital and the health center.

Do not dispense medicines on your own or buy them from the pharmacy without consulting your doctor.

Keep your health and fitness healthy and exercise.

Quitting smoking, abstaining from smoking, and avoiding factors leading to an asthma attack.

Take a seasonal flu vaccine to reduce the risk of flu.

REFERENCES

1. Agertoft, L, Pedersen, S. (1994). Effects of long-term treatment with an inhaled corticosteroid on growth and pulmonary function in asthmatic children. Respir Med 1994;88(5):373–81.

2. Busse, W. (1996). The role of leukotrienes in asthma and allergic rhinitis. Clin Exp Allergy 1996;26(8):868–79. Review

3. Covar, RA, Spahn JD, Murphy JR, Szefler SJ (2004); Childhood Asthma Management Program Research Group. Progression of asthma measured by lung function in the childhood asthma management program. Am J Respir Crit Care Med 2004;170(3):234–41. Epub March 2004.

4. Holgate, S (1999). Genetic and environmental interaction in allergy and asthma. J Allergy Clin Immunol 1999;104(6):1139–46. Review.

5. Gaur, A. S., and Gaur, S. S. (2006) *Statistical Methods for Practice and Research*. London: Sage Publication.

6. Howell, D. C. (2010) *Statistical Method forPsychology*. [online] available from <http://books.google.co.uk/books?hl=en&lr=&id=5WFohzuwzP0C&oi=fnd&pg=PR9&dq=statistical+method+for+pschology&ots=oRFhI4PlcQ&sig=ts4z_RgXNplCYK8cNJ0TLw2V1NY#v=onepage&q&f=false

7. Little. R & Rubin. D (1987). Statistical Analysis with Missing Data. 2nd

8. ed. John Wiley & Sons. Hoboken. New Jersy

9. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ (2007). Asthma and wheezing in the first six years of life. The Group Health Medical Associates. N Engl J Med 1995;332(3):133–8

10. Mann, P. S. (2007*) Introductory Statistics*. United States: Laurie Rosatone.

11. Spiegel, M. R. (1992) *Schaum's Outline of Theory and Problem of Statistics*. UK: MsGraw-Hill International.

Appendix A

R Codes

A.1 Exploratory data analysis

Histograms

```
hist(Child,breaks=12)

hist(Deuterom,breaks=10)

hist(ASC,breaks=14)

hist(Basidiom)

hist(Tree,breaks=12)

hist(Weed)

hist(Grass)

hist(Tpollen,breaks=14)

hist(MaxTemp)

hist(MinTemp)

hist(RH,breaks=14)

hist(Rain,breaks=14)

hist(O3)

hist(NO2)

hist(SO2)

hist(O3hours)
```

Boxplots

```
boxplot(Child,horizontal=T,main="Child")

boxplot(Deuterom,horizontal=T,main="Deuterom")

boxplot(ASC,horizontal=T,main="ASC")
```

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

```
boxplot(Basidiom,horizontal=T,main="Basidiom")
boxplot(Tree,horizontal=T,main="Tree")
boxplot(Weed,horizontal=T,main="Weed")
boxplot(Grass,horizontal=T,main="Grass")
boxplot(Tpollen,horizontal=T,main="Tpollen")
boxplot(MaxTemp,horizontal=T,main="MaxTemp")
boxplot(MinTemp,horizontal=T,main="MinTemp")
boxplot(RH,horizontal=T,main="RH")
boxplot(Rain,horizontal=T,main="Rain")
boxplot(O3,horizontal=T,main="O3")
boxplot(NO2,horizontal=T,main="NO2")
boxplot(SO2,horizontal=T,main="SO2")
boxplot(O3hours,horizontal=T,main="O3hours")
### QQ plots ###
sqqnorm(scale(Child,T,T),main="Child") abline(0,1)
qqnorm(scale(Deuterom,T,T),main="Deuterom")
abline(0,1)
qqnorm(scale(ASC,T,T),main="ASC") abline(0,1)
qqnorm(scale(Basidiom,T,T),main="Basidiom")
abline(0,1)
qqnorm(scale(Tree,T,T),main="Tree") abline(0,1)
qqnorm(scale(Weed,T,T),main="Weed") abline(0,1)
qqnorm(scale(Grass,T,T),main="Grass") abline(0,1)
qqnorm(scale(Tpollen,T,T),main="Tpollen")
abline(0,1)
qqnorm(scale(MaxTemp,T,T),main="MaxTemp")
abline(0,1)
qqnorm(scale(MinTemp,T,T),main="MinTemp")
abline(0,1)
qqnorm(scale(RH,T,T),main="RH") abline(0,1)
qqnorm(scale(Rain,T,T),main="Rain") abline(0,1)
qqnorm(scale(O3,T,T),main="O3") abline(0,1)
qqnorm(scale(NO2,T,T),main="NO2") abline(0,1)
qqnorm(scale(SO2,T,T),main="SO2") abline(0,1)
qqnorm(scale(O3hours,T,T),main="O3hours")
abline(0,1)
```

Scatter plots

```
plot(Child~Month,main="Child~Month")
plot(Child~Deuterom,main="Child~Deuterom")
plot(Child~ASC,main="Child~ASC")
plot(Child~Basidiom,main="Child~Basidiom")
plot(Child~Tree,main="Child~Tree")
plot(Child~Weed,main="Child~Weed")
plot(Child~Grass,main="Child~Grass")
plot(Child~Tpollen,main="Child~Tpollen")
plot(Child~MaxTemp,main="Child~MaxTemp")
plot(Child~MinTemp,main="Child~MinTemp")
plot(Child~RH,main="Child~RH")
plot(Child~Rain,main="Child~Rain")
plot(Child~O3,main="Child~O3")
plot(Child~NO2,main="Child~NO2")
plot(Child~SO2,main="Child~SO2")
plot(Child~O3hours,main="Child~O3hours")
```

A.2 Generate models

Poisson distribution

```
lm0<-glm(Child~1,family=poisson) #Null model,
Model 0.
summary(lm0) #Find deviance and degrees of
freedom for Model 0.
lm1<-glm(Child~Month,family=poisson) #Generate
Model 1.
```

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

```
summary(lm1) #Find deviance and degrees of freedom for Model 1.

anova(lm0,lm1,test="Chisq") #Compare Model 0 and Model 1.

#Models with one explanatory variable.

glm(Child~Deuterom,family=poisson)

glm(Child~ASC,family=poisson)

glm(Child~Basidiom,family=poisson)

glm(Child~Tree,family=poisson)

glm(Child~Weed,family=poisson)

glm(Child~Grass,family=poisson)

glm(Child~MaxTemp,family=poisson)

glm(Child~MinTemp,family=poisson)

glm(Child~RH,family=poisson)

glm(Child~Rain,family=poisson)

glm(Child~O3,family=poisson)

glm(Child~NO2,family=poisson)

glm(Child~SO2,family=poisson)

glm(Child~O3hours,family=poisson)

glm(Child~Tpollen,family=poisson)

lm2<-glm(Child~Month+Tree,family=poisson)
#Generate Model 2.

summary(lm2) #Find deviance and degrees of freedom for

Model 2.

anova(lm1,lm2,test="Chisq") #Compare Model 1 and Model 2.

lm3<-
glm(Child~Month+Tree+Tpollen,family=poisson)

#Generate Model 3.

summary(lm3)

anova(lm2,lm3,test="Chisq") #Compare Model 2 and Model 3.

lm4<-
glm(Child~Month+Tree+Rain,family=poisson)
#Generate Model 4.

summary(lm4)

anova(lm2,lm4,test="Chisq") #Compare Model 2 and Model 4.

lm5<-
glm(Child~Month+Tree+Rain+Month*Tree,family=
poisson)

#Generate Model 5.

summary(lm5)

anova(lm4,lm5,test="Chisq") #Compare Model 4 and Model 5.

lm6<-
glm(Child~Month+Tree+Rain+Month*Tree+Month
*Rain,family=poisson)

#Generate Model 6.

summary(lm6)

anova(lm5,lm6,test="Chisq") #Compare Model 5 and Model 6.

lm7<-
glm(Child~Month+Tree+Rain+Month*Tree+Tree*R
ain,family=poisson)

#Generate Model 7.

summary(lm7)

anova(lm5,lm7,test="Chisq") #Compare Model 6 and Model 7.

lm8<-glm(Child~Month*Tree*Rain,family=poisson)
#Generate Model 8.

summary(lm8)

anova(lm7,lm8,test="Chisq") #Compare Model 7 and Model 8.

### Negative binomial distribution ###
```

May – June 2020
ISSN: 0193-4120 Page No. 2904 – 2917

```
library(MASS)

nb1<-glm.nb(Child~1) #Null model, Model 9.

summary(nb1) #Find deviance and degrees of freedom for Model 9.

nb2<-glm.nb(Child~Month) #Generate Model 10.

summary(nb2)

anova(nb1,nb2,test="Chisq") #Compare Model 9 and Model 10.

nb3<-glm.nb(Child~Month+Tree) #Generate Model 11.

summary(nb3)

anova(nb2,nb3,test="Chisq") #Compare Model 10 and Model 11.

nb4<-glm.nb(Child~Month+Tree+Rain) #Generate Model 12.

summary(nb4)

anova(nb3,nb4,test="Chisq") #Compare Model 11 and Model 12.

nb5<-
glm.nb(Child~Month+Tree+Rain+Month*Tree)
#Generate Model 13.

summary(nb5)

anova(nb4,nb5,test="Chisq") #Compare Model 12 and Model 13.

nb6<-
glm.nb(Child~Month+Tree+Rain+Month*Rain)
#Generate Model 14.

summary(nb6)

anova(nb4,nb6,test="Chisq") #Compare Model 12 and Model 14.

nb7<-glm.nb(Child~Month+Tree+Rain+Tree*Rain)
#Generate Model 15.

summary(nb7)

anova(nb4,nb7,test="Chisq") #Compare Model 12 and Model 15.
```

A.3 Check Models

```
### Poisson distribution ###

fitted.values<-lm8$fitted.values #Find fitted values of Model 8.

standard.residuals<-(Child-
fitted.values)/sqrt(fitted.values)

#Calculate standard residuals.

plot(fitted.values,standard.residuals)

qqnorm(standard.residuals) abline(0,1)

### Negative binomial distribution ###

fv<-nb7$fitted.values #Find fitted values of Model 15.

theta<-nb7$theta #Get estimated theta of Model 15.

standard.residuals<-(Child-fv)/sqrt(fv+(fv^2)/theta)

#Calculate standard residuals.

plot(fv,standard.residuals,xlab="Fitted residuals",

ylab="Standard residuals")

qqnorm(standard.residuals) abline(0,1)
```