Journal of the Kerala Statistical Association
Vol.34, December 2023, 39-51 *ISSN 2249-4553*

A Mixture Regression Approach for Modelling Early Post Operative Hypocalcemia

Sreelaya K*[1], Yadev I.[2], and Sebastian George[1]

[1]Department of Statistical Sciences, Kannur University, India
[2]Government Medical College, Pariapally, Kollam, India

ABSTRACT

In this paper, we show that a two component Laplace mixture model is an appropriate distribution to model postoperative calcium levels of patients undergoing thyroidectomy or surgery for thyroid diseases. A mixture regression model is constructed to predict postoperative calcium levels based on a pre-determined set of potential predictors. The parameters of the model are estimated using EM algorithm. Our study based on a real data set shows that two component Laplace mixture regression model is suitable for prediction and interpretation compared to the usual Gaussian mixture regression model.

Key words and Phrases: *EM algorithm, Gaussian mixture models, Laplace mixture models, Laplace mixture regression models, Post operative hypocalcemia, Thyroidectomy*
.

1 INTRODUCTION

Thyroidectomy results in many morbid complications to the patients. Though the incidence of permanent complication like recurrent laryngeal nerve palsy and post operative hypocalcemia is rare (1 to 5%), it portends a devastating effect to the individual. However the most common immediate complication is transient postoperative hypocalcemia

*sreelaya1997@gmail.com

(10 to 45%). Hypocalcemia result in much anguish and frequent visit to emergency room in addition to added cost and lost work hours to the patient. Importance of this early post operative complication is the necessity to frequent visits to casualty due manifestations of hypocalcemia like tetany, cramps etc. Acute hypocalcemia, or low calcium levels in the blood, is a serious medical condition that can cause tetany, a condition characterized by neuromuscular irritability. Mild symptoms of tetany include numbness and muscle cramps, while severe symptoms include seizures and laryngospasm. Patients with acute hypocalcemia may also experience fatigue, anxiety, and depression, in addition to the specific symptoms listed above. Not every patient with severe hypocalcemia will exhibit neuromuscular symptoms. The underlying mechanism of tetany in acute hypocalcemia is the hyperexcitability of the nervous system at multiple levels, including the spinal reflexes, motor endplates, and the central nervous system (up to date). Temporary hypocalcemia, or low calcium levels in the bloodstream, is a common complication following thyroidectomy, but its exact cause is unclear. Prior research has demonstrated that patients who undergo unilateral lobectomy, or the removal of one thyroid lobe, may experience a decrease in total calcium levels due to a decrease in albumin levels, while free calcium, parathyroid hormone, and calcitonin levels remain unchanged. In contrast, patients who undergo bilateral lobectomy, or the removal of both thyroid lobes, may experience a reduction in total calcium levels due to a reduction in albumin-bound calcium levels and a reduction in free calcium levels due to a reduction in parathyroid hormone levels. The levels of calcitonin do not appear to differ between the two groups. Despite efforts to preserve the parathyroid glands and their blood supply during surgery, bilateral lobectomy frequently results in temporary hypoparathyroidism. However there are studies which implicated the role of calcitonin in post operative hypocalcemia. Many factors are implicated in the cause of immediate or early post operative hypocalcemia. Dissection of the tissues, vascular compromise, surgeon experience and use of diathermy devices are some of the causative factors associated with this temporary complication. Old age, the type of procedure, the amount of time spent in surgery, neck dissection, the histology of the surgical material, and vocal fold paralysis are also contributing factors. A decrease in calcium levels post-operatively in comparison to the immediate pre-operative values can act as a practical and straightforward predictor of hypocalcemia in patients following complete thyroidectomy. The assessment of postoperative calcium levels compared to immediate preoperative levels has been identified as a useful and simple predictor of hypocalcemia in patients undergoing total thyroidectomy. Transient hypocalcemia, a common complication after thyroidectomy, generally responds favorably to replacement therapy within a few days or weeks. The primary cause of hypocalcemia is typically sec-

ondary hypoparathyroidism, which results from damage or devascularization of one or more parathyroid glands during surgery. In some cases, erroneous parathyroid removal may also contribute to hypocalcemia. Several studies have investigated the use of serum calcium and parathyroid hormone levels as predictors of hypocalcemia after thyroidectomy. These findings highlight the importance of closely monitoring calcium levels in patients undergoing thyroid surgery to promptly identify and treat hypocalcemia.

In the medical literature, postoperative calcium levels are typically categorized as either hypocalcemia or normal calcemia based on established cut-off values. While this approach is useful as a clinical decision rule, it does result in the loss of important information and the inability to model continuous calcium values with appropriate models. Most studies on hypocalcemia have utilized logistic or Cox regression models to analyze the data. However, the distribution of postoperative calcium level look like a mixture distribution and a mixture regression approach to this problem seems to be more appropriate. The exploration of postoperative calcium as a mixture has not been attempted in the statistical / medical literature to date. The dichotomization of calcium levels into hypocalcemia or normal calcemia may not fully capture the complexity of this phenomenon and may limit our understanding of its underlying mechanisms and potential predictors. Therefore, there is a need for alternative approaches that can better model the continuous nature of postoperative calcium levels and provide a more comprehensive understanding of this condition. The primary objective of this study is to develop a statistical model using a mixture model approach to predict postoperative calcium levels based on a pre-determined set of potential predictors. Specifically, we will utilize mixture regression to predict postoperative calcium levels from a set of candidate predictors.

2 MATERIALS AND METHODS

2.1 Data Generation Methods

This study is a secondary data analysis of the data collected for another study on thyroid at Goverment Medical College, Trivandrum during the year 2017 to 2021. It was conducted in accordance with the declaration of Helsinki. The histology record from the Department of Pathology was appended to the list that was obtained by searching for diagnosis in the hospital information system in order to construct a list of all cases involving thyroidectomy. The registry kept at the operating theater was also abstracted in search of prospective thyroidectomy cases. In addition, a search was conducted on the computer at the recorded library using ICD codes. A consolidated list that was put together using the sources mentioned above. Case records were consecutively retrieved from the case record library, following the sequence in the register of thyroidectomy that was constructed from the sources that were discussed earlier in this paragraph. The retrieval of inpatient records for the primary study had both administrative and ethical clearance before proceeding.

For the purposes of this study, we collected data from the medical records of all patients who underwent thyroidectomy during the specified time period. In order to be eligible for inclusion in the study, patients had to be at least 13 years of age. Patients who had undergone thyroidectomy in combination with neck dissection or who had undergone revision thyroidectomy were not included in the study, as they had already received treatment for their condition. Thyroidectomy is a commonly performed procedure in hospitals, and it can be carried out by both experienced and inexperienced surgeons. At our facility, the preferred method for thyroidectomy is the capsular technique (Lemin 2011 and Delbridge 1992). This procedure has been shown to be effective in the treatment of thyroid conditions, and it does not typically result in recurrent laryngitis or any damage to the parathyroid glands.

2.2 Mixture Regression

2.2.1 Finite Mixture Models

Finite mixture models have received wide attention in Statistics literature as it is a versatile method for modelling data that comes from a heterogeneous population with multiple underlying sub populations, which is the primary cause of the heterogeneity of observed outcome. Finite mixture models are able to capture the more specific properties of the observed data, such as multi modality, heavy tails and/or skewness. However, the asymmetry and shape variations, possibly due to intrinsic aspects of the data, can indicate that not all observations come from the same population. In other words, a data set can be a mixture of observations from various populations. Mixture models can be applied in various fields such as astronomy, genetics, psychiatry, engineering, economics and in various physical and social sciences. Among these the most important application is in medical sciences. This motivated us to look into mixture models.

A finite mixture model is given by,

$$f(y) = \sum_{i=1}^{g} \pi_i f_i(y) \tag{2.1}$$

where $f_i(y)$, i=1,2,...,g are densities and π_i are non-negative quantities with $0 \leq \pi_i \leq 1;\ (i = 1, 2, ..., g)$ such that $\sum \pi_i = 1$.

$\pi_1, \pi_2, ..., \pi_g$ are known as mixing coefficients and $f_i(y)$, $(i = 1, , ..., g)$ are component densities of the mixture.

In general a probability model is said to be a mixture model if the density function can be represented as

$$f(y) = \int h(y; \theta) dG(\theta) \tag{2.2}$$

where the integral represents the Riemann-Stieltje's integral with respect to the distribution function G and $h(y; \theta)$ is a density function. G is known as the mixing distribution and (2.2) becomes of the form (2.1), when G is a discrete distribution function. For various theoretical issues and estimation procedures related to mixture distributions we refer to Mclachlan and Peel (2000). For the hypocalcemia data considered in this paper, we found that a two component Laplacian mixture model is an appropriate model compared to the conventional finite Gaussian mixture.

2.2.2 Laplace mixture models

The density function of a two parameter Laplace distribution is given by

$$f(x, \theta) = \frac{1}{2\sigma} exp\Big(- \frac{|y - \mu|}{\sigma} \Big) \tag{2.3}$$

where $-\infty < y < \infty$, $-\infty < \mu < \infty$ and $\sigma > 0$.

Now consider a two component Laplace mixture model. Let p be the mixing coefficient related to the first component and μ_1, σ_1 be the corresponding location and scale parameters . Let $q = 1 - p$ be the mixing coefficient of second component and μ_2, σ_2 be the corresponding location and scale parameters. Then, the two component Laplace mixture density is given by

$$f(y, \underline{\theta}) = \frac{p}{2\sigma_1} e^{(-\frac{|y-\mu_1|}{\sigma_1})} + \frac{1-p}{2\sigma_2} e^{(-\frac{|y-\mu_2|}{\sigma_2})} \tag{2.4}$$

where $-\infty < y < \infty$, $-\infty < \mu_1$, $\mu_2 < \infty$, $\sigma_1, \sigma_2 > 0$ and $0 < p < 1$.

The mean and variance of the distribution is given by

$$\mu = p\mu_1 + (1 - p)\mu_2 \tag{2.5}$$

$$\sigma^2 = p({\mu_1}^2 + 2{\sigma_1}^2) + (1 - p)\{{\mu_2}^2 + 2{\sigma_2}^2\} - \{p\mu_1 + (1 - p)\mu_2\}^2\}. \tag{2.6}$$

2.2.3 Mixture Regression

Due to the heterogeneity of units in the population as in the case of hypocalcimia as the response variable, the response variable will follow mixture distribution. The usual

regression procedures may produce an unreasonable result even when measurements on some useful explanatory variables are available as these procedures are not taking into account the group structure in the population. In mixture regression models we first identify the heterogeneity of units in the population and then explain the relationship between the response variable and the explanatory variable. Mainly the application of mixture regression models are in medical field.
Mathematical form of finite mixture regression model is given by,

$$f(y|x;\psi) = \sum_{i=1}^{g} \pi_i f_i(y|x;\theta_i) \tag{2.7}$$

$$\pi_i \geq 0 \quad \sum_{i=1}^{g} \pi_i = 1$$

Here y is the response variable with conditional density f, x is a vector of independent variables, π_i is the prior probability of component i, θ_i is the component specific parameter vector for the density function f_i, and $\psi = (\pi_1, \pi_2, ..., \pi_{g-1}, {\theta_1}^T, {\theta_2}^T, ..., {\theta_g}^T)^T$.
The conditional expectation or the mixture regression of Y on X is $E(Y|X)$ which will be a convex combination of g functions of the explanatory variables. Our aim is to estimating the parameters in these functions and the mixing coefficients $\pi_1, \pi_2, ..., \pi_{g-1}$.

2.2.4 Laplace Mixture Regression

Let x be a p dimensional explanatory variable and y be the response variable and z is the latent class variable which is independent of x deciding the group structure. If response variable y depends in a linear way on explanatory variable x given $Z = i$ as

$$y = X'\beta_i + \epsilon_i, \qquad i = 1, 2, ..., g. \tag{2.8}$$

Where β_i's are known as regression coefficients, g is the number of components, ϵ is the error term.
Let the sample be $S = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ to estimate the parameter θ. Thus the log likelihood function to estimate $\theta = (\beta_1, \sigma_1^2, \pi_1, \beta_2, \sigma_2^2, \pi_2, ..., \beta_g, \sigma_g^2, \pi_g)$ is given by,

$$logL = \sum_{j=1}^{n} log \sum_{i=1}^{g} \frac{\pi_i}{2\sigma_i} exp\left(\frac{-|y_j - {x_j}'\beta_i|}{\sigma_i}\right) \tag{2.9}$$

By maximizing the above likelihood, the MLE of θ can be obtained. But since the log likelihood function does not have an explicit solution, we have to use some kind of numerical methods such as EM algorithm to estimate θ. Let ϵ_i's in model (2.7) follows

Laplace distribution. Then let Z_{ij} denote latent Bernoulli variables such that,

$$Z_{ij} = \begin{cases} 1 & if\ j^{th}\ observation\ is\ from\ i^{th}\ component \\ o & otherwise \end{cases} \tag{2.10}$$

The complete log likelihood function of θ if the full data set is observable is given by

$$logL_c = \sum_{j=1}^{n}\sum_{i=1}^{g} Z_{ij} log \frac{\pi_i}{2\sigma_i} exp\left(\frac{-|y_j - x_j{}'\beta_i|}{\sigma_i}\right) \tag{2.11}$$

Now we will maximize the complete log likelihood function given by equation (10).

Andrews and Mallows (1974) showed that a Laplace distributed random variable can be expressed as a mixture of a normally distributed random variable and another variable related to exponential distribution. Let W_j be the latent variable coupled with (X_j, Y_j), $j = 1, 2, ..., n$, then complete log likelihood can be written as,

$$\begin{aligned} logL_c &= \sum_{j=1}^{n}\sum_{i=1}^{g} Z_{ij} log\pi_i \frac{W_j}{\sqrt{\pi}\sigma_i} exp\left(-\frac{W_j^2(Y_j - X_j'\beta_i)^2}{\sigma_i^2}\right)\frac{1}{W_j^3} exp(-\frac{1}{2W_j^2}) \\ &= \sum_{j=1}^{n}\sum_{i=1}^{g} Z_{ij} log\pi_i - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{g} Z_{ij} log\pi\sigma_i^2 - \sum_{j=1}^{n}\sum_{i=1}^{g} Z_{ij}\frac{W_j^2(Y_j - X_j'\beta_i)^2}{\sigma_i^2} \\ &\quad - \sum_{j=1}^{n}\sum_{i=1}^{g} Z_{ij} log W_j^2 - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{g}\frac{Z_{ij}}{W_j^2} \end{aligned} \tag{2.12}$$

Based on EM algorithm principle (Mclachlan and Krishanan, 2008 and Philips, 2002), for the E-step, we have to calculate the conditional expectation $E[L(\theta; D)|S, \theta^{(0)}]$, where $\theta^{(0)} = (\beta_1^{(0)}, \sigma_1^{2(0)}, \pi_1^{(0)}, ..., \beta_g^{(0)}, \sigma_g^{2(0)}, \pi_g^{(0)})$ and $S = (X_j, Y_j)_{j=1}^n$.
We need to find the following two terms to calculate $E[L(\theta; D)|S, \theta^{(n)}]$.
Hence we obtain

$$\tau_{ij}^{(1)} = \frac{\pi_i^{(0)}\sigma_i^{-1(0)} exp(-\sqrt{2}|Y_j - X_j'\beta_i^{(0)}|/\sigma_i^{(0)})}{\sum_{m=1}^{g}\pi_m^{(0)}\sigma_m^{-1(0)} exp(-\sqrt{2}|Y_j - X_j'\beta_m^{(0)}|/\sigma_m^{(0)})}$$

Similarly

$$\delta_{ij}^{(1)} = \frac{\sigma_i^{(0)}}{\sqrt{2}|Y_j - X_j'\beta_i^{(0)}|}$$

. In M step we maximize following expression with respect to $\pi_i, \beta_i, \sigma_i^2$.

$$\sum_{j=1}^{n}\sum_{i=1}^{g}\tau_{ij} log\pi_i - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{g}\tau_{ij} log\sigma_i^2 - \sum_{j=1}^{n}\sum_{i=1}^{g}\frac{\tau_{ij}\delta_{ij}(Y_j - X_j'\beta_i)^2}{\sigma_i^2}$$

Which can be maximized by using the following formulae

$$\Pi_i^{(k+1)} = \frac{1}{n}\sum_{j=1}^{n}\tau_{ij}^{(k+1)}$$

$$\beta_i^{(k+1)} = (\sum_{j=1}^{n}\tau_{ij}^{(k+1)}\delta_{ij}^{(k+1)}X_jX_j')^{-1}(\sum_{j=1}^{n}\tau_{ij}^{(k+1)}\delta_{ij}^{(k+1)}X_jY_j)$$

and

$$\sigma_i^{2(k+1)} = \frac{2\sum_{j=1}^{n}\tau_{ij}^{(k+1)}\delta_{ij}^{(k+1)}(Y_j - X_j'\beta_i^{(k+1)})^2}{\sum_{j=1}^{n}\tau_{ij}^{(k+1)}}$$

When all σ_i^2's are equal then σ^2, a common initial value of σ^2 is used. i.e.,

$$\sigma^{2(k+1)} = \frac{2\sum_{j=1}^{n}\sum_{i=1}^{g}\tau_{ij}^{(k+1)}\delta_{ij}^{(k+1)}(Y_j - X_j'\beta_i^{(k+1)})^2}{n}$$

. From the adoption of Least absolute Deviations (LAD) regression the above said EM algorithm follows robustness as noted by Song et al. (2014), Zakiah et al. (2019) and Bai et al. (2002).

3 RESULTS AND DISCUSSION

3.1 Preliminary Analysis

Figure 1 shows the distribution of post operative calcium levels of 777 patients undergone operations due thyroid complications. From the figure it is evident that the distribution is non-Gaussian and bimodal. Summary statistics related to the explanatory variables such as sex, type of tumor, age of the patients, post operative calcium levels and weight of thyroid specimen are given in Table 1. Table 2 shows the beta coefficients, their 95% confidence intervals and the P-values to check the significance of these beta coefficients related to a multiple linear regression analysis. The results show that none of these beta coefficients are significant. The correlation plot given in figure 2 also support this result. This is a usual scenario in the case of mixture regression models and hence we tried mixture regression for prediction and analysis.

Gender		Type of tumor	
Female	90%	Benign	17 %
Male	10%	Malignant	83 %

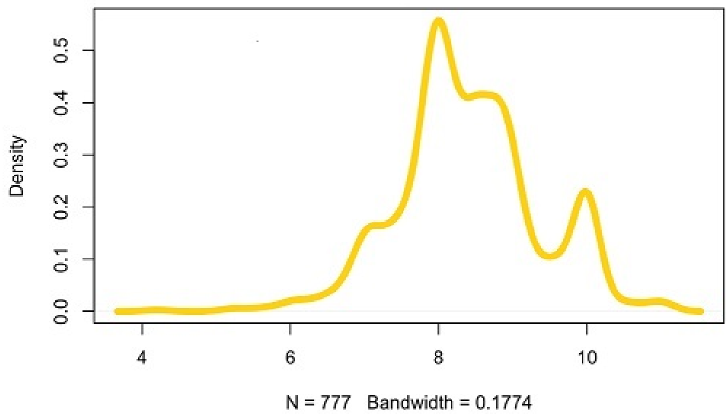

Figure 1: Distribution of postoperative calcium levels

	Age of the patient	Postoperative calcium levels	Weight of thyroid specimen
Mean	41	8.40	43.5
median	40	8.30	30
Range	[13,81]	[4.2,11]	[5,850]

Table 1: Data Characteristics

Independent variables	β	95 % CI	p value
Age	0.001	[-0.003,0.0006]	0.58
Gender	0.130	[-0.059,0.318]	0.18
Weight of thyroid specimen	0.000	[-0.001,0.001]	0.55
Malignant or Benign	-0.099	[-0.252,0.054]	0.20

Table 2: Linear regression

3.2 Mixture Models to Post Operative Calcium Levels of Thyroid Patients

Mixture models were fitted to the data described in table 1. First we fitted a two component Gaussian mixture model to the postoperative calcium level of patients by estimating the parameters using EM algorithm. A similar analysis was carried out under the assumption that the response variable follows a Laplace mixture model. We computed the AIC an BIC for each model and their values strongly support the use of a two component Laplacian mixture regression model for better prediction.

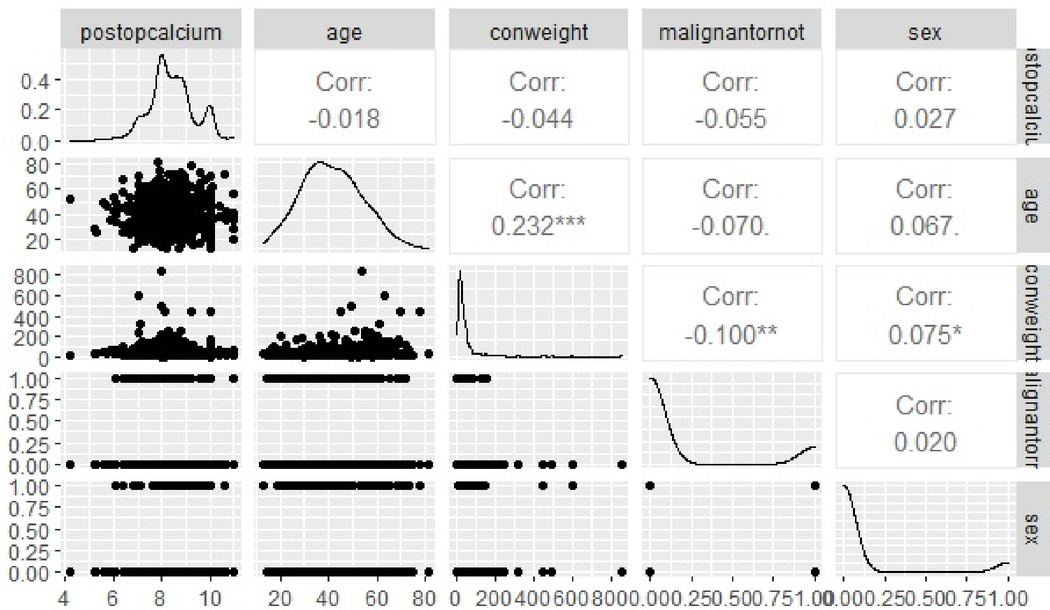

Figure 2: Correlation plot showing mixture structure

Models	p		μ		σ		AIC	BIC
	comp 1	comp 2	comp 1	comp 2	comp 1	comp 2		
Two component normal mixture	0.25185	0.74815	8.17697	8.47387	0.38635	1.0665	2207.376	2230.653
Two component Laplace mixture	0.52549	0.47451	8	8.9	0.78579	1.08138	2167.042	2191.581

Table 3: Comparison of Gaussian and Laplace Mixture Models .

3.3 Laplace Mixture Regression Model for the Thyroid Data

Taking the postoperative calcium level of the thyroid patients as the response variable and age, weight of the thyroid specimen, gender of the patient and whether the thyroid specimen is malignant or not as explanatory variables, we carried out mixture regression analysis under the assumption that the response variable follows a two component Laplacian model. For comparison purpose, we carried out the same type of anlysis under a two component Gaussian mixture.

Here p is the mixture proportion, β_i's are regression coefficients and σ is the standard deviation.

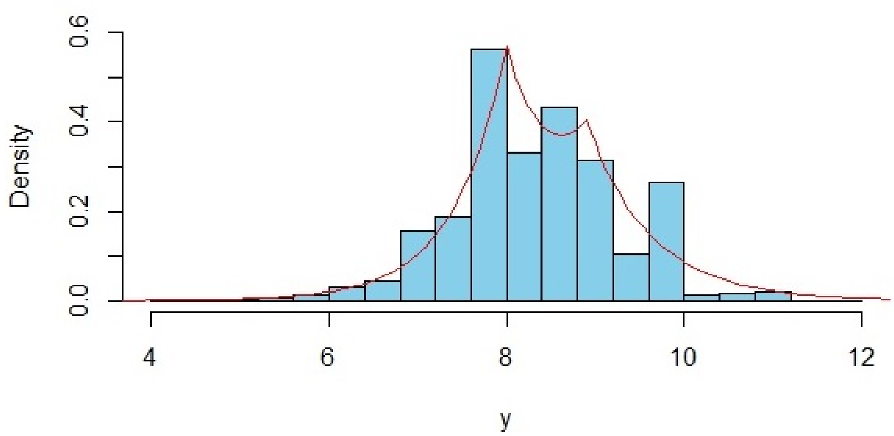

Figure 3: Plot of a Superimposed Laplace Mixture

	p	β_1	β_2	β_3	β_4	β_5	σ
Comp1	0.83658	8.31157	0.00377	-0.00401	0.176497	-0.18767	0.74234
Comp2	0.16342	8.86146	-0.011155	-0.00032	-0.06287	-0.06660	0.65893

Table 4: Laplace mixture regression model for thyroid data

4 Conclusion

As we mentioned earlier mostly hypocalcemia is modelled using Cox regression and logistic regression. In regression theory, when the response variable is non-Gaussian some transformation method such as Box-Cox method are recommended. However, when the response variable follows a mixture distribution, even the Box-Cox transformation may not lead to a reasonable result. We recommend the use of a two component Laplacian mixture regression model to predict the postoperative hypocalcemia level. We could show that our model performs better compared to the linear regression model and Gaussian mixture regression model based on AIC and BIC values.

REFERENCES

Andrews, D. F. and Mallows, C. L. (1974). Scale Mixtures of Normal Distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, **36 (1)**, 99-102.

Bai, X., Yao, W., and Boyer, J. E. (2012). Robust fitting of mixture regression models. *Computational Statistics and Data Analysis*, **56**, 2347-2359.

Benaglia, T., Chauveau, D., Hunter, D. R. & Young, D. S. (2010). mixtools: an R package for analyzing mixture models. *Journal of statistical software*, **32**, 1-29.

Bentrem, D.J., Rademaker, A., Angelos, P. (2001). Evaluation of serum calcium levels in predicting hypoparathyroidism after total/ near-total thyroidectomy or parathyroidectomy. *Am Surg* , **67**, 249-251.

Yang, C., Weijie, L. I. and Lemin, L. I. N. (2011). Capsular dissection technique in total thyroidectomy[J].*Chinese Journal of Clinical Medicine*, **18(5)**, 708-710.

Chisti, M. M., Nair, R.S., Kuttanchettiyar, K. G. and Yadev, I. (2017). Mechanisms behind Post-Thyroidectomy Hypocalcemia: Interplay of Calcitonin, Parathormone, and Albumin-A Prospective Study. *J Invest Surg*, **30**(4), 217-225.

Delbridge, L., Reeve, T. S., Khadra, M. & Poole, A. G. (1992). Total thyroidectomy: the technique of capsular dissection. *The Australian and New Zealand journal of surgery*, **62(2)**, 96–99.

Husein, M., Hier, M.P., Al-Abdulhadi, K., et al. (2002). Predicting calcium status post thyroidectomy with early calcium levels. *Otolaryngol Head Neck Surg*, **127**, 289-293.

Lindblom, P., Westerdahl, J. and Bergenfelz, A. (2002). Low parathyroid hormone levels after thyroid surgery: a feasible predictor of hypocalcemia. *Surgery*, **131**, 515-520.

McLachlan, G.J. and Krishnan, T. (2008). *The EM Algorithm and Extensions.* Wiley Series in Probability and Statistics, New Jersey.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*, Wiley Series in Probability and Statistics, New York.

Phillips, R. F. (2002). Least absolute deviations estimation via the EM algorithm. *Statistics and Computing*, **12**(3), 281-285.

Rio, P., Arcuri, M. F., Ferreri, G., et al. (2005). The utility of serum PTH assessment 24 hours after total thyroidectomy. *Otolaryngol Head Neck Surg.* **32**, 584-586.

Song, W., Yao, W. & Xing, Y. (2014). Robust mixture regression model fitting by Laplace distribution. *Computational Statistics & Data Analysis*,**71**, 128-137.

Tredici, P., Grosso, E., Gibelli, B., Massaro, M. A., Arrigoni, C. and Tradati, N. (2011). Identification of patients at high risk for hypocalcemia after total thyroidectomy. *Acta Otorhinolaryngol Ital*,**144**, 8.

Zakiah, I. K. and Faten, A. (2019). Two Component Laplace Mixture Model: Properties and Parametric Estimations. *Mathematics and Statistics*, **7(4A)**, 9-16.