Hacettepe Journal of Mathematics and Statistics
Volume 45 (6) (2016), 1805 – 1818

Analysis of covariance by assuming a skew normal distribution for response variable

Afshin Fallah*† and Zahra Goodarzi‡

Abstract

The traditional theory of analysis of covariance (ANCOVA) is based on normality assumption, while in many real world applications the data violate normality and this theory is not adequate. In this paper, we expand a model for analysis of covariance with a skew normal response variable. The maximum likelihood estimates of the model parameters are provided via an EM algorithm. We also developed asymptotic confidence intervals for parameters. A simulation study is performed to assess the performance of the proposed model. The methodology is illustrated using a real data set.

Keywords: Analysis of covariance, skew normal, maximum likelihood, EM algorithm.

2000 AMS Classification: 62H25, 62F12, 62P15

Received : 13.07.2015 *Accepted* : 25.10.2015 *Doi* : 10.15672/HJMS.20159314058

1. Introduction

Analysis of covariance is a widely used technique for exploring possible relation between a usually continuous response variable and a set of covariates and treatments. This methodology is a combination of regression and analysis of variance (ANOVA) that profits the benefits of both of these two efficient modeling methods. The ANCOVA can be employed for a wide range of different purposes. It can be used to filter out error variance, to explore pre-test vs. post-test effects, to control the variables, to finding significant difference between groups by reducing the within-groups variations etc. It also provides a useful approach to treat the potentially confounding variables. In many practical situations, one cannot provide the ideal homogenous experimental units for all treatments,

*Department of Statistics, Faculty of Basic Sciences, Imam Khomeini International University, Qazvin, Iran.
Email: a.fallah@sci.ikiu.ac.ir
†Corresponding Author.
‡Department of Statistics, Faculty of Basic Sciences, Imam Khomeini International University
Email: goodarzizahra67@yahoo.com

even after blocking, which is an essential requirement for comparative experiments analyzed by ANOVA. Thus one has to appeal to ANCOVA. In this context, adjusting treatment effects for nuisance covariates effects on the response variable is of paramount importance for the researcher. This practice allows finding the net effect of treatments under specified collection of covariates and provides a clear guidance for users of the results. This concept is foreign to proper regression analysis which does not discriminate between treatment and covariate. ANCOVA was firstly motivated by Fisher [14]. During the years many researchers have investigated different theoretical and applied aspects of ANCOVA in different sciences. Cochran [10] and Cox and McCullagh [11] and references therein are good sources for more information about ANCOVA. As it is pointed out by [19] the traditional theory of normal ANCOVA is not adequate when the data violate the normality assumption. This creates a strong motivation for considering ANCOVA under other distributions that are more flexible than normal distribution. Many researches have been recently focused to develop suitable methods for dealing with non-normality. These considerations are not limited to ANCOVA and other modeling techniques such as regression, ANOVA, discriminant analysis etc. have investigated repeatedly for use in situations that the normality assumption does not hold. In particular, the skew normal family of distributions as a generalization of the normal family has attracted considerable attentions in literature. Though the earlier appearance of skew normal distribution returns to Roberts [21] and O'Hagan and Leonard [20] and Aigner *et al.* [1], but the first formal definition of this family of distributions was provided by Azzalini [3]. The multivariate form of the skew normal distribution is expanded in [4] and [5]. During the three past decades many skew normal distribution have been introduced and discussed in literature. References [22, 15, 16, 4] are excellent sources of information about the skew normal family of distributions and their properties. Different modeling approaches such as regression analysis (Sahu *et al.* [22], Ferreira and Steel [13] and Cancho [9]), Bayesian nonlinear regression (De la Cruz and Branco [12]), linear mixed models (Arellano-Valle [2]) and analyzing longitudinal data (Baghfalaki *et al.* [7] and Lin and Lee [18]) have been developed under the assumption of skew normal distribution. This paper investigates ANCOVA under the assumption of skew normal distribution for response variable. We show that the skew normal ANCOVA model leads to the more efficient estimations of the model parameters than the traditional models.

The rest of paper is structured as follows. In section 2, we give some brief preliminaries and necessary background about the concept of ANCOVA and its formulation. The skew normal ANCOVA model is developed in section 3. We provide the ML estimates of the model parameters and their adjusted counterparts via EM algorithm. In section 4, we construct the asymptotic confidence intervals for the model parameters. A simulation study is performed to assess the performance of the proposed model, in section 5. In section 6, a real data set is analyzed to explain the proposed methodology.

2. Preliminaries and Notations

The aim of ANCOVA is to explore possible relation between a response variable and a set of treatments and covariates. Consider a balanced complete randomized design with t treatments and r replications. We treat the balanced design to avoid cluttered notations, but the problem can be cast in general unbalanced design, as it is explained by Meshkani *et. al.* [19]. In the simplest case, an ANCOVA model with a covariate and a two-level factor is given by

$$\boldsymbol{E}[Y_{ij}|\boldsymbol{x},\boldsymbol{z}] = \beta_0 + \beta_i + \gamma(z_{ij} - \bar{z}_i) \qquad i = 1,\dots,t, j = 1,\dots,r, \tag{2.1}$$

where β_0 shows the intercept term and $\beta_i, i = 1, \ldots, t$ denote the factor effects which satisfy the constraint $\sum_{i=1}^{t} \beta_i = 0$. The vector of model parameters is

$$\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')' = (\beta_0, \beta_1, \ldots, \beta_{t-1}, \gamma)'.$$

In model (2.1) the regression equation of Y on Z has a fixed slope γ for all treatments. If the slopes of the regression model for different treatments is not the same, the ANCOVA model would be of the form

$$\boldsymbol{E}[Y_{ij}|\boldsymbol{x}, \boldsymbol{z}] = \beta_0 + \beta_i + \gamma_i(z_{ij} - \bar{z}_i) \qquad i = 1, \ldots, t, j = 1, \ldots, r,$$

therefore there is a vector of slopes $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_t)'$. In general case, an ANCOVA model can be written as

$$\boldsymbol{E}(Y_{11}, \ldots, Y_{tr}) = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} = \boldsymbol{W\theta}, \tag{2.2}$$

where $\boldsymbol{X}$ denotes the design matrix, $\boldsymbol{Z}$ includes the observed covariates, $\boldsymbol{W} = [\boldsymbol{X}, \boldsymbol{Z}]$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ with $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{t-1})'$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$ and $p = t + q$. For example, in model (2.1) we have

$$\boldsymbol{W} = \left[\begin{array}{ccccc|c} \mathbf{1}_r & \mathbf{1}_r & \mathbf{0}_r & \ldots & \mathbf{0}_r & \tilde{\boldsymbol{z}}_1 \\ \mathbf{1}_r & \mathbf{0}_r & \mathbf{0}_r & \ldots & \mathbf{0}_r & \tilde{\boldsymbol{z}}_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{1}_r & \mathbf{0}_r & \mathbf{0}_r & \ldots & \mathbf{1}_r & \tilde{\boldsymbol{z}}_{t-1} \\ \mathbf{1}_r & -\mathbf{1}_r & -\mathbf{1}_r & \ldots & -\mathbf{1}_r & \tilde{\boldsymbol{z}}_t \end{array}\right] = [\boldsymbol{X}|\boldsymbol{Z}] \tag{2.3}$$

where $\mathbf{1}_r = (1, \ldots, 1)'$, $\mathbf{0}_r = (0, \ldots, 0)'$ and $\tilde{\boldsymbol{z}}_i = ((z_{i1} - \bar{z}_i), \ldots, (z_{ij} - \bar{z}_i), \ldots, (z_{ir} - \bar{z}_i))$. As it can be clearly seen, in an ANCOVA model the relationship between the mean of a response variable and treatments and covariates is determined by the structure of design matrix $\boldsymbol{X}$ and covariate matrix $\boldsymbol{Z}$. For model (2.2) the design matrix, $\boldsymbol{X}$, and the vector of treatments effects, $\boldsymbol{\beta}$, are the same as model (2.1), but the matrix of observed covariates is given by

$$\boldsymbol{Z} = \left[\begin{array}{ccccc} \tilde{\boldsymbol{z}}_1 & \mathbf{0}_r & \ldots & \mathbf{0}_r & \mathbf{0}_r \\ \mathbf{0}_r & \tilde{\boldsymbol{z}}_2 & \ldots & \mathbf{0}_r & \mathbf{0}_r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_r & \mathbf{0}_r & \ldots & \tilde{\boldsymbol{z}}_{t-1} & \mathbf{0}_r \\ \mathbf{0}_r & \mathbf{0}_r & \ldots & \mathbf{0}_r & \tilde{\boldsymbol{z}}_t \end{array}\right].$$

The unbalanced form of ANCOVA models can also be represented by the general form given in equation (2.2). Considering the constraint $\sum_{i=1}^{t} r_i\beta_i = 0$, it would suffice to replace the $-\frac{1}{r_t}(r_1, \ldots, r_{t-1})$ for $-\mathbf{1}_r$ in the last row of the matrix $\boldsymbol{W}$ where $r_i, , i = 1, \ldots, t$, denotes the number of replications for i-th treatment. For Other common designs such as split-plot, Latin squares, Greco-Latin etc., the modeling method is similar, i.e., the design matrix and the covariate matrix can be written in the general form of $\boldsymbol{W} = [\boldsymbol{X}, \boldsymbol{Z}]$. It should be noted that the constraint $\sum_{i=1}^{t} \beta_i = 0$ has been absorbed into the design matrix $\boldsymbol{W}$. More details and examples about other common designs can be found in [19]

Considering the general formulation of an ANCOVA model given in equation (2.2), the main goal of ANOCVA is to estimate the vector of parameters $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ using the vector of responses $\boldsymbol{y} = (y_{11}, \ldots, y_{tr})$ and the matrix of observations $\boldsymbol{W}$. In what follows we follow the notations of [19].

3. The model and parameter estimation

In this section, we develop an ANCOVA model under the assumption of skew normal distribution for response variable. Consider the general form of an ANCOVA model given in equation (2.2). Let the skew normal ANCOVA model be

$$Y_{ij}|\boldsymbol{w}_{ij} \sim SSN(\boldsymbol{w}_{ij}\boldsymbol{\theta} - \sqrt{\frac{2}{\pi}}\lambda, \sigma^2, \lambda) \quad i = 1, ..., s;\ j = 1, ..., r, \tag{3.1}$$

where s is the number of treatments, r is the number of replications, $\boldsymbol{w}_{ij} = (\boldsymbol{x}_i, z_{ij})$ denotes the ij-th row of matrix $\boldsymbol{W}$ and $SSN(\mu, \sigma^2, \lambda)$ denotes the Sahu skew normal distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, given by

$$f_{Y_{ij}}(y|\mu, \sigma^2, \lambda) = 2\phi(y; \mu - \sqrt{\frac{2}{\pi}}\lambda, \sigma^2 + \lambda^2)\Phi(\frac{\lambda}{\sigma}\frac{(y-\mu)}{(\sigma^2+\lambda^2)^{\frac{1}{2}}}), \tag{3.2}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the density and cumulative distribution function of the normal distribution. The likelihood function of the model (3.1) is

$$\begin{aligned} L(\boldsymbol{\theta}, \lambda, \sigma^2|\boldsymbol{y}, \boldsymbol{W}) &= \prod_{i=1}^{s}\prod_{j=1}^{r} f_{Y_{ij}}(y_{ij}|\boldsymbol{\theta}, \lambda, \sigma^2) \\ &= \prod_{i=1}^{s}\prod_{j=1}^{r} 2\phi(y_{ij}; \boldsymbol{w}_{ij}\boldsymbol{\theta} - \sqrt{\frac{2}{\pi}}\lambda, \sigma^2 + \lambda^2)\Phi(\frac{\lambda}{\sigma}\frac{(y_{ij} - \boldsymbol{w}_{ij}\boldsymbol{\theta})}{(\sigma^2+\lambda^2)^{\frac{1}{2}}}). \end{aligned} \tag{3.3}$$

Due to the complexity of likelihood function (3.3) there are no explicit form for the ML estimators of the model parameters. Therefore, we provide an EM algorithm to compute the numerical values of the ML estimates. For this, it is necessary to formulate the problem in terms of a missing data problem. The skew normal ANCOVA model (3.1) can be written in a hierarchical structure as a mixture of normal and halfnormal distributions given by

$$\begin{cases} Y_{ij}|T_{ij} = t_{ij} \sim N(\boldsymbol{w}_{ij}\boldsymbol{\theta} + \lambda(t_{ij} - \sqrt{\frac{2}{\pi}}), \sigma^2) \\ T_{ij} \sim HN(0,1) \qquad\qquad\qquad\qquad i = 1, ..., s;\ j = 1, ..., r. \end{cases}$$

Therefore, considering $\{T_{ij}; i = 1, \ldots, s; j = 1, \ldots, r\}$ and $\{y_{ij}; i = 1, .., s; j = 1, ..., r\}$, respectively, as missing and incomplete data, the joint density of the complete data (y_{ij}, T_{ij}) is given by

$$\begin{aligned} f_{(Y_{ij},T_{ij})}(y_{ij}, t_{ij}) &= f_{Y_{ij}|T_{ij}=t_{ij}}(y_{ij}) \times g_{T_{ij}}(t_{ij}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(y_{ij} - \boldsymbol{w}_{ij}\boldsymbol{\theta} - \lambda(t_{ij} - \sqrt{\frac{2}{\pi}}))^2 \right\} \\ &\times \sqrt{\frac{2}{\pi}} \exp\left\{ -\frac{t_{ij}^2}{2} \right\}. \end{aligned}$$

Hence, the complete data likelihood and log-likelihood functions are obtained to be

$$\begin{aligned} L_c(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W}, \boldsymbol{t}) &= \prod_{i=1}^{s}\prod_{j=1}^{r} f_{(Y_{ij},T_{ij})}(y_{ij}, t_{ij}) \\ &= \exp\left\{ -\frac{1}{2\sigma^2}\sum_{i=1}^{s}\sum_{j=1}^{r}\left[(y_{ij} - \boldsymbol{w}_{ij}\boldsymbol{\theta})^2 - 2\lambda(y_{ij} - \boldsymbol{w}_{ij}\boldsymbol{\theta} \right.\right. \\ &+ \left.\left. \lambda\sqrt{\frac{2}{\pi}})t_{ij} + (\lambda^2 + \sigma^2)t_{ij}^2 + 2\lambda\sqrt{\frac{2}{\pi}}(y_{ij} - \boldsymbol{w}_{ij}\boldsymbol{\theta}) + \lambda^2\frac{2}{\pi}\right]\right\} \end{aligned}$$

$$\times \quad (\pi)^{-sr}(\sigma^2)^{-\frac{sr}{2}}$$

and

$$\begin{aligned}
\ell_c(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W},\boldsymbol{t}) &= -\frac{1}{2\sigma^2}\Big[(\boldsymbol{y}-\boldsymbol{W\theta})'(\boldsymbol{y}-\boldsymbol{W\theta}) - 2\lambda(\boldsymbol{y}-\boldsymbol{W\theta}+\lambda\sqrt{\frac{2}{\pi}}\mathbf{1})'\boldsymbol{t} \\
&+ (\lambda^2+\sigma^2)(\boldsymbol{t}^2)'\mathbf{1} + 2\lambda\sqrt{\frac{2}{\pi}}(\boldsymbol{Y}-\boldsymbol{W\theta})'\mathbf{1} + \frac{2sr\lambda^2}{\pi}\Big] \\
&- sr\log\pi - \frac{sr}{2}\log\sigma^2,
\end{aligned}$$

respectively, where $\boldsymbol{t} = (t_{11},...,t_{sr})'$, $\boldsymbol{t}^2 = (t_{11}^2,...,t_{sr}^2)$ and $\mathbf{1}_{sr}$ denotes a $sr\times 1$ unit vector. To proceeds the EM algorithm, the conditional expectation of the complete data log-likelihood given incomplete data is obtained to be

$$\begin{aligned}
E(\ell_c(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})) &= -sr\log\pi - \frac{sr}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\Big[(\boldsymbol{y}-\boldsymbol{W\theta})'(\boldsymbol{y}-\boldsymbol{W\theta}) \\
&- 2\lambda(\boldsymbol{y}-\boldsymbol{W\theta}+\lambda\sqrt{\frac{2}{\pi}}\mathbf{1}_{sr})'\hat{\boldsymbol{t}} + (\lambda^2+\sigma^2)(\hat{\boldsymbol{t}^2})'\mathbf{1}_{sr} \\
&+ 2\lambda\sqrt{\frac{2}{\pi}}(\boldsymbol{Y}-\boldsymbol{W\theta})'\mathbf{1}_{sr} + \frac{2sr\lambda^2}{\pi}\Big], \qquad (3.4)
\end{aligned}$$

where $\hat{t}$ and $\hat{t^2}$ denote the first and second order conditional moments of random variable $T_{ij}|y_{ij}$, respectively. Using the equations of the truncated normal moments (see for example, Barr *et al.* 1999), these moments are given by

$$\begin{aligned}
\hat{t}_{ij} &= E(t_{ij}|\hat{\boldsymbol{\theta}},y_{ij}) = \eta_{ij}+\tau\delta_{ij}, \\
\hat{t^2}_{ij} &= E(t_{ij}^2|\hat{\boldsymbol{\theta}},y_{ij}) = \eta_{ij}^2+\tau^2+\tau\delta_{ij}\eta_{ij},
\end{aligned}$$

where $\eta_{ij} = \frac{\lambda}{\sigma^2+\lambda^2}(y_{ij}-\boldsymbol{w}_{ij}\boldsymbol{\theta}+\sqrt{\frac{2}{\pi}}\lambda)$, $\tau^2 = \frac{\sigma^2}{\sigma^2+\lambda^2}$ and $\delta_{ij} = \frac{\phi(\frac{\hat{\eta}_{ij}}{\hat{\tau}})}{\Phi(\frac{\hat{\eta}_{ij}}{\hat{\tau}})}$. The M-step of EM algorithm searches the parameter space to maximize the conditional expectation (3.4). Given the values of the parameters in k-th iteration of algorithm, the ML estimates of the parameters in (k+1)-th iteration are obtained as,

$$\begin{aligned}
\hat{\boldsymbol{\theta}}^{(k+1)} &= (\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\Big(\boldsymbol{y}+\hat{\lambda}^{(k)}(\sqrt{\frac{2}{\pi}}\mathbf{1}_{sr}-\hat{\boldsymbol{t}}^{(k)})\Big), \\
\hat{\sigma^2}^{(k+1)} &= \frac{1}{sr}\Big[(\boldsymbol{y}-\boldsymbol{W}\hat{\boldsymbol{\theta}}^{(k)})'(\boldsymbol{y}-\boldsymbol{W}\hat{\boldsymbol{\theta}}^{(k)}) \\
&- 2\hat{\lambda}^{(k)}(\boldsymbol{y}-\boldsymbol{W}\hat{\boldsymbol{\theta}}^{(k)}+\hat{\lambda}^{(k)}\sqrt{\frac{2}{\pi}}\mathbf{1}_{sr})'\hat{\boldsymbol{t}}^{(k)} + \hat{\lambda^2}^{(k)}(\boldsymbol{t}^2)'\mathbf{1} \\
&+ 2\lambda\sqrt{\frac{2}{\pi}}(\boldsymbol{Y}-\boldsymbol{W\theta})'\mathbf{1} + \frac{2sr\lambda^2}{\pi}\Big], \\
\hat{\lambda}^{(k+1)} &= \Big(\sqrt{\frac{2}{\pi}}\mathbf{1}_{sr}'(\boldsymbol{y}-\boldsymbol{W}\hat{\boldsymbol{\theta}}^{(k)}) - (\boldsymbol{y}-\boldsymbol{W}\hat{\boldsymbol{\theta}}^{(k)})'\hat{\boldsymbol{t}}^{(k)}\Big) \\
&\times \Big(\mathbf{1}_{sr}'(2\sqrt{\frac{2}{\pi}}\hat{\boldsymbol{t}}^{(k)} - \hat{\boldsymbol{t}^2}^{(k)}) - sr\frac{2}{\pi}\Big)^{-1}.
\end{aligned}$$

The E and M steps are repeated alternately until a convergence rule holds.

3.1. Adjusted effects. As it can be clearly seen, from equation (2.2), there are two types of parameters in an ANCOVA model. The first type, denoted by $\boldsymbol{\beta}$, corresponds to the treatment effects. Whereas the second type, denoted by $\boldsymbol{\gamma}$, corresponds to covariate effects. The EM algorithm expounded in previous section, provides the ML estimator for the vector of parameters, $\boldsymbol{\theta}$, without separating these two types of parameters. We may be interested in estimating either the effects of treatments adjusted for the effects of covariates or the effects of covariates adjusted for the effects of treatments. In this section, we provide the adjusted estimators for both covariates and treatments effects. For this purpose, we rewrite the equation (3.4) by using the equality of $\boldsymbol{w}_{ij}\boldsymbol{\theta} = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{\gamma}$, as:

$$\begin{aligned}
E(\ell_c(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{\beta}, \boldsymbol{\gamma})) \quad &\propto \quad \sum_{i=1}^{s}\sum_{j=1}^{r}\Big(y_{ij} - (\boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{\gamma})\Big)^2 \\
&- \quad 2\lambda\sum_{i=1}^{s}\sum_{j=1}^{r}\Big(y_{ij} - \boldsymbol{x}_i\boldsymbol{\beta} - \boldsymbol{z}_{ij}\boldsymbol{\gamma}\Big)\hat{t}_{ij} - 2\lambda^2\sqrt{\frac{2}{\pi}}\sum_{i=1}^{s}\sum_{j=1}^{r}\hat{t}_{ij}, \\
&+ \quad \lambda^2\sum_{i=1}^{s}\sum_{j=1}^{r}\hat{t^2}_{ij} + 2\lambda\sqrt{\frac{2}{\pi}}\sum_{i=1}^{s}\sum_{j=1}^{r}\Big(y_{ij} - \boldsymbol{x}_i\boldsymbol{\beta} - \boldsymbol{z}_{ij}\boldsymbol{\gamma}\Big) + sr\lambda^2\frac{2}{\pi} \\
&\propto \quad \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'(\boldsymbol{X\beta} + \boldsymbol{Z\gamma}) + (\boldsymbol{X\beta})'(\boldsymbol{X\beta}) + 2(\boldsymbol{X\beta})'(\boldsymbol{Z\gamma}) \\
&- \quad 2\lambda^2\sqrt{\frac{2}{\pi}}\boldsymbol{1}'_{sr}\hat{\boldsymbol{t}} + \lambda^2\boldsymbol{1}'_{sr}\widehat{\boldsymbol{t}^2} + 2\lambda\sqrt{\frac{2}{\pi}}\boldsymbol{1}'_{sr}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Z\gamma}) \\
&+ \quad (\boldsymbol{Z\gamma})'(\boldsymbol{Z\gamma}) + sr\lambda^2\frac{2}{\pi}. \qquad (3.5)
\end{aligned}$$

Equating the first order derivations of (3.5) with respect to the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to zero, leads to the following system of equations:

$$\begin{cases}
\dfrac{\partial(E(\ell_c(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{\beta}, \boldsymbol{\gamma}))}{\partial\boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X\beta} + 2\boldsymbol{X}'\boldsymbol{Z\gamma} - 2\lambda\sqrt{\tfrac{2}{\pi}}\boldsymbol{X}'\boldsymbol{1}_{sr} = \boldsymbol{0} \\
\dfrac{\partial(E(\ell_c(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{\beta}, \boldsymbol{\gamma}))}{\partial\boldsymbol{\gamma}} = -2\boldsymbol{Z}'\boldsymbol{y} + 2\boldsymbol{Z}'\boldsymbol{X\beta} + 2\boldsymbol{Z}'\boldsymbol{Z\gamma} - 2\lambda\sqrt{\tfrac{2}{\pi}}\boldsymbol{Z}'\boldsymbol{1}_{sr} = \boldsymbol{0}.
\end{cases}$$

Therefore, the adjusted ML estimators of treatments and covariates effects in $k-th$ iteration of the EM algorithm are obtained to be:

$$\hat{\boldsymbol{\beta}}^{(k+1)}_{ML.z} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\gamma}^{(k)}_{ML} + \lambda^{(k)}\sqrt{\frac{2}{\pi}}\boldsymbol{1}_{sr}),$$

$$\hat{\boldsymbol{\gamma}}^{(k+1)}_{ML.x} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(\hat{k})}_{ML} + \lambda^{(k)}\sqrt{\frac{2}{\pi}}\boldsymbol{1}_{sr}).$$

Note that, as it is well known, in the proper regression analysis each regression coefficient shows the effect of corresponding explanatory variable on the response variable, given all other explanatory variables (qualitative and quantitative) are kept fixed. But in ANCOVA, one needs the treatment effects for the situation that only the whole set of quantitative variables, i.e., covariates are kept fixed. Moreover, the partition used in ANCOVA is dictated by the context of each special experiment. For example, some experiments may have no covariate thus no partition is considered and some may have one or more covariates whose effects should be removed from the treatment effects. Thus, there is a natural partition of treatments and covariates correspond to each problem.

4. Asymptotic Confidence Intervals

To construct exact confidence intervals for the model parameters requires exact knowledge of the sampling distribution of the ML estimators. Due to the complexity of these estimators, derivation of their exact distributions is a challenging problem, if it be feasible at all. Therefore, in this section we use the asymptotic distributions of these estimators to construct asymptotic confidence intervals for the model parameters. The results of this section are valid when r or equivalently $n(=tr)$ goes to infinity.

Consider the skew normal ANCOVA model (3.1). Let

$$\ell_{ij}(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W}) = \log 2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log a - \frac{b_{ij}}{2} + \log\Phi(k_{ij}),$$

with $a = \sigma^2 + \lambda^2$ and

$$\begin{aligned} b_{ij} &= \frac{\left(y_{ij} - \boldsymbol{w}_{ij}\boldsymbol{\theta} + \sqrt{\frac{2}{\pi}}\lambda\right)^2}{\sigma^2 + \lambda^2}, \\ k_{ij} &= \frac{\lambda}{\sigma^{\frac{1}{2}}}\left(y_{ij} - \boldsymbol{w}_{ij}\boldsymbol{\theta} + \sqrt{\frac{2}{\pi}}\lambda\right). \end{aligned}$$

Then, the log-likelihood function of the model is given by

$$\ell(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W}) = \sum_{i=1}^{s}\sum_{j=1}^{r} \ell_{ij}(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W}).$$

The first order partial derivations of the log-likelihood function with respect to the model parameters $\boldsymbol{\theta}$, σ^2 and λ are given by

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W})}{\partial \boldsymbol{\theta}} &= \sum_{i=1}^{s}\sum_{j=1}^{r} \frac{\partial \ell_{ij}(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W})}{\partial \boldsymbol{\theta}}, \\ \frac{\partial \ell(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W})}{\partial \sigma^2} &= \sum_{i=1}^{s}\sum_{j=1}^{r} \frac{\partial \ell_{ij}(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W})}{\partial \sigma^2}, \\ \frac{\partial \ell(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W})}{\partial \lambda} &= \sum_{i=1}^{s}\sum_{j=1}^{r} \frac{\partial \ell_{ij}(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W})}{\partial \lambda}, \end{aligned}$$

respectively, where

$$\begin{aligned} \frac{\partial \ell_{ij}(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W})}{\partial \boldsymbol{\theta}} &= \frac{y_{ij} - \boldsymbol{w}_{ij}\boldsymbol{\theta} + \sqrt{\frac{2}{\pi}}}{a}\boldsymbol{w}'_{ij} - \delta_{\Phi(k_{ij})}\frac{\lambda}{\sigma\sqrt{a}}\boldsymbol{w}'_{ij}, \\ \frac{\partial \ell_{ij}(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W})}{\partial \sigma^2} &= -\frac{1}{2a} + \frac{b_{ij}}{2a} - \delta_{\Phi(k_{ij})}\frac{k_{ij}(2\sigma^2 + \lambda^2)}{\sigma^2 a}, \\ \frac{\partial \ell_{ij}(\boldsymbol{\theta}, \sigma^2, \lambda|\boldsymbol{y}, \boldsymbol{W})}{\partial \lambda} &= -\frac{\lambda}{a} - \frac{1}{a^2}\{a\sqrt{\frac{2}{\pi}}(y_{ij} - \boldsymbol{w}_{ij}\boldsymbol{\theta} + \sqrt{\frac{2}{\pi}}\lambda) - \lambda b_{ij}\} \\ &+ \delta_{\Phi(k_{ij})}k_{ij}(1 - \frac{\lambda^2}{a}) + \frac{\lambda}{\sigma\sqrt{a}}\sqrt{\frac{2}{\pi}}, \end{aligned}$$

and $\delta_{\Phi(u)} = \dfrac{\phi(u)}{\Phi(u)}$. The second order derivations of the log-likelihood function with respect to the parameters are similarly given by

$$\frac{\partial^2 \ell_{ij}}{\partial \boldsymbol{\nu}\partial \boldsymbol{\xi}'} = -\frac{1}{2}\frac{\partial^2 \log a}{\partial \boldsymbol{\nu}\partial \boldsymbol{\xi}'} - \frac{1}{2}\frac{\partial^2 b_{ij}}{\partial \boldsymbol{\nu}\partial \boldsymbol{\xi}'} + \frac{\partial^2 \log \Phi(k_{ij})}{\partial \boldsymbol{\nu}\partial \boldsymbol{\xi}'},$$

where $\boldsymbol{\nu}$ represents the parameters $\boldsymbol{\theta}, \sigma^2$ or λ and

$$\begin{aligned} \frac{\partial^2 \log \Phi(k_{ij})}{\partial \boldsymbol{\nu}\partial \boldsymbol{\xi}'} &= \delta_{\Phi}(k_{ij})\Big(\frac{\partial^2 k_{ij}}{\partial \boldsymbol{\nu}\partial \boldsymbol{\xi}'}\Big) + \Delta_{\Phi}(k_{ij})\Big(\frac{\partial k_{ij}}{\partial \boldsymbol{\nu}}\Big)\Big(\frac{\partial k_{ij}}{\partial \boldsymbol{\xi}}\Big)' \\ \frac{\partial^2 \log a}{\partial \lambda^2} &= 2\frac{(a - 2\lambda^2)}{a^2}, \quad \frac{\partial^2 b_{ij}}{\partial \lambda^2} = \frac{4}{\pi a} - 8\frac{\sigma k_{ij}}{\lambda a^{\frac{3}{2}}} + 2\frac{b_{ij}}{a^2}(4\lambda - a), \\ \frac{\partial^2 b_{ij}}{\partial \lambda \partial \sigma^2} &= 2\sqrt{\frac{2}{\pi}}\frac{\sigma k_{ij}}{\lambda a^{\frac{3}{2}}} + 4\frac{\lambda}{a^2}b_{ij}, \quad \frac{\partial^2 b_{ij}}{\partial {\sigma^2}^2} = -2\frac{b_{ij}}{a^2}, \\ \frac{\partial^2 b_{ij}}{\partial \lambda \partial \boldsymbol{\theta}} &= 2(\frac{k_{ij}}{a^{\frac{3}{2}}} - \sqrt{\frac{2}{\pi}}\frac{1}{a})\boldsymbol{w}'_{ij}, \quad \frac{\partial^2 b_{ij}}{\partial \sigma^2 \partial \boldsymbol{\theta}} = \frac{\sigma k_{ij}}{\lambda a^{\frac{3}{2}}}\boldsymbol{w}'_{ij}, \\ \frac{\partial^2 b_{ij}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} &= \frac{2}{a}\boldsymbol{w}'_{ij}(\boldsymbol{w}_{ij}), \quad \frac{\partial^2 k_{ij}}{\partial \lambda \partial \boldsymbol{\theta}} = -\frac{\sigma}{a^{\frac{3}{2}}}\boldsymbol{w}'_{ij}, \\ \frac{\partial^2 k_{ij}}{\partial \lambda^2} &= 2\sqrt{\frac{2}{\pi}}\frac{\sigma}{a^{\frac{3}{2}}} - 3\frac{\sigma^2}{a^2}k_{ij} \end{aligned}$$

$$\frac{\partial^2 k_{ij}}{\partial\lambda\partial\sigma^2} = \frac{(2\sigma^2+\lambda^2)}{2\sigma a^{\frac{3}{2}}}(\frac{k_{ij}}{\lambda a^{\frac{1}{2}}}-\sqrt{\frac{2}{\pi}}\lambda)+\frac{\lambda k_{ij}}{a^2}(\frac{3}{2}-\frac{a}{\sigma^2}), \quad \frac{\partial^2 k_{ij}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}=0.$$

with $\Delta_\phi(u)=\delta_{\Phi(u)}(u+\delta_{\Phi(u)})$. Therefore the Hessian matrix of model is obtained as:

$$\boldsymbol{H}(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})=\begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix},$$

where

$$\begin{aligned}
h_{11} &= \frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}, \quad h_{22}=\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\lambda^2}, \quad h_{33}=\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial(\sigma^2)^2},\\
h_{12} &= h_{21}=\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\boldsymbol{\theta}\partial\lambda}, \quad h_{13}=h_{31}=\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\boldsymbol{\theta}\partial\sigma^2},\\
h_{23} &= h_{32}=\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\sigma^2\partial\lambda}.
\end{aligned}$$

Consequently, the Fisher information matrix of model is given by

$$\boldsymbol{I}(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})=-H(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})=\begin{pmatrix} \boldsymbol{I}(\boldsymbol{\theta}) & \boldsymbol{I}(\boldsymbol{\theta}.\lambda) & \boldsymbol{I}(\sigma^2,\boldsymbol{\theta}) \\ \boldsymbol{I}(\boldsymbol{\theta},\lambda) & \boldsymbol{I}(\lambda) & \boldsymbol{I}(\sigma^2,\lambda) \\ \boldsymbol{I}(\sigma^2,\boldsymbol{\theta}) & \boldsymbol{I}(\sigma^2,\lambda) & \boldsymbol{I}(\sigma^2) \end{pmatrix},$$

where

$$\begin{aligned}
\boldsymbol{I}(\boldsymbol{\theta}) &= -\boldsymbol{E}(\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'})\\
&= -\frac{1}{a}\boldsymbol{W}'\boldsymbol{W}+\boldsymbol{E}\left(\frac{\phi(k_{ij})}{\Phi(\boldsymbol{k})}(k_{ij}+\frac{\phi(k_{ij})}{\Phi(k_{ij})})\right)-\frac{\lambda}{\sigma\sigma^2 a}\boldsymbol{W}'\boldsymbol{W},\\
\boldsymbol{I}(\lambda) &= -\boldsymbol{E}(\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\lambda^2})=-\frac{a-2\lambda^2}{a^2}\\
&- \left(\frac{2}{\pi a}-\frac{4}{a^2}\sqrt{\frac{2}{\pi}}\lambda+\frac{(4-a)}{a^2}\frac{1}{\sigma^2+\lambda^2}(\sigma^2+\frac{2\lambda^2}{\pi})\right)\\
&+ \boldsymbol{E}\left(\frac{\phi(k_{ij})}{\Phi(k_{ij})}(2\sqrt{\frac{2}{\pi}}\frac{\sigma}{a^{\frac{3}{2}}})-3\frac{\sigma^2}{a^2}k_{ij}\frac{\phi(k_{ij})}{\Phi(k_{ij})}(k_{ij}+\frac{\phi(k_{ij})}{\Phi(k_{ij})})k_{ij}\left(1-\frac{\lambda^2}{a}\right)\right.\\
&+ \left.\frac{\lambda}{\sigma a^{\frac{1}{2}}}\sqrt{\frac{2}{\pi}}\right),\\
\boldsymbol{I}(\sigma^2) &= -\boldsymbol{E}(\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\sigma^2}),\\
\boldsymbol{I}(\sigma^2,\lambda) &= -\boldsymbol{E}(\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\sigma^2\partial\lambda}),\\
\boldsymbol{I}(\sigma^2,\boldsymbol{\theta}) &= -\boldsymbol{E}(\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\sigma^2\partial\boldsymbol{\theta}}),\\
\boldsymbol{I}(\boldsymbol{\theta},\lambda) &= -\boldsymbol{E}(\frac{\partial^2\ell(\boldsymbol{\theta},\sigma^2,\lambda|\boldsymbol{y},\boldsymbol{W})}{\partial\boldsymbol{\theta}\partial\lambda}).
\end{aligned}$$

Now, one can use the inverse of expected Fisher information matrix to approximate the variance of ML estimators. Thus, the asymptotic distributions of ML estimators and asymptotic confidence intervals for the model parameters are given by

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{ML} &\sim AN(\boldsymbol{\theta},\boldsymbol{I}^{-1}(\boldsymbol{\theta})),\\
\hat{\sigma^2}_{ML} &\sim AN(\sigma^2,\boldsymbol{I}^{-1}(\sigma^2)),\\
\hat{\lambda}_{ML} &\sim AN(\lambda,\boldsymbol{I}^{-1}(\lambda)),
\end{aligned}$$

and

$$(\hat{\boldsymbol{\theta}}_{ML}-Z_{1-\frac{\alpha}{2}}\sqrt{\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}}_{ML})} \quad , \quad \hat{\boldsymbol{\theta}}_{ML}+Z_{1-\frac{\alpha}{2}}\sqrt{\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}}_{ML})}),$$

$$(\hat{\sigma^2}_{ML} - Z_{1-\frac{\alpha}{2}}\sqrt{\boldsymbol{I}^{-1}(\hat{\sigma^2}_{ML})} \quad , \quad \hat{\sigma^2}_{ML} + Z_{1-\frac{\alpha}{2}}\sqrt{\boldsymbol{I}^{-1}(\hat{\sigma^2}_{ML})}),$$

$$(\hat{\lambda}_{ML} - Z_{1-\frac{\alpha}{2}}\sqrt{\boldsymbol{I}^{-1}(\hat{\lambda}_{ML})} \quad , \quad \hat{\sigma^2}_{ML} + Z_{1-\frac{\alpha}{2}}\sqrt{\boldsymbol{I}^{-1}(\hat{\lambda}_{ML})}), \tag{4.1}$$

respectively. Notice that in (4.1) we substituted the ML estimates of parameters in Fisher information matrix to estimate it. According to the large sample theory results, if $\boldsymbol{I}(\xi)$ is a continuous function of ξ, as it is typically the case, then $\boldsymbol{I}(\hat{\theta}_{ML})$ is a consistent estimator of $\boldsymbol{I}(\theta_{ML})$. See, for example, Lehmann [17], p.p. 525, for more details.

5. Simulation Study

In this section, we perform a simulation study to assess the efficiency of the ML estimators of the model parameters for the proposed model. We consider an ANCOVA model with a covariate and a two-level treatment of the form

$$\mu_{ij} = E(y_{ij}|\boldsymbol{X}, \boldsymbol{Z}) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_i + \boldsymbol{\gamma}(\boldsymbol{z}_{ij} - \bar{\boldsymbol{z}}_i), \tag{5.1}$$

with $i = 1, 2$ and $j = 1, \ldots, r$. The values of the model parameters are set to be $\beta_0 = 2, \beta_1 = 5$ and $\gamma = 1$. The covariate values are simulated from normal distribution. Then, the values of $\mu_{ij}, i = 1, 2, j = 1, \ldots, r$ are computed using the equality of $\mu_{ij} = \boldsymbol{w}_{ij}\boldsymbol{\theta}$. Finally, the response variable observations $\{y_{ij}, i = 1, 2; j = 1, ..., r\}$ are simulated from $SSN(\boldsymbol{w}_{ij}\boldsymbol{\theta} - \sqrt{\frac{2}{\pi}}, \sigma^2, \lambda)$. In order to evaluate the effect of sample size, $n = sr$, on efficiency of the ML estimators, we consider the number of replications, r, to be $\{10, 25, 50, 100\}$.

Also, to assess the ability of the proposed model for modeling observations with both symmetric and asymmetric structures, we consider different values for the skewness parameter as $\{-2, -1, 0, 1, 2\}$. Taking these considerations into account, the values of the root mean square error (RMSE) for the ML estimators of the model parameters are computed and presented in Table 1. We also provide the corresponding values of the ML estimators under normal distribution as the usual traditional ANCOVA model in order to compare and evaluate the robustness of different models against violation from normality. The number of repetitions in simulations fixed to be 5000 in order to take into account the uncertainty in random number generating procedure. As it is expected, for $\lambda = 0$, there are no significant differences between the values of RMSE for normal and skew normal ANCOVA models. This is because for $\lambda = 0$ the skew normal distribution reduces to normal distribution. For positive and negative values of the skewness parameter, which respectively correspond to the right-skewed and left-skewed data, the skew normal model provides more efficiency (in terms of smaller RMSE) than the normal model because it truly takes into account the skewed structure of data. Obviously, due to the asymptotic optimality of ML estimators, the efficiency of estimators for both normal and skew normal models increase when the sample size increases.

6. Real Example

To illustrate the proposed methodology and to evaluate its applicability, we provide a real example in this section. Table 2 shows the salary data for 58 employees in a company in Iran by the level of proficiency and working experience. Our aim is to find the possible relation between salary as the response variable and working experience as a covariate for different levels of the proficiency factor. Therefore, we consider an ANCOVA model with a covariate and a two-level treatment as

$$\mu_{ij} = E(y_{ij}|\boldsymbol{X}, \boldsymbol{Z}) = \beta_0 + \beta_i + \gamma(\boldsymbol{z}_{ij} - \bar{\boldsymbol{z}}_i),$$

Table 1. The values of RMSE for ML estimators of the model parameters.

		ANCOVA Model						
		Skew Normal				Normal		
λ	Sample Size	λ	β_0	β_1	γ	β_0	β_1	γ
-2	20	4.5209	0.9088	4.7756	27.5127	1.0760	5.0431	27.6104
	50	4.6625	0.9106	4.6961	22.2627	1.0288	5.0194	22.9811
	100	4.4520	0.9126	4.7334	12.2603	1.0186	5.0084	12.6771
	200	3.9090	0.9206	4.8485	10.1249	1.0077	5.0029	9.6656
-1	20	3.5050	0.6819	4.7952	18.3095	1.0264	5.0178	18.3095
	50	1.1328	0.6630	4.9245	10.9095	1.0105	5.0063	11.4319
	100	3.7464	0.6995	4.7370	8.2709	1.0068	5.0068	8.6453
	200	3.4097	0.7170	4.7982	5.7749	1.0017	5.0040	6.1442
0	20	$< 1 \times 10^{-13}$	1.0098	5.0072	10.5319	1.0098	5.0072	10.5319
	50	$< 1 \times 10^{-13}$	1.0055	5.0018	7.5567	1.0055	5.0018	7.5567
	100	$< 1 \times 10^{-13}$	1.0036	5.0013	5.4093	1.0036	5.0013	5.4093
	200	$< 1 \times 10^{-13}$	1.0019	4.9981	3.6826	1.0019	4.9981	3.6826
1	20	1.4593	0.6506	4.8993	18.1243	1.0292	5.0137	19.0754
	50	1.1328	0.6630	4.9245	10.9095	1.0105	5.0063	11.4319
	100	0.7788	0.6711	4.9595	7.4625	1.0067	5.0032	7.8143
	200	0.9599	0.9989	4.9984	6.1875	1.0037	5.0002	6.2877
2	20	2.4142	0.6868	4.6382	29.2781	1.0835	5.0349	36.4821
	50	2.9663	0.6086	4.3624	15.1047	1.0414	4.9991	19.1748
	100	3.3649	0.5308	4.1750	9.6994	1.0131	5.0090	12.9511
	200	3.8389	0.4401	3.8845	7.6109	1.0100	5.0012	10.2676

with $i = 1, 2$ and $j = 1, \ldots, 29$. The histogram and box plot of the response variable observations presented in Figure 1, indicate unimodality and right-skewed structure of data. The result of goodness-of-fit tests indicate that skew normal, lognormal and inverse gaussian distributions could be fitted to the response observations at the 5% significance level . The ML estimates of model parameters and their corresponding 95% asymptotic confidence intervals are presented in Table 3. We also provide the corresponding values for the ML estimators of the parameters for inverse Gaussian ANCOVA model, developed by [19], and lognormal distribution as other possible candidates for modeling a positively skewed data. The normal model is also considered to assess the effect of ignoring the skewness in modeling process. As it can be clearly seen, in skew normal, lognormal and inverse Gaussian ANCOVA models both the effects of covariate and proficiency factor are significant. While the normal model incorrectly indicates that the working experience is not a significant covariate. The negative log-likelihood values along with AIC and BIC criteria for different models are provided in Table 4. Notice that as it is pointed out by [19] the regression coefficients and factor effects are not directly comparable for different models due to their different link functions. Therefore the predicted mean, $\hat{\boldsymbol{\mu}}$, under different models of interest can be compared via the root mean square error of prediction (RMSEP) criterion, presented in Table 4.

Table 2. Salary data for 58 employees in a company in Iran by the level of proficiency and working experience.

	Proficiency			
	Level I		Level II	
No.	Salary (1000,000 Rial)	Working Experience (Year)	Salary (1000,000 Rial)	Working Experience (Year)
1	13.21067	13	22.37932	16
2	17.06074	22	18.49741	11
3	15.59944	17	15.24213	10
4	14.71969	15	15.09468	11
5	15.36749	18	15.91385	10
6	16.06207	20	18.95504	15
7	19.89032	11	15.32940	9
8	13.40146	13	15.11344	8
9	11.32948	13	14.69974	10
10	16.11468	16	17.81728	11
11	12.02035	16	19.44098	17
12	11.43113	11	18.08340	13
13	15.06282	13	12.64286	4
14	14.66817	17	14.48483	8
15	13.40854	19	20.44149	14
16	14.63256	13	18.36701	12
17	16.38201	18	25.21042	22
18	13.41755	10	15.72097	10
19	13.49431	16	20.43385	14
20	19.67476	16	17.28177	11
21	12.79719	15	16.72111	11
22	13.80254	18	19.04832	12
23	17.15738	23	14.69331	8
24	16.31889	16	14.76271	10
25	12.91568	14	17.64608	12
26	12.68716	11	17.11542	10
27	11.24524	10	19.80369	17
28	14.62928	18	16.30396	7
29	17.96321	26	18.22075	6

Table 3. The ML estimates and 95% asymptotic confidence intervals of parameters for the skew normal ANCOVA model along with corresponding values for normal, lognormal and inverse Gaussian models.

	Parameters		
ANCOVA Model	β_0	β_1	γ
Normal	7.67	4.70	0.44
	(5.59,9.76)	(3.20,6.21)	(-0.07,0.96)
Lognormal	2.25	0.29	0.03
	(2.12,2.39)	(0.22,0.36)	(0.02,0.03)
Inverse Gaussian	0.063	0.005	-0.001
	(0.061,0.065)	(0.003,0.007)	(-0.002,-0.001)
Skew Normal	14.65	2.95	0.46
	(13.65,15.65)	(1.93,3.97)	(0.28,0.65)

It is seen that the skew normal model has better fit to the data than those of the other models. Moreover, the values of the RMSEP for different models indicate that the skew normal model leads to a model with higher predictive power than other models. Of course, it is clear that the main advantage of the skew normal model to lognormal and inverse Gaussian models is its applicability for symmetric, right-skewed and left-skewed

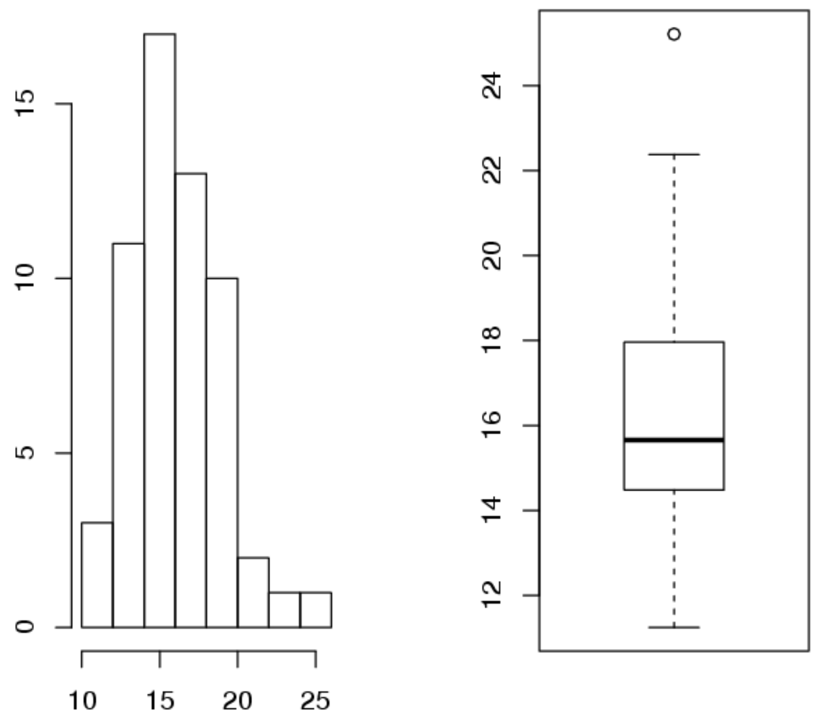

Figure 1. The histogram and box plot of the response variable observations.

Table 4. The values of negative log-likelihood, AIC and BIC and RMSEP criteria for skew normal ANCOVA model along with corresponding values for normal, lognormal and inverse Gaussian mode.

	Goodness-of-fit Criteria			
ANCOVA Model	-loglike	AIC	BIC	RMSEP
Normal	408.65	821.31	825.43	6.5931
Lognormal	291.40	588.80	594.98	5.2246
Inverse Gaussian	156.59	319.19	325.37	2.7994
Skew Normal	140.34	286.68	292.86	1.8062

data, whereas lognormal and inverse Gaussian models can be used only for analyzing right-skewed data.

7. Conclusions

In many real world applications the normality assumption does not hold. Too many researches have been recently focused to develop suitable methods for dealing with non-normality. Particularly, in many real world applications the response variable reflects a unimodal skewed structure. In these cases, the skew normal family of distributions due to its flexibility can be used for data analysis. The results show that in this situations the skew normal ANCOVA model leads to the more efficient estimations of the model parameters than the normal model. Moreover it is a considerably good rival for other traditional models such as lognormal and inverse Gaussian for analyzing skewed data. It is obvious that, the proposed ANCOVA model can be used when the data are symmetric, because the skew normal family of distribution includes the normal distribution as a special case. In this paper, we employed Sahu [21] skew normal distribution among other families of skew normal distributions due to its interesting distributional properties such as simple implementing of the EM algorithm. But other families of the skew normal distribution can be employed in a similar manner.

Acknowledgements

We thank the reviewers for their helpful suggestions, which improved our manuscript.

References

[1] Aigner, J. Lovell, C. A. K. and Schmidt, P. *Formulation and estimation of stochastic frontier production function model*, Journal of Econometrics **12**, 21-37, 1977.

[2] Arellano-Valle, R. B., Bolfaine, H., and Lachos, V. H. *Skew normal linear mixed models*, J. Data Sci. **3**, 415-438, 2005a.

[3] Azzalini, A. *A class of distributions which includes the normal ones*, Scandinavian Journal Statistics **12** 171-178, 1985.

[4] Azzalini, A. *The skew-normal distribution and related multivariate families*, Scandinavian Journal of Statistics **32** 159-188, 2005.

[5] Azzalini, A. and Capitanio, A. *Statistical application of the multivariate skew-normal Distribution*, Journal of Royal Statistical Society, Ser. B **61**, 579-602, 1999.

[6] Azzalini, A. and Dalla Valle, A. *The multivariate skew-normal distribution*, Biometrika **83** 715-726, 1996.

[7] Baghfalaki, T., Ganjali1, M. and Khounsiavash, M. *A Non-random dropout model for analyzing longitudinal skew-normal response*, JIRSS **11**, 101-129, 2012.

[8] Barr, R., Donald, E. and Sherril. *Mean and variance of truncated normal distribution*, The American Statistician. **53** 357-361, 1999.

[9] Cancho, V. Lachos, V. and Ortega, E. *A nonlinear regression model with skew-normal errors*, Statistical Paper. **51** 547-558, 2008.

[10] Cochran, W. G. *Analysis of covariance: Its nature and uses*, Biometrics, **13**, 261-281, 1957.

[11] Cox, D. R. and McCullagh, P. *Some aspects of analysis of analysis of covariance*, Biometrics **38**, 541-561, 1982.

[12] De la Cruz R. and Branco M. *Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves*, Biometrical Journal **51**, 588-609, 2009.

[13] Ferreira, J. T. A. S. and Steel, M. F. J. *Bayesian multivariate skewed regression modelling with an application to firm size*, in Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality (Edited by M. G. Genton), (CRC Chapman & Hall, Boca Raton, 2008), 175-190.

[14] Fisher, R. A. *Statistical methods for research workers*, it 4th ed., (Oliver and Boyed, Edinburgh, 1932).

[15] Genton, M. G. Skew-elliptical distributions and their applications. (Chapman and Hall, New York, 2004).

[16] Gupta, A. K., Gonzalez-Faras, G. and Domnguez-Monila, J. A. *A multivariate skew normal distribution*, J. Multivariate Anal **89**, 181-190, 2004.

[17] Lehmann, E. L. Elements of Large Sample Theory. (Springer-Verlag, New York, 1999).

[18] Lin, T. I., Lee, J. C., *Estimation and prediction in linear mixed models with skew normal random effects for longitudinal data*. Stat. Med **27**, 14901507, 2008.

[19] Meshkani, M. R., Fallah, A. and Kavousi, A. *Analysis of Covariance under inverse Gaussian model*, Journal of Applied Statistics **41**, 1189-1202, 2014.

[20] O'Hagan, A. and Leonard, T. *Bayes estimation subject to uncertainty about parameter constraints*, Biometrika **63**, 201-202, 1976.

[21] Roberts, C. *A correlation model useful in the study of twins.* Journal of American Statistical Association **61** 1184-1190, 1966.

[22] Sahu, S. Dey, D. and Branco, M. *A new class of multivariate distributions with applications to Bayesian regression models*, Canadian Journal of Statistics. **29**, 129-150, 2003.