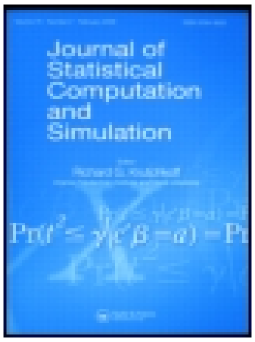

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: http://www.tandfonline.com/loi/gscs20

Robust estimation and testing in one-way ANOVA for Type II censored samples: skew normal error terms

Nuri Çelik & Birdal Şenoğlu

To cite this article: Nuri Çelik & Birdal Şenoğlu (2018): Robust estimation and testing in one-way ANOVA for Type II censored samples: skew normal error terms, Journal of Statistical Computation and Simulation, DOI: 10.1080/00949655.2018.1433670

To link to this article: https://doi.org/10.1080/00949655.2018.1433670

Published online: 07 Feb 2018.

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2018
https://doi.org/10.1080/00949655.2018.1433670

Robust estimation and testing in one-way ANOVA for Type II censored samples: skew normal error terms

Nuri Çelik[a] and Birdal Şenoğlu[b]

[a]Department of Statistics, Bartin University, Bartin, Turkey; [b]Department of Statistics, Ankara University, Ankara, Turkey

ABSTRACT

Censoring can be occurred in many statistical analyses in the framework of experimental design. In this study, we estimate the model parameters in one-way ANOVA under Type II censoring. We assume that the distribution of the error terms is Azzalini's skew normal. We use Tiku's modified maximum likelihood (MML) methodology which is a modified version of the well-known maximum likelihood (ML) in the estimation procedure. Unlike ML methodology, MML methodology is non-iterative and gives explicit estimators of the model parameters. We also propose new test statistics based on the proposed estimators. The performances of the proposed estimators and the test statistics based on them are compared with the corresponding normal theory results via Monte Carlo simulation study. A real life data is analysed to show the implementation of the methodology presented in this paper at the end of the study.

ARTICLE HISTORY
Received 1 November 2017
Accepted 24 January 2018

KEYWORDS
One-way ANOVA; Type II censoring; modified likelihood; skew normal; efficiency

1. Introduction

Censoring is a condition in which some data cannot be observed precisely because of the time and cost constraints. It is frequently encountered in many statistical analyses. For example, it may not be possible to get all clinical data related to some specific disease or to observe all life time data of the components of some expensive systems. In such occasions, censored data are carried after the experiments or the measurements. Censoring is frequently used in various areas of applications such as medicine, insurance, engineering and quality control etc., see [1–9]. There are several types of censoring. In this study, Type II censoring is considered. Type II censoring occurs, if a predetermined number (or proportion) of the smallest or a predetermined number (or proportion) of the largest observation are censored.

Consider the one-way analysis of variance (ANOVA) model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \ldots, a; \; j = 1, 2, \ldots, n, \tag{1}$$

where y_{ij} are the responses corresponding to the jth observation in the ith treatment, μ is the overall mean, α_i is the effect of the ith treatment and ϵ_{ij} are the independent and identically distributed (iid) random error terms.

CONTACT Nuri Çelik ncelik@bartin.edu.tr Department of Statistics, Bartin University, 74100 Bartin, Turkey

It is generally assumed that the errors are normally distributed with mean 0 and the variance σ^2 in order to estimate the unknown parameters in model (1). Maximum likelihood (ML) methodology is used to obtain the estimators of the unknown parameters, since it is the most efficient methodology when the regularity conditions are satisfied. It should be noted that the ML methodology is equivalent to the least squares (LS) methodology under the normality assumption. However, non-normal distributions are more prevalent than the normal distribution in practice. It is known that the ML methodology has some computational difficulties because of the intractable terms in the likelihood equations, see for example [10]. To overcome the computational difficulties encountered in solving the likelihood equations, we use the modified maximum likelihood (MML) methodology proposed by Tiku [11]. Unlike the ML methodology, the MML methodology is non-iterative and gives closed form estimators of the model parameters. It has been widely used in many statistical applications including experimental design, see [12–20].

In the context of Type II censored data, the smallest r_{1i} $(i = 1, 2, \ldots, a)$ and the largest r_{2i} $(i = 1, 2, \ldots, a)$ observations in ith treatment cannot be observed because of the time and the cost constraints. Therefore, the r_{1i} observations from the below and the r_{2i} observations from the above are censored. Then, we have $n - r_{1i} - r_{2i}$ observations in each treatment. Tiku [4] and Tiku and Stewart [10] considered MML methodology to estimate the parameters when the distribution of the error terms is Type II censored normal. Senoglu and Tiku [21] used MML methodology in one-way ANOVA model under Type I and Type II censoring when the distribution of the error terms is Weibull and Arslan and Senoglu [22] considered one-way ANOVA model under the Jones and Faddy's skew t error distribution.

In this study, we assume that the distribution of the error terms is skew normal (SN). The probability density function (pdf) of the SN distribution is given by

$$h(z) = 2\phi(z)\Phi(\lambda z), \quad -\infty < z < \infty, \ -\infty < \lambda < \infty, \tag{2}$$

where $\phi(z)$ and $\Phi(z)$ are the pdf and the cumulative distribution function (cdf) of the standard normal distribution, respectively [23,24]. λ is the skewness parameter, it is also known as the shape parameter since it regulates the shape of the distribution. If a random variable is distributed SN with the skewness parameter λ, then it is denoted by SN(λ). The reason for choosing the SN(λ) as an error distribution in this study is that it includes the normal distribution as well as the plausible alternatives of it with different levels of skewness and kurtosis. SN distribution is widely used in various different areas of science such as reliability, engineering, finance and climatology, see [25–30].

It may be noted that for $\lambda = 0$, SN(λ) reduces to the well-known standard normal distribution N(0,1). When $\lambda \to \mp\infty$, SN(λ) converges to the half-normal distribution. The location-scale form of SN(λ) distribution is given below

$$h(y) = 2\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right), \tag{3}$$

where $\mu \in R$ is the location parameter and $\sigma \in R^+$ is the scale parameter. If the random variable Y has SN(λ) distribution with the parameters μ, σ and λ, then it is denoted by

$Y \sim \mathrm{SN}(\mu, \sigma, \lambda)$. Expected value and the variance of the random variable Y are given by

$$E(Y) = \mu + \sqrt{\frac{2\lambda^2}{\pi(1+\lambda^2)}}\sigma \quad \text{and} \quad V(Y) = \left(1 - \frac{2\lambda^2}{\pi(1+\lambda^2)}\right)\sigma^2, \tag{4}$$

respectively. The cdf of $\mathrm{SN}(\lambda)$ distribution is defined as follows;

$$H(z) = \Phi(z) - 2T(z, \lambda), \tag{5}$$

where $T(z,\lambda) = \int_z^\infty \int_0^{\lambda s} \phi(s)\phi(t)\,\mathrm{d}s\,\mathrm{d}t$ is the Owen function.

The rest of the paper is organized as follows; in Section 2, the MML estimators of the model parameters are obtained and a new test statistic based on the MML estimators for testing the equality of treatment means is proposed. Monte Carlo simulation results for comparing the performeances of the MML estimators with the corresponding LS estimators and also simulated Type I errors for the tests based on them are given in Section 3. In Section 4, we give an application to illustrate the implementation of the proposed methodology. Concluding remarks and comments are given at the end of the paper.

2. Estimation of the parameters and hypothesis testing

In this section, we estimate the unknown parameters in Equation (1) when the distribution of the error terms is Type II censored $\mathrm{SN}(\lambda)$. We use two different methodologies; one is iterative and the other one is non-iterative. Additionally, we propose test statistics based on these estimators in order to test the equality of the treatment effects. We now give the details of these methodologies.

2.1. Maximum likelihood estimation

ML methodology is based on the idea of maximizing the likelihood function with respect to the parameters of interest. Likelihood function (L) is given by

$$L(z_{ij}, \mu, \alpha_i, \sigma) = \prod_{i=1}^{a} \prod_{j=r_{1i}+1}^{n-r_{2i}} \frac{n!}{r_{1i}!r_{2i}!} f(z_{ij}) \prod_{i=1}^{a} \left[F(z_{ij})\right]^{r_{1i}} \prod_{i=1}^{a} \left[1 - F(z_{ij})\right]^{r_{2i}}, \tag{6}$$

where $z_{ij} = (y_{ij} - \mu - \alpha_i)/\sigma$. Based on Equation (6), likelihood function for the SN distribution under Type II censored samples is obtained as shown below

$$L(z_{ij}, \mu, \alpha_i, \sigma, \lambda) = \prod_{i=1}^{a} \prod_{j=r_{1i}+1}^{n-r_{2i}} \frac{n!}{r_{1i}!r_{2i}!} \frac{2}{\sigma} \phi(z_{ij})\Phi(\lambda z_{ij}) \prod_{i=1}^{a} \left[\Phi(z_{ij}) - 2T(z_{ij}, \lambda)\right]^{r_{1i}}$$
$$\cdot \prod_{i=1}^{a} \left[1 - (\Phi(z_{ij}) - 2T(z_{ij}, \lambda))\right]^{r_{2i}}. \tag{7}$$

The log-likelihood function (lnL) can be obtained by taking the natural logarithm of L

$$lnL = \left(\sum_{i=1}^{a}(n - r_{1i} - r_{2i})\right)\ln\sigma - \frac{1}{2}\sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} z_{ij}^2 + \sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} ln\left(\Phi(\lambda z_{ij})\right)$$
$$+\sum_{i=1}^{a} r_{1i}\ln\left(\Phi(z_{ij}) - 2T(z_{ij},\lambda)\right) + \sum_{i=1}^{a} r_{2i}\ln\left(1 - (\Phi(z_{ij}) - 2T(z_{ij},\lambda))\right). \quad (8)$$

By differentiating lnL with respect to the parameters of interest and setting them equal to zero, we obtain the following likelihood equations;

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} z_{ij} - \frac{\lambda}{\sigma}\sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} g(z_{ij}) - \frac{1}{\sigma}\sum_{i=1}^{a} r_{1i}g_1(z_{i(r_{1i}+1)})$$
$$+\frac{1}{\sigma}\sum_{i=1}^{a} r_{2i}g_2(z_{i(n-r_{2i})}) = 0,$$
$$\frac{\partial \ln L}{\partial \alpha_i} = \frac{1}{\sigma}\sum_{j=r_{1i}+1}^{n-r_{2i}} z_{ij} - \frac{\lambda}{\sigma}\sum_{j=r_{1i}+1}^{n-r_{2i}} g(z_{ij}) - \frac{1}{\sigma}r_{1i}g_1(z_{i(r_{1i}+1)}) + \frac{1}{\sigma}r_{2i}g_2(z_{i(n-r_{2i})}) = 0 \quad (9)$$

and

$$\frac{\partial \ln L}{\partial \sigma} = \sum_{i=1}^{a}\frac{(n - r_{1i} - r_{2i})}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} z_{ij}^2 - \frac{\lambda}{\sigma}\sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} g(z_{ij})z_{ij}$$
$$-\frac{1}{\sigma}\sum_{i=1}^{a} r_{1i}g_1(z_{i(r_{1i}+1)})z_{i(r_{1i}+1)} + \frac{1}{\sigma}\sum_{i=1}^{a} r_{2i}g_2(z_{i(n-r_{2i})})z_{i(n-r_{2i})} = 0,$$

where $g(z) = \phi(\lambda z)/\Phi(\lambda z)$, $g_1(z) = f(z)/F(z)$, $g_2(z) = f(z)/(1 - F(z))$, $f(z) = (2/\sigma)\phi(z)\Phi(\lambda z)$, $F(z) = \int_{-\infty}^{z} f(z)\,\mathrm{d}z$, $z_{ij} = (y_{ij} - \mu - \alpha_i)/\sigma$, $z_{i(r_{1i}+1)} = (y_{i(r_{1i}+1)} - \mu - \alpha_i)/\sigma$ and $z_{i(n-r_{2i})} = (y_{i(n-r_{2i})} - \mu - \alpha_i)/\sigma$.

Solutions of these equations are called as the ML estimators. However, they have no algebraic solutions, we therefore resort to iterative methods because of the a+1 nonlinear expressions in Equation (9), see [31] in the context of complete data. Iterative methods are problematic because of the following reasons: (i) non-convergence of iterations (ii) convergence to the wrong root (iii) convergence to multiple roots, see [32]. Hence, in order to get rid of the computational difficulties encountered in iterative methods and maintain the good properties of ML estimators, we use the MML methodology.

2.2. MML estimation

MML methodology is based on the idea of linearizing the intractable terms in the likelihood equations. The methodology proceeds as follows: Let,

$$y_{i(r_{1i}+1)} < y_{i(r_{1i}+2)} < \cdots < y_{i(n-r_{2i})}, \quad i = 1, 2, \ldots, a \quad (10)$$

be the order statistics arranged in the ascending order. Then, the ordered observations are used in the likelihood equations, since the complete sums are invariant to ordering, that is, $\sum_{i=1}^{a} f(y_i) = \sum_{i=1}^{a} f(y_{(i)})$, where $f(.)$ is any function.

In the next step, nonlinear functions in the likelihood equations in (9) are linearized by using the first two terms of Taylor series expansion around the expected values of the standardized order statistics, that is, $t_{i(j)} = E(z_{i(j)})$, $t_{1i} = E(z_{i(r_{1i}+1)})$ and $t_{2i} = E(z_{i(n-r_{2i})})$. This linearization yields

$$\begin{aligned} g(z_{i(j)}) &\cong \alpha_{i(j)} - \beta_{i(j)} z_{i(j)}, \quad j = r_{1i+1}+1, \ldots, n-r_{2i}, \\ g_1(z_{i(r_{1i}+1)}) &\cong \alpha_{i(r_{1i}+1)} - \beta_{i(r_{1i}+1)} z_{i(r_{1i}+1)} \end{aligned} \tag{11}$$

and

$$g_2(z_{i(n-r_{2i})}) \cong \alpha_{i(n-r_{2i})} + \beta_{i(n-r_{2i})} z_{i(n-r_{2i})},$$

where

$$\alpha_{ij} = \frac{\phi(\lambda t)}{\Phi(\lambda t)}\left[1 + t\left(\frac{\lambda^2 t \Phi(\lambda t) + \lambda\phi(\lambda t)}{\Phi(\lambda t)}\right)\right],$$

$$\beta_{ij} = 1 + \frac{\phi(\lambda t)}{\Phi(\lambda t)}\left(\frac{\lambda^2 t \Phi(\lambda t) + \lambda\phi(\lambda t)}{\Phi(\lambda t)}\right), \quad t = t_{i(j)}$$

$$\alpha_{i(r_{1i}+1)} = \frac{f(t_1)}{q_{1i}} + t_1\beta_{1i}, \quad \beta_{1i} = -\left[\frac{f'(t_1)}{q_{1i}} - \frac{(f(t_1))^2}{q_{1i}^2}\right], \quad q_{1i} = \frac{r_{1i}}{(n+1)}, \quad t_1 = t_{1i}$$

and

$$\alpha_{i(n-r_{2i})} = \frac{f(t_2)}{q_{2i}} - t_2\beta_{2i}, \quad \beta_{2i} = \frac{f'(t_2)}{q_{2i}} - \frac{(f(t_2))^2}{q_{2i}^2}, \quad q_{2i} = 1 - \frac{r_{2i}}{(n+1)}, \quad t_2 = t_{2i}.$$

The expected values of the standardized order statistics can not be obtained exactly. Therefore, we obtain their approximate values by solving the following equality.

$$\int_{-\infty}^{t_{i(j)}} f_{\mathrm{SN}}(z)\,\mathrm{d}z = \frac{j}{n+1}, \quad j = 1, 2, \ldots, n.$$

Also, t_1 and t_2 are called population quantiles and obtained by solving the following equalities $F(t_1) = r_1/n$ and $F(t_2) = 1 - r_2/n$, respectively. The use of approximate values does not effect the efficiencies of the MML estimators.

Incorporating Equation (11) into Equation (9), we obtain the following modified likelihood equations

$$\begin{aligned} \frac{\partial \ln L^*}{\partial \mu} = &\frac{1}{\sigma}\sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} z_{i(j)} - \frac{\lambda}{\sigma}\sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} (\alpha_{i(j)} - \gamma_{i(j)} z_{i(j)}) \\ &- \frac{1}{\sigma}\sum_{i=1}^{a} r_{1i}(\alpha_{i(r_{1i}+1)} - \beta_{i(r_{1i}+1)} z_{i(r_{1i}+1)}) \\ &+ \frac{1}{\sigma}\sum_{i=1}^{a} r_{2i}(\alpha_{i(n-r_{2i})} + \beta_{i(n-r_{2i})} z_{i(n-r_{2i})}) = 0, \end{aligned}$$

$$\begin{aligned}\frac{\partial \ln L^*}{\partial \alpha_i} &= \frac{1}{\sigma}\sum_{j=r_{1i}+1}^{n-r_{2i}} z_{i(j)} - \frac{\lambda}{\sigma}\sum_{j=r_{1i}+1}^{n-r_{2i}} (\alpha_{i(j)} - \gamma_{i(j)} z_{i(j)}) \\ &\quad - \frac{1}{\sigma} r_{1i}(\alpha_{i(r_{1i}+1)} - \beta_{i(r_{1i}+1)} z_{i(r_{1i}+1)}) \\ &\quad + \frac{1}{\sigma} r_{2i}(\alpha_{i(n-r_{2i})} + \beta_{i(n-r_{2i})} z_{i(n-r_{2i})}) = 0 \end{aligned} \tag{12}$$

and

$$\begin{aligned}\frac{\partial \ln L^*}{\partial \sigma} &= \sum_{i=1}^{a} \frac{(n-r_{1i}-r_{2i})}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} z_{i(j)}^2 - \frac{\lambda}{\sigma}\sum_{i=1}^{a}\sum_{j=r_{1i}+1}^{n-r_{2i}} (\alpha_{i(j)} - \gamma_{i(j)} z_{i(j)}) z_{i(j)} \\ &\quad - \frac{1}{\sigma}\sum_{i=1}^{a} r_{1i}(\alpha_{i(r_{1i}+1)} - \beta_{i(r_{1i}+1)} z_{i(r_{1i}+1)}) z_{i(r_{1i}+1)} \\ &\quad + \frac{1}{\sigma}\sum_{i=1}^{a} r_{2i}(\alpha_{i(n-r_{2i})} + \beta_{i(n-r_{2i})} z_{i(n-r_{2i})}) z_{i(n-r_{2i})} = 0. \end{aligned}$$

It should be noted that if $g(z)$ is bounded, $z_{i(j)}$ tends to its expected value $t_{i(j)}$ and $[g(z_{i(j)}) - (\alpha_{i(j)} + \beta_{i(j)} z_{i(j)})]$, $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, n$ tends to zero as n tends to infinity, see [33]. Similar statements can also be done for the differences $[g_1(z_{i(r_{1i}+1)}) - (\alpha_{i(r_{1i}+1)} - \beta_{i(r_{1i}+1)} z_{i(r_{1i}+1)})]$ and $[g_2(z_{i(n-r_{2i})}) - (\alpha_{i(n-r_{2i})} + \beta_{i(n-r_{2i})} z_{i(n-r_{2i})})]$. Therefore, equations given in (9) are asymptotically equivalent to the equations given in (12). In other words, under some regularity conditions, the modified likelihood equations are asymptotically equivalent to the corresponding likelihood equations, that is, $\lim_{n\to\infty}(1/n)(\partial \ln L/\partial\mu - \partial \ln L^*/\partial\mu) = 0$, $\lim_{n\to\infty}(1/n)(\partial \ln L/\partial\alpha_i - \partial \ln L^*/\partial\alpha_i) = 0$ and $\lim_{n\to\infty}(1/n)(\partial \ln L/\partial\sigma - \partial \ln L^*/\partial\sigma) = 0$, see [34]. The solutions of these modified likelihood equations are called as the MML estimators.They are given by

$$\hat{\mu} = M + \Delta\hat{\sigma}, \quad \hat{\alpha}_i = (M_i - M) - (\Delta_i - \Delta)\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \frac{-B + \sqrt{B^2 + 4AC}}{2\sqrt{A(A-a)}}, \tag{13}$$

where

$$M_i = \left(\sum_{j=r_{1i}+1}^{n-r_{2i}} \beta_{ij} y_{i(j)} - r_{1i}\beta_{i(r_{1i}+1)} y_{i(r_{1i}+1)} + r_{2i}\beta_{i(n-r_{2i})} y_{i(n-r_{2i})}\right) \Bigg/ m_i,$$

$$M = \sum_{i=1}^{a} m_i M_i / m,$$

$$\Delta_i = \left(-\lambda \sum_{j=r_{1i}+1}^{n-r_{2i}} \alpha_{ij} - r_{1i}\alpha_{i(r_{1i}+1)} + r_{2i}\alpha_{i(n-r_{2i})}\right) \Bigg/ m_i, \quad \Delta = \sum_{i=1}^{a} m_i \Delta_i \Bigg/ m,$$

$$m_i = \sum_{j=r_{1i}+1}^{n-r_{2i}} \beta_{ij} - r_{1i}\beta_{i(r_{1i}+1)} + r_{2i}\beta_{i(n-r_{2i})}, \quad m = \sum_{i=1}^{a} m_i,$$

$$A = \sum_{i=1}^{a} (n - r_{1i} - r_{2i}),$$

$$B = \lambda \sum_{i=1}^{a} \sum_{j=r_{1i}+1}^{n-r_{2i}} \alpha_{ij}(y_{i(j)} - M_i) + \sum_{i=1}^{a} r_{1i}(y_{i(r_{1i}+1)} - M_i) - \sum_{i=1}^{a} r_{2i}(y_{i(n-r_{2i})} - M_i),$$

$$C = \sum_{i=1}^{a} \sum_{j=r_{1i}+1}^{n-r_{2i}} \alpha_{ij}(y_{i(j)} - M_i)^2 + \sum_{i=1}^{a} r_{1i}(y_{i(r_{1i}+1)} - M_i)^2 - \sum_{i=1}^{a} r_{2i}(y_{i(n-r_{2i})} - M_i)^2.$$

These estimators reduce to the estimators obtained by Tiku [4] when $\lambda = 0$ and to the estimators obtained by Celik et al. [31] when $r_{1i} = r_{2i} = 0$, $i = 1, 2, \dots, a$.

The MML estimators have good properties such that they are the functions of the sample observations and are easy to compute. Additionally, the MML estimators of the parameters are asymptotically equivalent to the ML estimators. Therefore, they are asymptotically minimum variance bound (MVB) estimators and unbiased besides being asymptotically normal and consistent [34,35]. Even for small sample sizes they have high efficiencies [33].

2.3. Hypothesis testing

The null and alternative hypothesis for testing the equality of the treatment means in one-way ANOVA are given below

$$\begin{aligned} &H_0 : \alpha_i = 0, \quad i = 1, 2, \dots, a \\ &H_1 : \alpha_i \neq 0 \quad \text{for at least one i} \end{aligned} \tag{14}$$

For testing the hypothesis, we traditionally use the following test statistic based on the LS estimators.

$$F_{\text{LS}} = \frac{n \sum_{i=1}^{a} \tilde{\alpha}_{i,\text{LS}}^2}{(a-1)\tilde{\sigma}_{\text{LS}}^2}. \tag{15}$$

In this study, we propose the following test statistic based on the MML estimators.

$$F_{\text{MML}} = \frac{\sum_{i=1}^{a} m_i \hat{\alpha}_{i,\text{MML}}^2}{(a-1)\hat{\sigma}_{\text{MML}}^2} \tag{16}$$

F_{MML} has asymptotically central F distribution with degrees of freedom $\nu_1 = a - 1$ and $\nu_2 = \sum_{i=1}^{a}(n - r_{1i} - r_{2i} - 1)$, since the distribution of $\hat{\alpha}_i$ is asymptotically normal and the distribution of the $\hat{\sigma}$ is asymptotically a multiple of chi-square [33]. For small n values, we use Monte Carlo simulation study to verify how close the null distribution of them is to central F in the following section.

3. Simulation study

In this part, we compare the MML estimators with the traditional LS estimators with respect to the mean, variance, mean square error (MSE) and relative efficiency (RE) criteria. First, it should be noted that we are interested in λ values satisfying the property

$0.4 < P(X > E(X)) < 0.6$ in the context of experimental design. We, therefore use λ values satisfying the mentioned condition, that is, we take $-1 < \lambda < 1$ from now on. We use $a = 3$, $n = 10$ and $\lambda = 0.4, 0.7$ and 1.0 for illustration. It is clear that SN distribution is positively skewed when $\lambda > 0$, therefore we only consider the right censoring scheme and use $q = r_{2i}/n,\ i = 1, 2, \ldots, a$ in the simulation study. We obtain the similar results when SN is negatively skewed and $\lambda < 0$, however we did not reproduce them for the sake of brevity. We use three different censoring schemes such as $(0.1, 0.1, 0.1),\ (0.1, 0.2, 0.1)$ and $(0.2, 0.2, 0.2)$. Without loss of generality, we use the simulation setting $\mu_i = 0$ and $\sigma = 1$ in our study. $[|100, 000/n|]$ Monte Carlo runs is used by using Matlab. Here, $[|.|]$ denotes rounding a decimal to the nearest integer number. Simulation results are shown in Tables 1 and 2 for the parameters μ_i and σ, respectively. The REs are defined as follow

$$\mathrm{RE}_{\mathrm{MML}} = \frac{\mathrm{MSE}_{\mathrm{MML}}}{\mathrm{MSE}_{\mathrm{LS}}}.$$

As can be noticed from Table 1, the MML estimator of μ_i is more efficient than the corresponding LS estimator of μ_i for each censoring proportion. Also, it is seen that RE of the MML estimator of μ_i increases as the skewness parameter increases. On the other hand, from Table 2, the MML estimator of the scale parameter is much more efficient than the corresponding LS estimator. However, for the scale parameter, the efficiency of the MML estimator increases as the skewness parameter decreases.

Table 1. Means, variances, MSEs and REs for the LS and the MML estimators of μ_i.

	(q)			Mean		Variance		MSE		RE
	I	II	III	LS	MML	LS	MML	LS	MML	MML
	0.1	0.1	0.1	0.043	0.041	0.096	0.096	0.098	0.098	100
$\lambda = 0.4$	0.1	0.2	0.1	0.035	0.030	0.092	0.092	0.093	0.093	100
	0.2	0.2	0.2	−0.069	−0.030	0.095	0.096	0.100	0.096	96
	0.1	0.1	0.1	0.096	0.038	0.086	0.087	0.095	0.088	93
$\lambda = 0.7$	0.1	0.2	0.1	0.109	0.047	0.080	0.082	0.092	0.094	91
	0.2	0.2	0.2	0.009	0.009	0.084	0.080	0.084	0.080	96
	0.1	0.1	0.1	0.177	−0.028	0.071	0.075	0.102	0.076	74
$\lambda = 1.0$	0.1	0.2	0.1	0.169	−0.039	0.069	0.073	0.097	0.075	77
	0.2	0.2	0.2	0.156	0.061	0.074	0.074	0.098	0.078	79

Table 2. Means, variances, MSEs and REs for the LS and the MML estimators of σ.

	(q)			Mean		Variance		MSE		RE
	I	II	III	LS	MML	LS	MML	LS	MML	MML
	0.1	0.1	0.1	1.246	0.884	0.036	0.018	1.589	0.799	50
$\lambda = 0.4$	0.1	0.2	0.1	1.214	0.872	0.039	0.020	1.513	0.780	52
	0.2	0.2	0.2	1.243	0.833	0.046	0.020	1.592	0.715	45
	0.1	0.1	0.1	1.103	0.885	0.028	0.018	1.244	0.801	64
$\lambda = 0.7$	0.1	0.2	0.1	1.078	0.878	0.028	0.018	1.191	0.789	66
	0.2	0.2	0.2	1.133	0.847	0.035	0.019	1.318	0.737	56
	0.1	0.1	0.1	1.026	0.894	0.025	0.019	1.077	0.819	76
$\lambda = 1.0$	0.1	0.2	0.1	0.997	0.885	0.024	0.019	1.017	0.802	79
	0.2	0.2	0.2	1.015	0.849	0.028	0.019	1.059	0.740	70

Table 3. The simulated Type I errors for F_{LS} and F_{MML} ($a = 3, n = 10, \alpha = 0.050$).

	q				
	I	II	III	LS	MML
	0.1	0.1	0.1	0.059	0.048
$\lambda = 0.4$	0.1	0.2	0.1	0.077	0.051
	0.2	0.2	0.2	0.081	0.053
	0.1	0.1	0.1	0.084	0.052
$\lambda = 0.7$	0.1	0.2	0.1	0.089	0.050
	0.2	0.2	0.2	0.119	0.057
	0.1	0.1	0.1	0.091	0.058
$\lambda = 1.0$	0.1	0.2	0.1	0.096	0.055
	0.2	0.2	0.2	0.116	0.057

In Table 3, simulated Type I errors are given for the F_{MML} and the corresponding F_{LS} tests. We simulate the Type I errors for each test statistic for different censoring proportion to see how close their distribution is to central F with degrees of freedom $\nu_1 = a - 1$ and $\nu_2 = \sum_{i=1}^{a}(n - r_{1i} - r_{2i} - 1)$. It can be easily seen from Table 3 that Type I errors of the test statistics are not affected too much for small deviations of λ from 0. However, Type I error of the F_{LS} test becomes larger than the predetermined nominal value of α (here 0.050) as the skewness parameter λ gets close to 1. It can be concluded that the test statistic based on the LS estimators are not distributed as central F for λ values not close to zero. On the other hand, F distribution provides a satisfactory approximation to the percentage points for F_{MML} test for all λ values and censoring proportion.

4. Application

The data set is about the drying times of three types of the cements. This application is a typical example of one-way ANOVA with three treatments. All types of the cements are applied to the certain place and the drying times based on minutes are recorded. The data set is given in Table 4.

The drying times of cement data is censored with the rule of Type II censoring schemes. In all treatments, there is one outlier that should be censored above. Therefore, $r_{21} = r_{22} = r_{23} = 1$ is observed. Before starting to analyse data, the shape parameter (λ) must be identified. For this reason, we use EM algorithm to estimate the shape parameter. The detailed information about identifying the shape parameter with EM algorithm for SN distribution is given by Celik et al. [31]. The shape parameter is determined to be

Table 4. Drying times of the cements (minutes).

The cements		
A	B	C
23	25	34
21	27	39
36	29	143
153	159	40
25	35	31

Table 5. The parameter estimates and the calculated F values for the censored data.

	μ	α_1	α_2	α_3	σ	F
LS	29.380	−4.167	−1.417	5.584	9.132	1.21
MML	28.478	−4.328	−1.378	5.706	5.705	3.79

0.6147. Kolmogrov–Smirnov (K–S) test is also applied in order to test the following null hypothesis:

$$H_0 : \text{Data set has the SN}(\lambda) \text{ distribution.}$$

We use the test statistics

$$D = \sup_x [|F_n(x) - F_0(x)|],$$

where $F_0(.)$ is the cdf of an SN distribution with known skew parameter λ and $F_n(.)$ is the empirical cdf. K–S test statistic is calculated to be 0.337 by using residuals obtained from

$$\tilde{\epsilon_{ij}} = y_{ij} - \tilde{\mu} - \tilde{\alpha_i}; \quad i = 1, 2, 3, \;\; j = 1, 2, \ldots, n - r_{2i}.$$

Therefore, it can be concluded that the distribution is SN with the skewness parameter 0.614 at $\alpha = 0.05$ significance level since the table value for the K–S test is 0.375.

Table 5 shows the result for the censored data. As can be seen from the table, the ML estimate of μ_i are very close to the LS estimate of μ_i. However, $\hat{\sigma}_{\text{MML}} = 5.705 < \tilde{\sigma}_{\text{LS}} = 9.132$. Therefore, test statistics based on MML estimators is more reliable than the test statistics based on LS estimators. Additionally, the null hypothesis

$$H_0 : \text{There is no difference between the drying times of the cements.}$$

is rejected according to the $F_{\text{MML}} = 3.79$, but it is not rejected according to the $F_{\text{LS}} = 1.21$ in the significance level 0.10. Therefore, p-value obtained by F_{MML} is greater than the one obtained by F_{LS} due to the smaller standard errors of the MML estimato. This is another indication of the superiority of F_{MML} to F_{LS}.

5. Conclusion

Traditionally, LS estimators and the test statistics based on them are used in the context of experimental design. However, when the normality assumption is not satisfied, the efficiencies of these estimators and the power of the test statistics based on them are low. In this study, we assume the distribution of the error terms is SN in one-way ANOVA models under Type II censoring. It is too complicated to find the explicit estimators of the unknown parameter even the distribution of the error terms is normal. Therefore, we propose another estimation methodology (MML) which gives explicit solutions and easy to compute. Simulation study shows that the MML estimators have better efficiencies for different censoring proportion and λ values. Additionally, the test statistics based on the MML estimators provide accurate approximation to the F distribution, while the test statistics based on the LS estimators diverges from F distribution for some censoring proportion and λ values. Therefore, the MML estimation methodology is more preferable and reliable in terms of efficiency and the approximation to the true distribution.

Disclosure statement

We thank to the editor and the reviewers for their valuable comments and suggestions which greatly improved the paper.

References

[1] Balakrishnan N, Asgharzadeh A. Inference for the scaled half-logistic distribution based on progressively Type II censored samples. Comm Statist Theory Methods. 2005;34:199–208.
[2] Cohen AC. On the solution of estimating equations for truncated and censored samples from normal populations. Biometrika. 1957;44:225–236.
[3] Sampford MR, Taylor J. Censored observations in randomized block experiments. J R Stat Soc. 1959;21B:214–237.
[4] Tiku ML. Testing group effects from Type II censored normal samples in experimental design. Biometrics. 1973;29:25–33.
[5] Asgharzadeh A. Approximate MLE for the scaled generalized exponential distribution under progressive type-II censoring. J Korean Statist Soc. 2009;38:223–229.
[6] Stotvig JG. Censored Weibull distributed data in experimental design. Norway: Norwegian University of Science and Technology; 2014.
[7] Sue-Chu AM. Methods for dealing with censored data using experimental design. Norway: Trondheim, Norwegian University of Science and Technology; 2013.
[8] Parsi S, Ganjali M, Farsipour NS. Simultaneous confidence intervals for the parameters of Pareto distribution under progressive censoring. Comm Statist Theory Methods. 2010;39:94–106.
[9] Aly HM, Ismail AA. Optimum simple time-step stress plans for partially accelerated life testing with censoring. Far East J Theor Stat. 2008;24(2):175–200.
[10] Tiku ML, Stewart DE. Estimating and testing group effects from Type I censored normal samples in experimental design. Comm Statist Theory Methods. 1977;A6-15:1485–1501.
[11] Tiku ML. Estimating the mean and standard deviation from a censored normal sample. Biometrika. 1967;54:155–165.
[12] Chen W, Xie M, Wu M. Modified maximum likelihood estimator of scale parameter using moving extremes ranked set sampling. Comm Statist Simulation Comput. 2014;45(6):2232–2240.
[13] Cohen AC, Whitten B. Modified maximum likelihood and modified moment estimators for the three-parameter Weibull distribution. Comm Statist Simulation Comput. 2010;41:2631–2656.
[14] Islam T, Shaibur MR, Hossein SS. Effectivity of modified maximum likelihood estimators using selected ranked set sampling data. Austrian J Statist. 2009;38:109–120.
[15] Kasap P, Senoglu B, Arslan O, et al. Estimating the location and scale parameters of the GT distribution. New World Sci Acad. 2011;6(3):103–111.
[16] Senoglu B. Robust factorial design with Weibull error distributions. J Appl Stat. 2005;32:1051–1066.
[17] Senoglu B. Estimating parameters in one-way analysis of covariance model with short-tailed symmetric error distributions. J Comput Appl Math. 2007;201:275–283.
[18] Senoglu B. Robust estimation and hypothesis testing of linear contrasts in analysis of covariance with stochastic covariates. J Appl Stat. 2007;34:141–151.
[19] Senoglu B, Kasap P, Acitas S. Robust 2^k designs: nonnormal symmetric distributions. Pakistan J Statist. 2012;28:93–114.
[20] Tiku ML, Senoglu B. Estimation and hypothesis testing in BIB design and robustness. Comput Statist Data Anal. 2009;53:3439–3451.
[21] Senoglu B, Tiku ML. Censored and truncated samples in experimental design under nonnormality. Stat Methods. 2004;6(2):173–199.
[22] Arslan T, Senoglu B. Type II censored samples in experimental design under Jones and Faddy's skew t distribution. Iran J Sci Technol Trans A Sci. 2017.
[23] Azzalini A. A class of distributions which includes the normal ones. Scand J Statist. 1985;12:171–178.

[24] Azzalini A. Further results on a class of distributions which includes the normal ones. Statistica. 1986;46:199–208.
[25] Abtahi A, Towhidi M, Behboodian J. An appropriate empirical version of skew-normal density. Statist Papers. 2011;52:469–489.
[26] Bolance C, Guillen M, Pelican E, et al. Skewed bivariate models and nonparametric estimation for the CTE risk measure. Insurance . 2008;43:386–393.
[27] Eling M. Fitting insurance claims to skewed distributions: are the skew-normal and skew-student good models? Insurance: Mathematics and Economics. 2012;51(2):239–248.
[28] Franceschini C, Loperfido N. A skewed GARCH-type model for multivariate financial time series. In: Corazza M, Pizzi C, editors. Mathematical and statistical methods for actuarial sciences and finance. Vol. 12; 2010. p. 143–152.
[29] Lien D, Shrestha K. Estimating optimal hedge ratio: a multivariate skewnormal distribution approach. Appl Finan Econ. 2010;20:627–636.
[30] Lio YL, Park C. A bootstrap control chart for inverse Gaussian percentiles. J Stat Comput Simul. 2010;80(3):287–299.
[31] Celik N, Senoglu B, Arslan O. Estimation and testing in one-way ANOVA when the errors are skew-normal. Colombian J Statist. 2015;38(1):75–91.
[32] Barnett VD. Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots. Biometrika. 1966;52:151–165.
[33] Tiku ML, Suresh RP. A new method of estimation for location and scale parameters. J Statist Plann Inference. 1992;30:281–292.
[34] Vaughan DC, Tiku ML. Estimation and hypothesis testing for a nonnormal bivariate distribution with applications. Math Comput Modelling. 2000;32:53–67.
[35] Bhattacharyya GK. The asymptotics of maximum likelihood and related estimators based on type II censored data. J Amer Statist Assoc. 1985;80:398–404.