Int J Adv Manuf Technol (2006) 30: 887–896
DOI 10.1007/s00170-005-0123-9

ORIGINAL ARTICLE

Chien-Hui Yang · Lee-Ing Tong

Predicting type II censored data from factorial experiments using modified maximum likelihood predictor

Received: 16 November 2004 / Accepted: 8 April 2005 / Published online: 21 January 2006

Abstract Censored data are often found in industrial experiments. The censored data are usually predicted by constructing complex statistical models or neural networks. Although a maximum likelihood predictor (MLP) was developed to predict Type II censored data, the likelihood equation may not be obtained for a closed-form solution. A modified maximum likelihood predictor (MMLP) was derived to overcome the problems of MLP. However, because MMLP requires normality assumption with unknown mean and known variance, and because the population variance of real-world experimental data is generally unknown, the MMLP has little practical use. Therefore, this study develops a modified maximum likelihood predictor (MMLP) for Type II censored data obtained from a normal distribution with unknown mean and variance. The predicted censored data using the proposed MMLP are merged with the uncensored data as a pseudo-complete data set. The analysis of variance (ANOVA) method is then employed to determine the optimal factor–level combination settings. The proposed method can also be employed to predict the Type II censored data obtained from Taguchi's parameter designs. Two examples are given to demonstrate the proposed method and the comparisons of the proposed method with existing methods of predicting the Type II censored data are made to demonstrate the effectiveness of the proposed method.

Keywords Type II censored data · Modified maximum likelihood predictor · Factorial design · Taguchi's parameter design · Prediction · Order statistics · Normal distribution

C.-H. Yang · L.-I. Tong (✉)
Department of Industrial Engineering and Management,
National Chiao Tung University,
30050 Hsinchu, Taiwan, Republic of China
e-mail: litong@cc.nctu.edu.tw

1 Introduction

When employing design of experiments (DOE) or Taguchi's parameter design to find the optimal factor-level combination, the experimenter often shortens the experimental time to bring the new product to the market or to reduce the experimental cost. Other restrictions such as lack of material resources, machine malfunctioning and measuring instrument limitations are sometimes encountered in the experiments. In these cases, censored data occurs. When the exact value of an observation cannot be obtained but is known to be less than or equal to a predetermined value, *L*, this observation is said to be left censored at *L*. Similarly, when the exact value of an observation cannot be obtained but is greater than *L*, this observation is said to be right censored at *L* [1]. When the left and right censoring situations arise together, the observation is said to be doubly censored. Generally, two censoring schemes are often used in reliability analysis to save experimental time. These two schemes are Type I and Type II censoring schemes. In the Type I censoring scheme, the experimental ending time *L* is predetermined. The experiment items *n* placed on test are set first, then a lifetime testing experiment is conducted and individual lifetimes of experiments that ended before time *L* are recorded [2]. Lifetimes from the testing experiment are known exactly only for those items that fail before time *L*. Therefore, the number of censored data is random. That is, the Type I censored data is defined as $Y_i=\min(X_i, L)$, $i = 1, 2, \ldots, n$, where X_i is the failure time of the *i*th item. For example, suppose that 50 items of the power supply used in a computer are to be tested. The experimental ending time is set to be 2,000 hours, and individual lifetimes of the power supply that ended before 2,000 hours are then recorded. If the lifetime of the power supply is longer than 2,000 hours, the datum is considered as censored, each with observed time in 2,000 hours. Where the lifetime is less than 2,000 hours, the datum is considered as uncensored [3, 4]. In the Type II censoring scheme, the experiment items *n* placed on test are set first and the number of uncensored data *r* is predetermined. Instead of continuing experiments until all n items have failed, the experiment is terminated when the rth item fails [1]. The remaining $n-r$ items are regarded as censored data, each having the observed failure time of the rth item, $X_{r:n}$. For example, suppose that 50 items of the power supply used in a computer are to be tested. Rather than wasting the items of the power supply, the experiment may be terminated when a predetermined number of the power supplies have burnt out. The experiment is terminated when the 40th lifetime datum is obtained, in which 40 is set in advance. The remaining lifetime data of ten items are regarded as censored data [4]. In the case of censored experimental data, the traditional analysis of variance (ANOVA) method is inappropriate to analyze the effects of factors. Various methods have been developed to predict the censored data and are described in the next section. These methods are computationally complex, or they need to construct complicated statistical models or neural networks. Therefore, these methods are not easy to use for a practitioner. A maximum likelihood predictor (MLP) was derived to predict Type II censored data, but it may not be able to solve the likelihood equation to obtain a closed-form solution for the MLP in most cases. To solve the problems of MLP, a modified maximum likelihood predictor (MMLP) [5] under the normality assumption with unknown mean and known variance was proposed. However, the population variance of experimental data is usually unknown in the real world. Therefore, MMLP has little practical use.

Hence, this study first develops a MMLP that assumes that Type II censored data are obtained from a normal distribution with unknown mean and variance. A novel method for predicting the Type II censored data from repetitious experiments or Taguchi's parameter designs using MMLP is then proposed. The motivation of the proposed method is that the complete experimental data can be analyzed easily; therefore, the censored data can be predicted by the MMLP before analyzing the experimental data. The predicted censored data and the uncensored data are merged as a pseudo-complete data set. The ANOVA method is conducted to determine the optimal settings of factor–level combinations.

The rest of this paper is organized as follows. Section 2 reviews literature relating to predicting Type II censored data and introduces some techniques for analyzing Type II censored data from repetitious experiments. Section 3 proposes a modified maximum likelihood predictor to predict the censored data from a normal distribution with unknown mean and variance. This predictor is extended from the MMLP [5] which assumes that the Type II censored samples are obtained from a normal distribution with unknown mean and known variance. Section 4 describes the implementation of the proposed procedures in factorial designs and Taguchi's experiments with Type II censored data. Section 5 presents examples to illustrate the effectiveness of the proposed method. Section 6 draws conclusions. The proposed method is simper than the procedures used to construct complex statistical models or when using neural networks, and therefore it is useful for the practitioner.

2 Literature review

2.1 MMLP

Kaminsky and Rhodin [6] employed the principle of maximum likelihood to predict future order statistics and estimate the unknown parameters. The predictor is called the MLP. Their study predicted higher order statistics from the lower ones in Type II censored samples. Furthermore, Kaminsky and Rhodin listed conditions for the existence of a unique MLP and gave several examples with various distributions to illustrate their method. However, the likelihood equation cannot be solved to obtain a closed-form solution for the MLP in most cases.

Since solving the predictive likelihood functions to obtain a closed-form expression for the MLP is not feasible in most cases. Raqab [5] proposed MMLP for predicting the future order statistics from normal samples. He replaced $h(x)$, $h_1(x, y)$, and $h_2(x, y)$ with their respective expected values to obtain the MMLP, in which $h(x)$, the hazard rate function, is defined in Eq. 1 and $h_1(x, y)$, and $h_2(x, y)$, the extended hazard rate functions, are defined in Eq. 2. However, Raqab only derived the MMLP from the Type II censored sample drawn from a normal distribution with unknown mean and known variance.

$$h(x) = \frac{f(x)}{1 - F(x)} \tag{1}$$

$$h_1(x, y) = \frac{f(x)}{F(y) - F(x)}, h_2(x, y) = \frac{f(y)}{F(y) - F(x)}, x < y. \tag{2}$$

where $f(\cdot)$ denotes the probability density function (pdf) and $F(\cdot)$ denotes the cumulative distribution function (CDF).

2.2 Analyzing type II censored data from repetitious experiments

Nelson and Hahn [7, 8] developed linear estimation methods for the parameters in the regression models using the ordered observations of censored data. Nelson and Hahn [7] presented two techniques for simple (but not minimum variance) linear estimation on the parameters of a linear regression model involving a single independent variable when the data on the dependent variable are censored and the observations have unequal variances and are correlated. Therein, a method for obtaining the best linear unbiased estimates is not applied since their method leads to a problem of weighted least squares estimation which involves complicated computations. Nelson and Hahn [8] also presented a method for finding the best linear unbiased estimates of the parameters of a linear regression model using censored data on the dependent variable when only one uncensored observation is made and the sample size under each test condition is the same. However, in both [7] and [8], methods of linear estimation were applicable only to Type II censored data. Hahn and Nelson [9] reviewed graphical, maximum likelihood (ML) and linear estimation methods for analyzing censored lifetime data in an investigation of the relationships between stress and product lifetime. They compared the advantages and disadvantages of these methods and selected a preferred one. Graphical methods provide subjective procedures, causing different results according to different individuals. The ML method can be applied to any model, that is, to almost any lifetime distribution and relationships of the distribution parameters. Furthermore, it is appropriate for many types of censored data, including Type I, Type II and single or multiple censoring. Linear estimation methods are computationally simpler than ML methods. Nevertheless, linear estimation methods are applicable only to appropriate Type II censoring. Krall et al. [10] proposed a forward selection procedure for selecting the most important variables associated with survival time data using MLE. The method proposed therein involves fewer computations than the original MLE method. However, the forward selection procedure still involves many calculations and iterations. Schmee and Hahn [11] proposed an iterative model selection method based on an iterative least squares procedure for estimating the censored data. An initial least squares fit is obtained, in which the censored values are considered as failures. Next, the initial fit is employed to estimate the expected failure time for each censored observation. These estimates are then employed to obtain a revised least squares fit and new expected failure times are estimated for the censored values. This procedure is iterated until convergence is reached. However, treating the censored data as if they were uncensored may bias the estimates of the regression line. Therefore, Hahn et al. [12] developed the iterative least squares method to analyze the data from a fractional factorial experiment involving censoring to the left. However, the initial model selection significantly influences the final model. Tong and Su [13] developed an effective procedure based on the transformation of the rank of the responses and regression analysis for analyzing an experiment with singly censored data. The non-parametric method and regression analysis are employed to analyze the censored data from a multi-factor and multi-level experiment. Their procedure not only concurrently analyzes censored data obtained in replicated and unreplicated experiments but also considers the variability of control factors.

Su and Miao [14] developed procedures based on the back-propagation neural network to analyze censored data in replicated experiments. The proposed procedures must construct three neural networks to find the optimal settings of factor-level combinations. Although their procedure does not require any statistical assumption, the network parameters must be selected by trial-and-error. In any case, the procedure cannot identify the significant factors.

2.3 Some results for order statistics

This section reviews three lemmas developed by Raqab [5]. These lemmas are important in approximating h, h_1, h_2 in the predictive likelihood equations (PLEs). Raqab assumed that the derivative of the pdf, $f'(x)$, exists almost everywhere and is absolutely integrable. Under these assumptions, $E\,\Psi\,(X_{i:n})$ exists where

$$\Psi(x) = -\{f'(x)/f(x)\}. \tag{3}$$

These lemmas are given as follows:

Lemma 2.3.1

$$Ef(X_{i:n}) = \frac{1}{n+1}\sum_{k=i+1}^{n+1} E\Psi(X_{k:n+1}),\ i \le n.$$

Lemma 2.3.2

$$Eh(X_{i:n}) = \frac{1}{n-i}\sum_{k=i+1}^{n} E\Psi(X_{k:n}),\ i \le n-1.$$

Lemma 2.3.3

(1) $Eh_1\left(X_{i:n}, X_{j:n}\right) = \dfrac{1}{j-i-1}\displaystyle\sum_{k=i+1}^{n} E\Psi(X_{k:n})(j-i \ge 2),$

(2) $Eh_2\left(X_{i:n}, X_{j:n}\right) = \dfrac{1}{j-i-1}\displaystyle\sum_{k=j}^{n} E\Psi(X_{k:n})(j-i \ge 2),$

where h_1 and h_2 are given by Eq. 2.

3 Modified maximum likelihood predictor

The MMLP can be employed to predict the censored data when the experiment is terminated before finished and not all observations in this experiment are observed. Assume that $X=(X_{1:n}, X_{2:n},\ldots,X_{r:n})$ are observed in the experiment and the goal is to predict $X_{s:n}$ where $1\le r<s\le n$; then under the normality assumption with unknown mean μ and unknown variance σ^2, the predictive likelihood functions (PLF) of μ and σ^2 (i.e. which is denoted as L_1) and the PLF of $X_{s:n}$ (i.e. which is denoted as L_2) are given as follows, respectively:

$$L_1(\mu, \sigma; X) = \frac{n!}{(n-r)!}\prod_{j=1}^{r} f\left(X_{j:n}\right)[1 - F(X_{r:n})]^{n-r}$$

and

$$L_2(X_{s:n};\mu,\sigma,\boldsymbol{X}) = \frac{(n-r)!}{(s-r-1)!(n-s)!}\frac{[F(X_{s:n})-F(X_{r:n})]^{s-r-1}}{[1-F(X_{r:n})]^{n-r}} \times [1-F(X_{s:n})]^{n-s}f(X_{s:n}) \quad (4)$$

For notational clarity, L_1 (μ, σ; $\boldsymbol{X}$) and L_2 ($X_{s:n}$; μ, σ; $\boldsymbol{X}$) can be rewritten as

$$L_1(\mu,\sigma;Z) = \frac{n!}{(n-r)!}\prod_{j=1}^{r}\sigma^{-1}f(Z_{j:n})[1-F(Z_{r:n})]^{n-r}$$

and

$$L_2(Z_{s:n};\mu,\sigma,Z) = \sigma^{-1}\frac{(n-r)!}{(s-r-1)!(n-s)!} \times \frac{[F(Z_{s:n})-F(Z_{r:n})]^{s-r-1}}{[1-F(Z_{r:n})]^{n-r}} \times [1-F(Z_{s:n})]^{n-s}f(Z_{s:n}) \quad (5)$$

where $Z_{j:n}$=($X_{j:n}$−μ)/σ for j=1,...,n.

To derive the MMLP for Type II censored data when μ and σ are unknown, take logarithm of Eq. 5 and differentiate Eq. 5 with respect μ and σ in L_1, and $Z_{s:n}$ in L_2, the PLEs can be derived as follows:

$$\frac{\partial \ln L_1}{\partial\mu} = \frac{1}{\sigma}\left[-\sum_{j=1}^{r}\frac{f'(Z_{j:n})}{f(Z_{j:n})} + (n-r)\frac{f(Z_{s:n})}{1-F(Z_{s:n})}\right] = 0 \quad (6)$$

$$\frac{\partial \ln L_1}{\partial\sigma} = \frac{1}{\sigma}\left[-r-\sum_{j=1}^{r}\frac{f'(Z_{j:n})}{f(Z_{j:n})}Z_{j:n} + (n-r)\frac{f(Z_{r:n})\cdot Z_{r:n}}{1-F(Z_{r:n})}\right] = 0 \quad (7)$$

$$\frac{\partial \ln L_2}{\partial Z_{s:n}} = (s-r-1)\frac{f(Z_{s:n})}{F(Z_{s:n})-F(Z_{r:n})} + \frac{f'(Z_{s:n})}{f(Z_{s:n})} -(n-s)\frac{f(Z_{s:n})}{1-F(Z_{s:n})} = 0 \quad (8)$$

Rewriting Eqs. 6, 7 and 8, these equations can be expressed as

$$\frac{\partial \ln L_1}{\partial\mu} = \frac{1}{\sigma}\left[\sum_{j=1}^{r}\Psi(Z_{j:n}) + (n-r)\cdot h(Z_{r:n})\right] = 0 \quad (9)$$

$$\frac{\partial \ln L_1}{\partial\sigma} = \frac{1}{\sigma}\left[-r+\sum_{j=1}^{r}\Psi(Z_{j:n})\cdot Z_{j:n} + (n-r)\cdot h(Z_{r:n})\cdot Z_{r:n}\right] = 0 \quad (10)$$

$$\frac{\partial \ln L_2}{\partial Z_{s:n}} = (s-r-1)\cdot h_2(Z_{r:n},Z_{s:n}) - \Psi(Z_{s:n}) -(n-s)\cdot h(Z_{s:n}) = 0 \quad (11)$$

Equations 9–11 can be employed to produce the modified maximum likelihood estimators (MMLEs) of μ and σ and the MMLP of $X_{s:n}$. Because the distribution of the observations is assumed to be normal, the function Ψ(z) defined in Eq. 3 is equal to z. The following results can be obtained:

$\partial f(z)/\partial\mu = z\cdot f(z)/\sigma$ and $\partial f(z)/\partial\sigma = z^2\cdot f(z)/\sigma$.

Consequently, Eqs. 9–11 can be reduced, respectively, to

$$\sum_{j=1}^{r} Z_{j:n} + (n-r)\cdot h(Z_{r:n}) = 0, \quad (12)$$

$$-r+\sum_{j=1}^{r} Z_{j:n}^2 + (n-r)\cdot h(Z_{r:n})\cdot(Z_{r:n}) = 0, \quad (13)$$

$$(s-r-1)\cdot h_2(Z_{r:n},Z_{s:n}) - Z_{s:n} - (n-s)\cdot h(Z_{s:n}) = 0. \quad (14)$$

The first step of estimating μ and σ is to replace the hazard function appearing in Eqs. 12 and 13 by its expected value, then the PLEs in Eqs. 12 and 13 can be modified as

$$\sum_{j=1}^{r} Z_{j:n} + \sum_{j=r+1}^{n} E(Z_{j:n}) = 0, \quad (15)$$

$$-r+\sum_{j=1}^{r} Z_{j:n}^2 + Z_{r:n}\cdot\sum_{j=r+1}^{n} E(Z_{j:n}) = 0, \quad (16)$$

respectively. Solving Eqs. 15 and 16 to estimate μ and σ, the MMLE of μ and σ can be obtained as

$$\hat{\mu} = \bar{X}_r + \frac{\hat{\sigma}}{r}\sum_{j=r+1}^{n} E(Z_{j:n}), \quad (17)$$

$$\hat{\sigma} = \frac{a \cdot (X_{r:n} - \bar{X}_r) + \sqrt{\left[a \cdot (X_{r:n} - \bar{X}_r)\right]^2 + 4r \cdot \left[\sum_{j=1}^{r} (X_{j:n} - \bar{X}_r)^2\right]}}{2r}, \quad (18)$$

where $\bar{X}_r = (1/r)\sum_{j=1}^{r} X_{j:n}$ and $a = \sum_{j=r+1}^{n} E(Z_{j:n})$.

In the second step, $\hat{\mu}$ and $\hat{\sigma}$ are obtained using the MMLEs to substitute μ and σ, respectively, to obtain $Z_{s:n}$ in Eq. 14, and $h(Z_{s:n})$ and $h_2(Z_{r:n}, Z_{s:n})$ are replaced by their respective expected values. The PLE in Eq. 14 can then be changed as

$$E(Z_{s:n}) - Z_{s:n} = 0 \quad (19)$$

Solving the equation for predicting $X_{s:n}$, $\hat{X}_{s:n}$ can be derived as

$$\hat{X}_{s:n} = \hat{\mu} + \hat{\sigma} \cdot E(Z_{s:n}) \quad (20)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are given by Eqs. 17 and 18, respectively.

Note that $E(Z_{j:n})s$ are known and tabulated by Teichroew [15]. Although in Lemma 2.2.1, $r+1<s\leq n-1$ is assumed, Eq. 20 clearly holds even when $s=n$.

Next, the case where $s=r+1$ is considered. By replacing $h(Z_{s:n})$ with its expected value in Eq. 14, this equation can be reduced to:

$$-Z_{r+1:n} - \sum_{j=r+2}^{n} E(Z_{j:n}) = 0 \quad (21)$$

Solving this equation for predicting $X_{s:n}$, $\hat{X}_{s:n}$ can be obtained as

$$\hat{X}_{r+1:n} = \hat{\mu} - \hat{\sigma} \cdot \sum_{j=r+2}^{n} E(Z_{j:n}) \quad (22)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are given by Eqs. 17 and 18, respectively.

Because $E(Z_{j:n})s$ are monotonically increasing and $\sum_{j=1}^{n} E(Z_{j:n}) = 0$, $\sum_{j=k}^{n} E(Z_{j:n}) > 0$ for $k>1$; hence, the MMLP $\widehat{X}_{r+1:n}$ is less than $\overline{X}_r$ and $X_{r:n}$, for $s=r+1$, and it may be less than $X_{r:n}$ even for $s>r+1$. Thus, the MMLP of $X_{s:n}$ can be obtained as follows:

$$\hat{X}_{s:n} = \begin{cases} \hat{\mu} + \hat{\sigma} \cdot E(Z_{s:n}) & \text{if } \hat{X}_{s:n} > X_{r:n} \text{ and } r+1 < s \leq n, \\ X_{r:n} & \text{if } \hat{X}_{s:n} \leq X_{r:n} \text{ or } s = r+1, \end{cases} \quad (23)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are given by Eqs. 17 and 18, respectively.

4 Proposed procedure for analyzing censored data from factorial design and from Taguchi's parameter design

The main objective of experimental designs or Taguchi's parameter designs is to identify the important factors that influence the quality characteristics of a process or product and to select the optimal setting of factor-level combinations. Accordingly, this study proposes a procedure for optimizing Type II censoring experiments. The following sections first describe the statistical assumption of the Type II censored data, then present the proposed procedures of predicting Type II censored data from factorial designs and Taguchi's parameter designs.

4.1 Statistical assumption of the type II censored data

Normally, when the ANOVA method is employed to determine the optimal settings of factor-level combinations, the assumption of normality of the responses is required. Therefore, in this study, the Type II censored data are also assumed to be normally distributed. If the distribution of the response is not normal, then the transformations such as Box-Cox transformation, square root transformation, logarithmic transformation and arcsin transformation can be utilized to bring the distribution of the response closer to normal. Additionally, although the responses of Taguchi's parameter designs need not be normally distributed, the proposed procedure described in the following section requires that the responses of Taguchi's parameter designs be normally distributed. The goodness of fit test, such as chi-square test or Kolmogorov-Smirnov test [16], can be employed to verify whether the responses are normally distributed.

4.2 Proposed procedure for analyzing type II censored data from factorial designs

The steps of the procedure for analyzing Type II censored data from factorial designs are given as follows:

(1) Predict the censored data of every treatment in a factorial experiment using the MMLP derived in Sect. 3.
(2) Combine the predicted censored data and the uncensored data of each treatment as a pseudo-complete data set. Then, analyze the pseudo-complete data of all treatments using the ANOVA method.
(3) Determine the optimal settings of factor-level combinations.

Proposed procedure for analyzing type II censored data from Taguchi's parameter designs

The steps of applying the proposed procedure to analyze the Type II censored data from Taguchi's parameter designs are as follows.

Table 1 Factors and levels in the experiment

Label	Factor	Low level	High level
A	Lamination Temperature(°C)	55	75
B	Lamination Time (s)	10	25
C	Lamination Pressure (Ton)	5	10
D	Firing Temperature (°C)	1,580	1,620
E	Firing Cycle Time (h)	17.5	29
F	Firing Dew Point (°C)	20	26

(1) Use the MMLP derived in Sect. 3 to predict the censored data of every treatment.
(2) Merge the predicted censored data and the uncensored data of each treatment as a pseudo-complete data set, and compute the signal-to-noise (S/N) ratio from this data set. Use the ANOVA method to analyze factor effects and graph the factor effects to visually identify significant factors.
(3) Determine the optimal settings of factor-level combinations.

The following section gives numerical examples to illustrate the utility and effectiveness of the proposed method.

5 Numerical examples

5.1 Problem description

This example is adopted from Montgomery [17]. A 16-run experiment was conducted in a semiconductor manufacturing plant to investigate the effects of six factors on the curvature or camber of the substrate devices produced. Table 1 presents the six variables and their levels. Each run was replicated four times, and a camber measurement was made on the substrate. Table 2 presents all the relevant data. If the largest datum of each treatment is considered to be censored data, then the right Type II censored data are generated with n=4, r=3 and s=4. The Kolmogorov-Smirnov test is used to verify the normality assumption of the experimental data. The test result indicates that the experimental data is normally distributed.

5.2 Data analysis

To analyze the right Type II censored data described in Section 5.1, the MMLP shown in Eq. 23 is used to predict the censored data of each treatment. Because s=r+1, $X_{s:n}$ of each treatment is substituted by $X_{r:n}$. Table 3 presents the pseudo-complete data.

After the pseudo-complete data are obtained, the ANOVA method is used to determine the optimal setting of factor-level combinations. Table 4 shows the ANOVA result, clearly indicating a significant interaction between factors B and D because the p-value is less than α=0.05. Because other interactions of factors are insignificant (p-values are larger than α=0.05), Table 4 only shows the significant B and D interaction. Furthermore, the main

Table 2 Design and data for the experiment

Run	A	B	C	D	E	F	Camber for replicate			
							X_1	X_2	X_3	X_4
1	−1*	−1	−1	−1	−1	−1	167	128	149	185
2	1	−1	−1	−1	1	−1	62	66	44	20
3	−1	1	−1	−1	1	1	41	42	43	50
4	1	1	−1	−1	−1	1	73	81	39	30
5	−1	−1	1	−1	1	1	47	47	40	89
6	1	−1	1	−1	−1	1	219	258	147	296
7	−1	1	1	−1	−1	−1	121	90	92	86
8	1	1	1	−1	1	−1	191	186	162	106
9	−1	−1	−1	1	−1	1	32	23	77	69
10	1	−1	−1	1	1	1	78	158	60	45
11	−1	1	−1	1	1	−1	43	27	28	28
12	1	1	−1	1	−1	−1	186	137	159	158
13	−1	−1	1	1	1	−1	110	86	101	158
14	1	−1	1	1	−1	−1	65	109	126	71
15	−1	1	1	1	−1	1	155	158	145	145
16	1	1	1	1	1	1	93	124	110	133

*:1 and −1 are used to represent the high and low levels of each factor

Table 3 Censored data and MMLP

No.	$X_{1:4}$	$X_{2:4}$	$X_{3:4}$	$\hat{X}_{4:4}$
1	128	149	167	167
2	20	44	62	62
3	41	42	43	43
4	30	39	73	73
5	40	47	47	47
6	147	219	258	258
7	86	90	92	92
8	106	162	186	186
9	23	32	69	69
10	45	60	78	78
11	27	28	28	28
12	137	158	159	159
13	86	101	110	110
14	65	71	109	109
15	145	145	155	155
16	93	110	124	124

effects of factors A, C, and E are significant because their p-values are also less than α=0.05.

Figure 1 presents a graph of the average responses of factors B and D at every treatment combination. The significant interaction is expressed by the lack of parallelism of the lines. Figure 1 demonstrates that changing from low level of factor B to high level of factor B, average camber with high level of factor D decreases, whereas it increases for low level of factor D. Since the response is the smaller-the-better, the combination of low level of factor B and high level of factor D seems to give the best result.

Figures 2, 3 and 4 depict plots of significant factors A, C, and E. The main effect plots are graphs of the marginal response averages at the levels of the three factors. Since smaller responses are better, Figs. 2, 3 and 4 indicate that the optimal setting of factor-level combinations is A and C at their low levels and E at its high level. Hence, the optimal setting of factor-level combinations can be determined as $A\_B\_C\_D_+E_+F_+$ (minus and plus signs are used to denote the low and high levels of a factor, respectively).

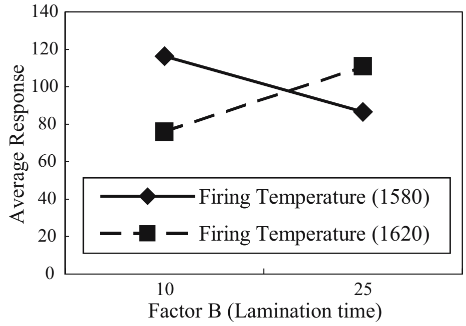

Fig. 1 Plot of factors B and D

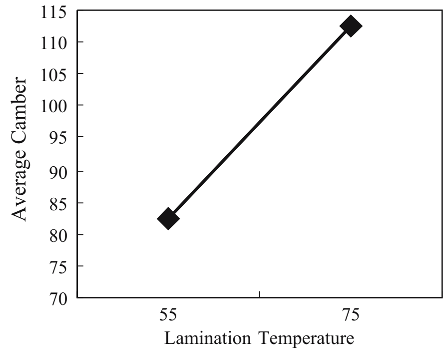

Fig. 2 Plot of factor A

5.3 Comparison of the result of analyzing censored data with complete data

Where the complete experimental data are analyzed, the significant factors are A, B, C, D, and E. The optimal setting of factor-level combinations is factors A, B, and C at their low levels and factors D and E at their high levels. This result is the same as that obtained by the proposed method.

5.4 Comparison of the result of analyzing censored data with another method

This section compares the results of analyzing censored data from the proposed method and the procedure proposed

Table 4 Analysis of variance for experimental data

Source of variation	Sum of squares	Degree of freedom	Mean square	F_0	P-value
A	14762.2500	1	14762.2500	7.66	0.0076
B	105.0625	1	105.0625	0.05	0.8162
C	35815.5625	1	35815.5625	18.59	<.0001
D	1024.0000	1	1024.0000	0.53	0.4691
E	31506.2500	1	31506.2500	16.35	0.0002
F	1722.2500	1	1722.2500	0.89	0.3485
BD	16835.0625	1	16835.0625	8.74	0.0046
Error	107909.3125	56	1926.9520		
Total	209679.7500	63			

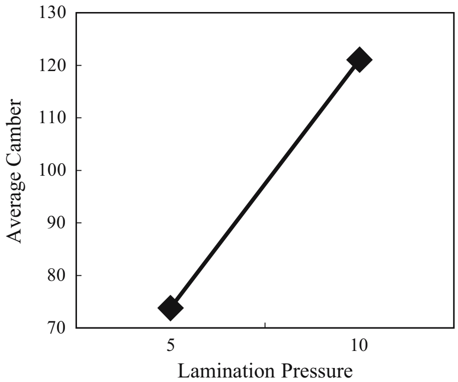

Fig. 3 Plot of factor C

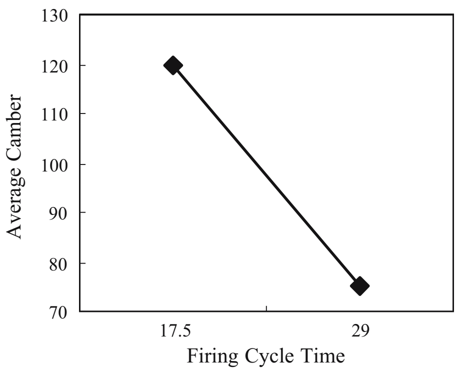

Fig. 4 Plot of factor E

Table 5 Factors and levels of each factor in the experiments

Label	Factor	Level		
A	Interference	Low	Medium	High
B	Connector wall thickness	Thin	Medium	Thick
C	Insertion depth	Shallow	Medium	Deep
D	Percent adhesive in connector pre-dip	Low	Medium	High

by Su and Miao [14] to illustrate the effectiveness of the proposed method. The optimal setting of factor-level combinations obtained by Su and Miao is $A_{+}B\_C_{+}D_{+}E\_F\_$. The levels of factors A, C and E are different from those of the proposed method. Moreover, the procedure proposed by Su and Miao cannot determine significant factors. Therefore, the proposed method is more effective.

5.5 Problem description

This example is adopted from Byrne and Taguchi [18]. An experiment was conducted to find a method to assemble economically an elastomeric connector to a nylon tube that would provide the requisite pull-off performance for use in automotive engine components. The objective of the experiment was to maximize the pull-off force.

Four control factors and three noise factors that could influence the assembly's pull-off force have been identified. Table 5 lists the four control factors and their levels. The three noise factors were conditioning time, conditioning temperature and conditioning relative humidity; these were not included in the experiments performed herein. Each run was replicated eight times, and the pull-off forces in pounds were recorded. Table 6 lists the complete experimental data of this experiment [18]. Suppose the largest two data of each treatment are censored data, the right Type II censored data are formed with n=8, r=6, s=7 and s=8. The Kolmogorov-Smirnov test is adopted to verify the normality assumption of the experimental data. The test result indicates that the experimental data is normally distributed.

5.6 Data analysis

According to the procedure described in Sect. 4, the MMLP shown in Eq. 23 is applied to predict the censored data of each treatment. Since the largest two data of each treatment are set to be censored, $X_{7:8}$ is substituted by $X_{6:8}$ because s=r+1 and $X_{8:8}$ is predicted using Eq. 20. After combining the uncensored data and the predicted censored

Table 6 The complete data of the experiment [18]

Run	A	B	C	D	Pull-off force for replicate								S/N ratio
					X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	
1	1*	1	1	1	15.6	9.5	16.9	19.9	19.6	19.6	20.0	19.1	24.045
2	1	2	2	2	15.0	16.2	19.4	19.6	19.7	19.8	24.2	21.9	25.522
3	1	3	3	3	16.3	16.7	19.1	15.6	22.6	18.2	23.3	20.4	25.335
4	2	1	2	3	18.3	17.4	18.9	18.6	21.0	18.9	23.2	24.7	25.904
5	2	2	3	1	19.7	18.6	19.4	25.1	25.6	21.4	27.5	25.3	26.908
6	2	3	1	2	16.2	16.3	20.0	19.8	14.7	19.6	22.5	24.7	25.326
7	3	1	3	2	16.4	19.1	18.4	23.6	16.8	18.6	24.3	21.6	25.711
8	3	2	1	3	14.2	15.6	15.1	16.8	17.8	19.6	23.2	24.4	24.832
9	3	3	2	1	16.1	19.9	19.3	17.3	23.1	22.7	22.6	28.6	26.152

*:1 represents the low level of a factor, 2 represents the medium level of a factor, and 3 represents the high level of a factor

Table 7 The pseudo-complete data obtained from MMLP and S/N ratio

Run	Pull-off force for replicate								S/N ratio
	$X_{1:8}$	$X_{2:8}$	$X_{3:8}$	$X_{4:8}$	$X_{5:8}$	$X_{6:8}$	$\hat{X}_{7:8}^{(2)}$	$\hat{X}_{8:8}^{(2)}$	
1	9.5	15.6	16.9	19.1	19.6	19.6	19.6	24.2	24.12
2	15	16.2	19.4	19.6	19.7	19.8	19.8	22.3	25.38
3	15.6	16.3	16.7	18.2	19.1	20.4	20.4	21.8	25.21
4	17.4	18.3	18.6	18.9	18.9	21	21	21.7	25.72
5	18.6	19.4	19.7	21.4	25.1	25.3	25.3	27.9	26.91
6	14.7	16.2	16.3	19.6	19.8	20	20	22.4	25.17
7	16.4	16.8	18.4	18.6	19.1	21.6	21.6	22.8	25.60
8	14.2	15.1	15.6	16.8	17.8	19.6	19.6	21.0	24.62
9	16.1	17.3	19.3	19.9	22.6	22.7	22.7	25.2	26.06

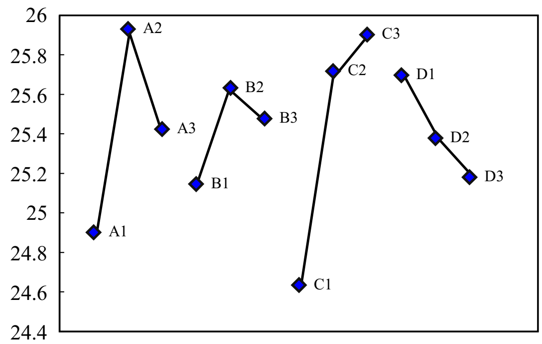

Fig. 5 Average S/N ratios for each level of the four controllable factors

data, the pseudo-complete data are shown in Table 7. The S/N ratios of each treatment are then calculated to analyze the data. The last column of Table 7 lists the S/N ratios.

Figure 5 plots the average S/N ratios for each level of the four controllable factors. Figure 5 indicates that the factors A and C are more significant than factors B and D. Larger average S/N ratios are better, so factor A is set at its intermedium level, and factor C is set at its high level. The optimal setting of factor-level combinations is determined as $A_2B_2C_3D_1$. The ANOVA method is utilized to test the significance of four factors. Table 8 lists the results, clearly showing that factors A and C are significant and factors B and D are not significant and that the optimal setting of factor-level combinations can be determined as $A_2B_2C_3D_1$.

5.7 Comparison of the result of analyzing censored data with complete data

When all of the complete data are analyzed, factors A and C are identified to be significant, and the optimal setting of factor-level combinations is determined to be $A_2B_2C_3D_1$. This result is the same as that obtained by the proposed method.

5.8 Comparison of the result of analyzing censored data with another method

This section compares the analytical results of the proposed method and the procedure of Su and Miao [14] with censored data described in Section 5.2 to illustrate the effectiveness of the proposed method. The optimal setting of factor-level combinations obtained by Su and Miao is $A_2B_2C_3D_1$. This result is the same as that obtained by the proposed method. However, the procedure proposed by Su and Miao must construct three neural networks, and selecting proper values of parameters in the networks to minimize the mean square error is difficult. By contrast, the proposed method is simpler to use.

Table 8 ANOVA table for S/N ratio

Source	Degree of freedom	SS	MS	F value	Pure sum of squares	Percent contribution (%)
A	2	1.5893	0.7947	4.0836	1.2001	23.18
B	2	0.3711*	–	–	–	–
C	2	2.8097	1.4049	7.2194	2.4205	46.75
D	2	0.4073*	–	–	–	–
Error	0	–	–	–	–	–
Pooling error	(4)	(0.7784)	(0.1946)		1.5568	30.07
Total	8	5.1774			5.1774	100.00

6 Conclusion

Many studies have proposed methods to predict the censored data from factorial experiments. However, these methods are computationally complex, or require complex statistical models or neural networks, and therefore have little practical use. Although the MLP method was derived to predict Type II censored data, a closed-form solution for the MLP may not be obtained in most cases. The MMLP under the normality assumption with unknown mean and known variance was derived by Raqab [5] to overcome the drawback of MLP. However, the population variance of experimental data in the real world is usually unknown in practice. The MMLP also has little practical use. Hence, assuming that both the true mean and variance are unknown for the experimental data, this study proposes a novel method. Engineers with little knowledge of statistics can adopt the proposed method to handle Type II censored data from repetitious experiments. This study first derived the MMLP, which assumes that the Type II censored samples are obtained from normal distribution with unknown mean and unknown standard deviation. Then the MMLP is employed to predict the censored data, and the uncensored data and predicted censored data are merged as the pseudo-complete data. The ANOVA method is employed to analyze the pseudo-complete data to determine the optimal settings of factor-level combinations. Moreover, this procedure can also be used to analyze the censored data from Taguchi's parameter designs. Two case studies are given to illustrate the proposed method. The proposed method is compared with existing methods to demonstrate the effectiveness of the proposed method. The results of the comparison demonstrate that the proposed method is more effective and simpler than the existing methods.

Notably, regardless of whether the proposed method is used to analyze the factorial experiments or the Taguchi experiments, the censored data of each treatment combination must include at least two observations.

References

1. Lawless JF (1982) Statistical models and methods for lifetime data. Wiley, New York
2. Miller RG (1981) Survival analysis. Wiley, New York
3. Bain LJ, Hardt ME (1991) Statistical analysis of reliability and life-testing model: theory and methods, 2nd edn. Marcel Dekker, New York
4. Chiou KC, Tong LI (2001) Average type II censoring times for two parameter Pareto and Rayleigh distributions. Int J Qual Reliab Manage 18(6):643–656
5. Raqab MZ (1997) Modified maximum likelihood predictors of future order statistics from normal samples. Comput Stat Data Anal 25:91–106
6. Kaminsky KS, Rhodin LS (1985) Maximum likelihood prediction. Ann Inst Stat Math 37:507–517
7. Nelson W, Hahn GJ (1972) Linear estimation of a regression relationship from censored data, part I—simple methods and their application. Technometrics 14:247–269
8. Nelson W, Hahn GJ (1973) Linear estimation of a regression relationship from censored data, part II—best linear unbiased estimation and theory. Technometrics 15:133–150
9. Hahn GJ, Nelson W (1974) A comparison of methods for analyzing censored life data to estimate relationships between stress and product life. IEEE Trans Reliab 23:2–11
10. Krall JM, Uthoff VA, Harley JB (1975) A step-up procedure for selection variables associated with survival. Biometrics 31:49–57
11. Schmee J, Hahn GJ (1979) A simple method for regression analysis with censored data. Technometrics 21:417–434
12. Hahn GJ, Morgan CB, Schmee J (1981) The analysis of a fractional factorial experiment with censored data using iterative least squares. Technometrics 23:33–36
13. Tong LI, Su CT (1997) A non-parametric method for experimental analysis with censored data. Int J Qual Reliab Manage 14(5):456–463
14. Su CT, Miao CL (1998) Neural network procedures for experimental analysis with censored data. Int J Qual Sci 3(3):239–253
15. Teichroew D (1956) Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution. Ann Math Stat 27:410–426
16. Johnson RA (2000) Probability and statistics for engineers, 6th edn. Prentice-Hall, Englewood Cliffs, NJ
17. Montgomery DC (2001) Design and analysis of experiments, 5th edn. Wiley, New York
18. Byrne DM, Taguchi S (1987) The Taguchi approach to parameter design. Quality Progress 19–26