THE COST OF USING INCOMPLETE
EXPONENTIAL DATA

M.S. Hamada

IIQP Research Report
RR-89-02

January 1989

THE COST OF USING INCOMPLETE EXPONENTIAL DATA

Michael Hamada

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario
N2L 3G1

ABSTRACT

In reliability studies, incomplete data whose values are not known exactly are frequently collected. This paper investigates the asymptotic and small sample costs of using several types of incomplete exponential data. Situations are identified where the information loss is substantial. Moreover, the small sample properties of the estimators are even worse than suggested by their asymptotic counterparts. These results provide guidance regarding the severity of the costs that can be incurred. This is especially helpful when it is possible to choose the type of incomplete data to be observed.

THE COSTS OF USING INCOMPLETE EXPONENTIAL DATA

Michael Hamada

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada N2L3G1
January 1989

Abstract

In reliability studies, incomplete data whose values are not known exactly are frequently collected. This paper investigates the asymptotic and small sample costs of using several types of incomplete exponential data. Situations are identified where the information loss is substantial. Moreover, the small sample properties of the estimators are even worse than suggested by their asymptotic counterparts. These results provide guidance regarding the severity of the costs that can be incurred. This is especially helpful when it is possible to choose the type of incomplete data to be observed.

1. Introduction

Before an experimenter performs a reliability study, he must choose the type of observation scheme he will use to collect the lifetime data. Data which are known exactly are called complete data and contain the most information for inference. Because complete data may be too expensive and sometimes are impossible to collect, incomplete data are frequently observed. These data arise for example when the experiment is stopped before all units have failed or when a unit cannot be monitored

1

continuously so that periodic inspection is required until the unit fails. There are costs incurred when using incomplete data in terms of information loss and realized in maximum likelihood estimators (MLE) with poorer small sample properties. The aim of this paper is to investigate these costs for different types of incomplete data.

We investigate four observation schemes: censoring, grouping, and two special cases of grouping, rounding and rounding-censoring, where the latter refers to inspecting at evenly spaced intervals up to a censoring point. The exponential distribution, one of the basic lifetime distributions, was successfully used by Davis (1952) to model several data sets. The focus of our study is the maximum likelihood estimator (MLE) of θ, the mean of the exponential distribution with density $f_\theta(t) = (1/\theta)\exp(-t/\theta)$.

This paper presents both the asymptotic and small sample costs of using several types of incomplete data by bringing together previous and new results. Situations are identified where the information loss is substantial through the use of asymptotic relative efficiencies (ARE's). Moreover, the small sample properties of the estimators are even worse than suggested by their asymptotic counterparts. These results are useful to the experimenter when he can choose the type of incomplete data that he collects. The expressions for some of the MLE's and ARE's are more natural than previous ones. Also, new interpretations are given to some of the ARE's. One contribution is a study of the small sample properties of the MLE's for censored, rounded, rounded-censored, and grouped data. Although formulas exist for small sample properties of MLE's from censored data (Mendenhall and Lehman 1960), they cannot be calculated by computer for sample sizes as small as 100. While there are asymptotic approximations for small sample properties of the MLE's from rounded and rounded-censored data (Kulldorff 1961), their use was recommended for sample sizes larger than 50. It is shown that certain criteria depend on the sample size and only on the degree of rounding, censoring, or the grouping boundaries relative to θ. Therefore, relatively few evaluations or simulations are needed to characterize the small sample behavior of the MLE's from these types of incomplete data.

Section 2 gives some notation and presents the estimators and their Fisher information for the different types of incomplete data. Section 3 focuses on the asymptotic

loss of information using asymptotic relative efficiency (ARE). Small sample properties from a simulation study are presented in Section 4. Section 5 concludes with a summary and discussion.

2. MLE's and Their Fisher Information

In this section we establish some notation and present the MLE's from complete and incomplete data and their Fisher information. Let T_1, ..., T_n be a random sample of size n from the exponential distribution with density $f_\theta(t) = (1 / \theta)\exp(-t / \theta)$.

For complete data, the T_i's are exactly known, and the MLE $\hat{\theta}$ is simply $\hat{\theta} = \overline{T}$. The Fisher information for $\hat{\theta}$ is

$$I(\hat{\theta}) = n / \theta^2 . \tag{2.1}$$

For singly Type I censored data with common censoring point L, the T_i's are either censored at L or exactly known. Letting r denote the number of exactly known observations, the MLE from censored data $\hat{\theta}_c = \sum T_i / r$. The Fisher information for $\hat{\theta}_c$ from Deemer and Votaw (1955) is

$$I(\hat{\theta}_c) = E(r) / \theta^2 = (n / \theta^2)(1 - \exp(-L / \theta)) . \tag{2.2}$$

We take rounded data to be recorded as $\{(2i-1)h / 2\}$, the midpoints of the intervals $\{[(i-1)h, ih)\}$ for $i = 1, 2, \dots$. The interval width h determines the degree of rounding. Equivalently, rounded data can be viewed as data from equally spaced inspections of length h in the form of intervals $\{[(i-1)h, ih)\}$. For rounded-censored data, we take the common censoring point L to be a multiple of h. The MLE from rounded-censored data is

3

$$\hat{\theta}_{rc} = h / \log((z+1)/(z-1)) \text{ , where } z = (2/h)(\textstyle\sum T_i / r) \text{ ,} \qquad (2.3)$$

where the T_i's are rounded or censored. With no censoring, the MLE from rounded data $\hat{\theta}_r$ is given in (2.3) with $z = (2/h)\overline{T}$. The Fisher information for $\hat{\theta}_{rc}$ is

$$I(\hat{\theta}_{rc}) = (E(r)/\theta^2)g(h/\theta) = (n/\theta^2)(1-\exp(-L/\theta))g(h/\theta) \text{ ,} \qquad (2.4)$$

where $g(a) = a^2\exp(-a)/(1-\exp(-a))^2$.

From (2.4), the Fisher information for $\hat{\theta}_r$ is

$$I(\hat{\theta}_r) = (n/\theta^2)g(h/\theta) \text{ .} \qquad (2.5)$$

McNeil (1966) considered rounding where $\{(i-1)h\}$ are recorded for the intervals $[0, h/2)$ and $\{[(2i-1)h/2, (2i+1)h/2)\}$ for i=1, 2, ... He showed that the Fisher information from this rounding scheme is larger than that studied here.

Grouped data are recorded as intervals $\{(x_{i-1}, x_i)\}$, where the interval boundaries partitioning $(0, \infty)$ into k groups are $x_0 = 0$, x_1, ..., $x_k = \infty$. Note that rounded data and rounded-censored data are special cases of grouped data. For arbitrary intervals, the MLE from grouped data $\hat{\theta}_g$ does not have a closed form and can be found iteratively via a Newton-Raphson procedure. The Fisher information for $\hat{\theta}_g$ is

$$I(\hat{\theta}_g) = (n/\theta^2)\textstyle\sum (1/p_i)((\exp(-x_i/\theta)(x_i/\theta) - \exp(-x_{i-1}/\theta)(x_{i-1}/\theta))^2 \text{ ,} \qquad (2.6)$$

where $\exp(-x_k/\theta)(x_k/\theta)$ is set to zero. Note that

$$I(\hat{\theta}_g) = (n/\theta^2)M(\{x_i/\theta\}) \text{ ,} \qquad (2.7)$$

where $M(\bullet)$ is a function of only $\{x_i/\theta\}$, the group boundaries relative to θ. Formulas (2.3)-(2.6) are equivalent to expressions found in Kulldorff (1961). However, (2.3)-(2.5)

4

are more natural for the setup given in this paper. When using the computer to evaluate $I(\hat{\theta}_g)$, (2.6) is preferred because it offers more stability.

3. Comparison of the ARE

The existence and the uniqueness of the MLE's were established in Kulldorff (1961). Hamada and Tse (1988) give results for the regression setting which can be applied to the univariate setting by using a single covariate vector of ones. Kulldorff (1961) also established the consistency and asymptotic efficiency of the MLE's. Therefore, the asymptotic relative efficiency (ARE), which is related to the increase in asymptotic variance incurred, is a meaningful quantity to compare the MLE's from the different data types. Let $\hat{\theta}$ and $\hat{\theta}_x$ denote the MLE's from complete and incomplete data, respectively. Then, the ARE for $\hat{\theta}_x$ (relative to $\hat{\theta}$) is $\mathrm{ARE}(\hat{\theta}_x)$ $= \mathrm{AsVar}(\hat{\theta}) / \mathrm{AsVar}(\hat{\theta}_x) = I(\hat{\theta}_x) / I(\hat{\theta})$, where $\mathrm{AsVar}(\cdot)$ and $I(\cdot)$ denote the asymptotic variance and Fisher information, respectively. To compare the MLE's from two incomplete data types, the ARE of $\hat{\theta}_x$ relative to $\hat{\theta}_y$ is denoted by $\mathrm{ARE}(\hat{\theta}_x, \hat{\theta}_y)$. We now use the results from the previous section to assess the asymptotic costs of using different incomplete data types.

3.1. The Effect of Censoring

From (2.1), (2.2), (2.4), and (2.5), the ARE for $\hat{\theta}_c$ and the ARE for $\hat{\theta}_{rc}$ relative to $\hat{\theta}_r$ are equal to

$$\mathrm{ARE}(\hat{\theta}_c) = \mathrm{ARE}(\hat{\theta}_{rc}, \hat{\theta}_r) = 1 - \exp(-L / \theta) . \tag{3.1}$$

Notice that $\mathrm{ARE}(\hat{\theta}_c) = P(T < L)$, so that if L is the 100αth percentile, $\mathrm{ARE}(\hat{\theta}_c) = \alpha$. $\mathrm{ARE}(\hat{\theta}_c)$ is a function of only L / θ. Table 3.1 presents the ARE's for several values of L / θ and shows that the information loss can be serious for censoring before the mean θ. These results also indicate that the extra costs for censoring in addition to

rounding can be substantial if the censoring is heavy.

Table 3.1 ARE($\hat{\theta}_c$), ARE($\hat{\theta}_{rc}$, $\hat{\theta}_r$)

Censored versus Complete, Rounded-Censored versus Censored

L / θ	.5	.75	1	1.5	2	2.5	3
ARE	.393	.528	.632	.777	.865	.912	.950

3.2. The Effect of Rounding

From (2.1), (2.2), (2.4), and (2.5), the ARE for $\hat{\theta}_r$ and the ARE for $\hat{\theta}_{rc}$ relative to $\hat{\theta}_c$ are equal to

$$\text{ARE}(\hat{\theta}_r) = \text{ARE}(\hat{\theta}_{rc}, \hat{\theta}_c) = g(h / \theta) \qquad (3.2)$$

for g(a) from (2.4). ARE($\hat{\theta}_r$) is directly related to the reduced variation in the random variable T_r obtained from rounding T: ARE($\hat{\theta}_r$) = Var(T_r)/Var(T). Note that (3.2) is a function of h / θ only. Table 3.2 presents the ARE's for $\hat{\theta}_r$ and the ARE's for $\hat{\theta}_{rc}$ relative to $\hat{\theta}_c$ for different values of h / θ. Note the small information loss for moderately heavy rounding. For example, there is only an 8% loss of information when the degree of rounding is equal to the mean θ. These results also indicate that there is little extra cost for rounding in addition to censoring even when the rounding is moderately heavy.

Table 3.2 ARE($\hat{\theta}_r$), ARE($\hat{\theta}_{rc}$, $\hat{\theta}_c$)

Rounded versus Complete, Rounded-Censored versus Censored

h / θ	.5	1	1.5	2	2.5	3
ARE	.979	.921	.832	.724	.609	.496

6

3.3. The Effect of Rounding and Censoring

Using (2.1) and (2.4), the ARE for $\hat{\theta}_{rc}$ is

$$\text{ARE}(\hat{\theta}_{rc}) = (1 - \exp(-L/\theta))g(h/\theta) \qquad (3.3)$$

for $g(a)$ from (2.4). Note that effect of both censoring and rounding is multiplicative. That is, $\text{ARE}(\hat{\theta}_{rc}) = \text{ARE}(\hat{\theta}_c)\,\text{ARE}(\hat{\theta}_r)$. $\text{ARE}(\hat{\theta}_{rc})$ is a function of L/θ and h/θ. Table 3.3 presents the ARE's for $\hat{\theta}_{rc}$ and the number of groups to the left of the censoring point for several combinations of L/θ and h/θ. The first column is $\text{ARE}(\hat{\theta}_c)$ and the last row is $\text{ARE}(\hat{\theta}_r)$. Table 3.3 shows that the information loss is small for moderately heavy rounding (equivalently, a small number of groups) provided the amount of censoring is not severe.

Table 3.3 $\text{ARE}(\hat{\theta}_{rc})$ and Number of Intervals Left of Censoring Point
Rounded-Censored versus Complete

L/θ	h/θ					
	0	.1	.5	1	2	4
.5	.393	.393	.385			
	∞	5	1			
1	.632	.631	.619	.582		
	∞	10	2	1		
2	.865	.864	.847	.797	.626	
	∞	20	4	2	1	
4	.982	.981	.961	.904	.711	.299
	∞	40	8	4	2	1
∞	1.000	.999	.979	.921	.724	.304
	∞	∞	∞	∞	∞	∞

3.4. The Effect of Grouping

From (2.1) and (2.7), the ARE for $\hat{\theta}_g$ is

$$ARE(\hat{\theta}_g) = M(\{x_i / \theta\}) . \qquad (3.4)$$

For finite number of groups k, x_{k-1} can be viewed as a common censoring point L. Using (2.2) and (2.7), the ARE for $\hat{\theta}_g$ relative to $\hat{\theta}_c$ for $x_{k-1} = L$ is

$$ARE(\hat{\theta}_g, \hat{\theta}_c) = M(\{x_i / \theta\}) / (1 - \exp(-L / \theta)) . \qquad (3.5)$$

From (2.4) and (2.7), the ARE for $\hat{\theta}_{rc}$ relative to $\hat{\theta}_g$ for $x_{k-1} = L$ is

$$ARE(\hat{\theta}_{rc}, \hat{\theta}_g) = M(\{x_i / \theta\}) / ((1 - \exp(-L / \theta))g(h / \theta)) \qquad (3.6)$$

for g(a) from (2.4).

We have already examined two special cases of grouping schemes, rounding and rounding-censoring. We next consider three schemes with a finite number of groups k. The ARE are studied for optimum unconstrained (OU), optimum equally spaced (OEQ), and equal probability (EP) designs. For the OEQ designs, the widths of the first k – 1 intervals are equal. For the EP designs, an observation has equal probability of being in any of the k groups. Hughs (1949) and Kulldorff (1961) gave solutions for the OU and OEQ group boundaries, respectively. From (3.4), only the interval boundaries relative to θ need be considered. Let a_i denote x_i / θ.

The ARE's and the a_i's for the OU, OEQ, and EP designs when k = 2, 3 are presented in Table 3.4. The optimal interval designs have similar ARE's, whereas that of the equal probability interval designs are somewhat lower. For a larger number of groups, say 10 or more, the superiority of the optimal designs is negligible.

Table 3.4 $\text{ARE}(\hat{\theta}_g)$ for Some Optimal Group and
Equal Probability Designs: Grouped versus Complete

Design	a_i		$\text{ARE}(\hat{\theta}_g)$
EP(k=2)	0.693		.480
OU(k=2)	1.594		.648
EP(k=3)	0.405	1.100	.650
OEQ(k=3)	1.207	2.414	.808
OU(k=3)	1.018	2.611	.820

The ARE of the optimal designs are surprisingly high for even a small number of groups. For example, the ARE for the OU(k=2) design is .648 when only the number of observations before the censoring point $L = x_1$ is recorded. Bartholomew (1963) observed that there is little additional information in knowing the uncensored lifetimes when the censoring is heavy. Table 3.5 demonstrates this where .288, .693, and 1.386 correspond to the exponential quartiles.

Table 3.5 $\text{ARE}(\hat{\theta}_g, \hat{\theta}_c)$ for k = 2
Grouped versus Censored

a_1	.288	.5	.693	1	1.386	2	4
ARE	.993	.979	.961	.921	.854	.724	.304

In practice the optimum intervals can be used if there is a priori knowledge about θ. Table 3.6 presents the ARE for several guessed values of θ under the OU, OEQ, and EP designs for k = 3.

9

Table 3.6 ARE($\hat{\theta}_g$) for Some Group Designs for Guessed Values of θ
Optimum Unconstrained, Optimum Equally Spaced, Equal Probability

Guessed Value of θ	Optimum Unconstrained	Optimum Equally Spaced	Equal Probability
0.25θ	.475	.450	.240
0.50θ	.704	.680	.420
1.00θ	.820	.808	.650
1.50θ	.772	.746	.764
1.75θ	.728	.689	.794
2.00θ	.680	.624	.811
2.50θ	.575	.491	.819
3.00θ	.471	.370	.806

There is a small decrease in information for guessed values larger but near θ. The two optimal designs behave similarly for values less than θ, whereas the OU design performs better for values greater than θ. The EP design behaves poorly for values less than θ, but behaves exceptionally well for values between 1.5θ and 3θ; it even outperforms the optimal designs (for true θ) in the latter region. Kulldorff (1961) also considered the two optimal designs for larger k and observed that the effect of making a wrong guess is much smaller. In practice, this suggests using a larger k and equally spaced intervals for convenience. This is further confirmed by observing the performance of $\hat{\theta}_r$ in Table 3.2. The results for the EP design suggest that these designs are also useful for a liberal guess of θ.

Optimal constrained and equal probability constrained designs, where $x_{k-1} = L$, were studied in Hamada (1987). It was shown that these designs gave practically the same ARE as the rounding-censoring schemes studied here for three or more groups.

10

4. Comparison of Small Sample Properties

In this section small sample properties of the MLE's from rounded, censored, rounded-censored, and grouped data are investigated. While assessment of interval estimation properties may be preferable, point estimation properties are simpler to calculate and can be directly compared with the ARE. Although the MLE's are asymptotically unbiased and efficient, their small sample properties may be quite different. The MLE's can be seriously biased or its variance may be much larger than its asymptotic variance. Only the MLE from complete data is unbiased. The criteria we focus on are relative bias (RBIAS $=$ bias$(\hat{\theta}_x)/\theta$) and relative MSE (RELMSE $=$ MSE$(\hat{\theta})/$MSE$(\hat{\theta}_x)$). Small negative RBIAS is desired, since negative bias gives a conservative estimate of reliability. RELMSE is compared with its asymptotic counterpart ARE to determine if the small sample costs are significantly larger than that suggested by the ARE.

For the exponential distribution with mean θ, RBIAS and RELMSE for the MLE's from incomplete data depend on n and θ only through the quantities L $/$ θ for $\hat{\theta}_c$, h $/$ θ for $\hat{\theta}_r$, L $/$ θ and h $/$ θ for $\hat{\theta}_{rc}$, and $\{x_i/\theta\}$ for $\hat{\theta}_g$. This follows from the exponential distribution being a scale family and the MLE's being scale equivariant. Therefore, these criteria share a property which implies that few calculations or simulations are required to characterize the behavior of the MLE's from incomplete data.

A simulation study was performed to evaluate RBIAS and RELMSE for the MLE's from censored, rounded, rounded-censored, and grouped data. 5000 samples of size 25, 50, and 100 from an exponential distribution were drawn using the IMSL subroutine GGEXP. Note that for each simulation, a sample of size 100 was drawn first. Then the first 25 and first 50 observations of the 100 were the samples of size 25 and 50, respectively. The results for the MLE from censored data for n $=$ 25 and n $=$ 50 were calculated using the formulas given in Mendenhall and Lehman (1960). The results for n $=$ 100 are from a simulation since the the formulas could not be calculated by the computer for this sample size. While asymptotic approximations exist for small sample properties of the MLE's from rounded and rounded-censored data (Kulldorff 1961), their use was recommended for sample sizes larger than 50.

Graphs of RBIAS and RELMSE for the MLE's from the incomplete data types appear in Figures 4.1-4.5. A summary for each incomplete data type is given next with an emphasis on the criteria behavior for small sample size (n = 25).

4.1. Results for Censoring

The exact and simulation results for $\hat{\theta}_c$ are displayed in Figures 4.1 for $L / \theta = .5$, .75, 1(.5)3. The MLE $\hat{\theta}_c$ is overbiased. For small sample size, the bias is substantial when censoring is heavy (small L / θ). However bias is minimal for heavy censoring if the sample size is increased. For light censoring, the bias is negligible for small sample size so that little is gained by increasing the sample size. A similar pattern can be seen for RELMSE. For small sample size and heavy censoring, RELMSE is substantially smaller than ARE; the main reason is that $\text{Var}\,\hat{\theta}_c$ is much larger than $\text{Var}\,\hat{\theta}$.

4.2. Results for Rounding

Simulation results for $\hat{\theta}_r$ are presented in Figures 4.2 for $h / \theta = .5(.5)3$. The MLE $\hat{\theta}_r$ is underbiased. For small sample size, the bias is substantial when rounding is heavy (large h / θ). However, the bias is small for heavy rounding if the sample size is substantially increased. For light rounding, the bias is negligible for even small sample size. A similar pattern can be seen for RELMSE. However for heavy rounding, RELMSE is substantially smaller than ARE even when the sample size is large.

4.3. Results for Rounding-Censoring

Simulation results for $\hat{\theta}_{rc}$ are displayed in Figures 4.3, where $L / \theta = 2$ and $h / \theta = 0$, .125, .25, .5, 1, and 2. These values of h / θ correspond to the following number of intervals left of L: ∞, 16, 8, 4, 2, and 1, respectively. The bias is negligible even for small sample size and heavy rounding. The MLE $\hat{\theta}_{rc}$ is overbiased for $L / \theta = 2$. For lighter censoring, say $L / \theta = 4$, $\hat{\theta}_{rc}$ is underbiased for heavy rounding. Recall that $\hat{\theta}_r$ is underbiased whereas $\hat{\theta}_c$ is overbiased, so that the sign of the bias of $\hat{\theta}_{rc}$ depends on

the degree of rounding and censoring. The RELMSE are close to the ARE even for small sample size.

4.4. Results for Grouping

Simulation results for $\hat{\theta}_g$ based on the OU, OEQ, and EP designs are presented in Figures 4.4. The MLE $\hat{\theta}_g$ is overbiased for these designs, although the bias is negligible for even small sample size. The two optimal designs perform better and are less sensitive to the sample size. In fact, $\hat{\theta}_g$ performs better for OU(2) than EP(3); more intervals do not necessarily lead to better results. Similar patterns can be seen for RELMSE. The EP designs perform poorly for small sample size where the RELMSE is much smaller than the ARE.

Simulation results for $\hat{\theta}_g$ based on the OU(3), OEQ(3), and EP(3) designs for different guessed values of θ are presented in Figures 4.5 (n = 25). For guessed $\theta^* > 2\theta$, the EP design outperforms the optimal designs. Whereas the overbias for the optimal designs is substantial, the EP design has negligible overbias. Furthermore, the probability of the MLE not existing for the optimal designs in this region is substantial. However, for guessed $\theta^* < \theta$, the EP design performs poorly with significant overbias and substantially smaller RELMSE.

5. Summary and Discussion

The MLE's of the mean of the exponential distribution from censored, rounded, rounded-censored, and grouped data have been studied. Intuitively, these data contain less information than complete data. The loss of asymptotic information is minimal in many situations. Surprisingly, it can be quite small for coarse grouping if the interval boundaries are chosen optimally. In practice, the optimal designs are not useful unless good a priori knowledge about the mean exists. The equal probability designs perform surprisingly well for a liberal guess of the mean. However, there is a significant asymptotic loss of information for heavy rounding, heavy censoring, and coarse grouping (far from optimal grouping).

13

Small sample properties should also be considered. While assessment of interval estimation properties may be preferable, point estimation properties are simpler to calculate and can be directly compared with the ARE. Consequently, this paper focused on bias and MSE. The bias is substantial for the same situations where there is a significant loss of information. Furthermore, the loss of information as measured by the increase in MSE is much larger than its asymptotic counterpart ARE suggests. For these extreme situations, the probability that the MLE does not exist can also be substantial. Increasing the sample size can offer some improvement, but the small sample properties can remain poor.

Results of this study have practical implications.

(1) Censoring too soon produces overbiased estimators with inflated variances which result in liberal reliability estimates.

(2) A measuring device with low precision or rounding data too heavily yields underbiased estimators with inflated variances which result in conservative reliability estimates. However, the costs for moderately heavy rounding as large as the mean are minimal.

(3) Grouping data too coarsely and far from optimal grouping leads to estimators with poor small sample properties as above.

(4) Given that one is resigned to do some censoring, the additional costs of making a relatively few inspections (resulting in grouped data) is negligible.

Good a priori information about the mean is important in designing an experiment when incomplete data will be taken. Careful consideration must be made about what type of data should be taken. The time or cost savings of using incomplete data must be weighed against the bias, increased variance, and possible non-existence of their estimators, for it is too late after the experiment is finished.

The exponential distribution is a special case of the Weibull distribution. The results above can be applied to the Weibull distribution with known shape parameter since a power transformation of exponential data using the shape parameter yields Weibull data. For unknown shape parameter, Meeker (1986) has studied the efficiency of several grouping designs and Type I censoring for estimating a given quantile.

14

Ostrouchov and Meeker (1987) study the small sample properties of the equal probability designs. Comparison of information loss and small sample properties of MLE's from other types of incomplete data is a topic for future study.

Acknowledgements

References

Bartholomew, D. J. (1963). The sampling distribution of an estimate arising in life testing. *Technometrics, 5*, 361-374.

Davis, D. J. (1952). An analysis of some failure data. *Journal of the American Statistical Society, 47*, 113-150.

Deemer, W. L. and D. F. Votaw (1955). Estimation of parameters of truncated or censored exponential distributions. *Annals of Mathematical Statistics, 26*, 498-504.

Hamada, M. (1987). Studies on incomplete and ordered categorical data from industrial experiments. Unpublished Ph.D. thesis, University of Wisconsin-Madison.

Hamada, M. and S. K. Tse (1988). A note on the existence of maximum likelihood estimates in linear regression models using interval-censored data. *Journal of the Royal Statistical Society,* Series B, 50, 293-296.

Hughs, H. M. (1949). Estimation of the variance of the bivariate normal distribution. *University of California Publications in Statistics,* 1, 37-51.

Kulldorff, G. (1961). *Estimation from Grouped and Partially Grouped Samples.* John Wiley, New York.

McNeil, D. R. (1966). Consistent statistics for estimating and testing hypotheses from grouped samples. *Biometrika,* 53, 545-557.

Meeker, W. Q. (1986). Planning life tests in which units are periodically inspected for

failure. *IEEE Transactions on Reliability,* R-35, 571-578.

Mendenhall, W. and E. H. Lehman (1960). An approximation to the negative moments of the positive binomial useful in life-testing. *Technometrics, 2,* 227-242.

Ostrouchov, G. and W. Q. Meeker (1987). Accuracy of approximate confidence bounds computed from interval censored weibull data. Submitted for publication.
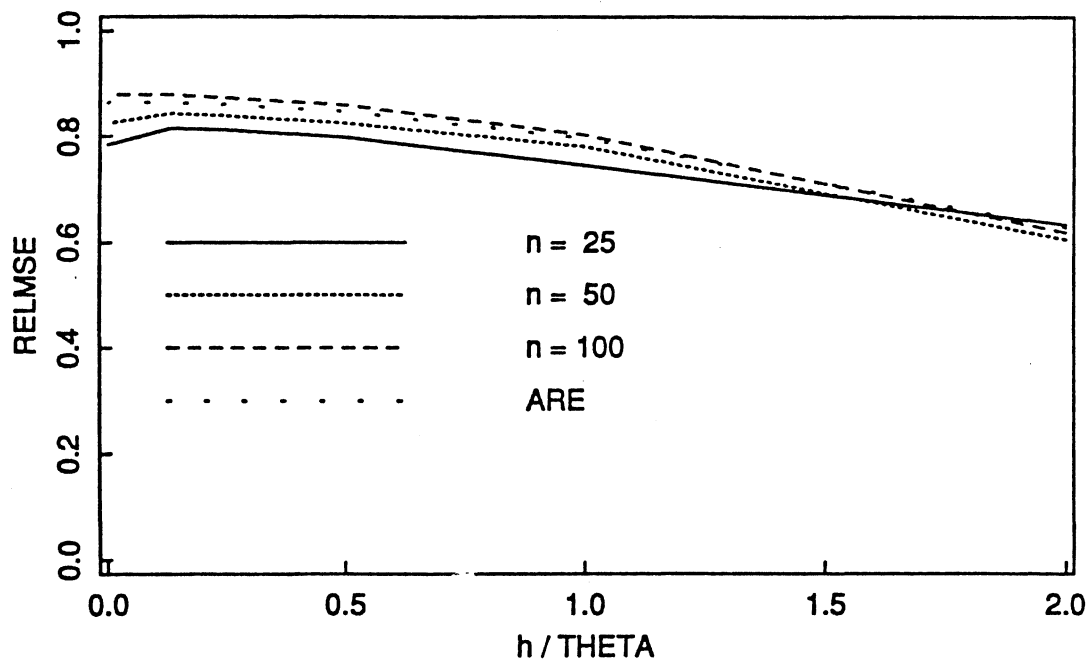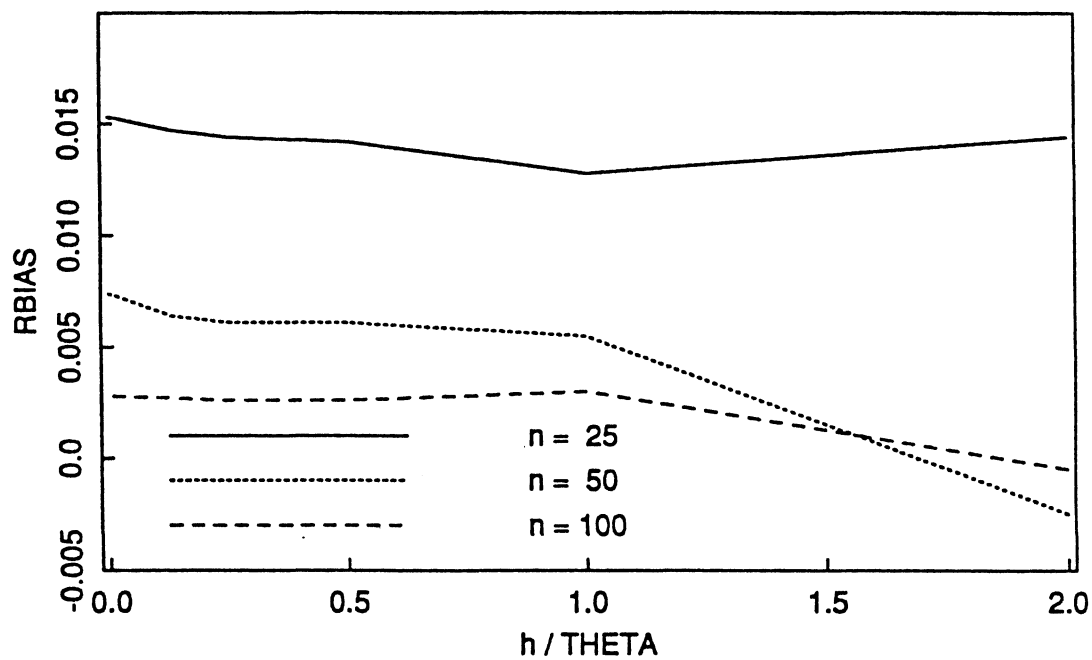
FIGURES 4.1 CENSORED DATA SIMULATIONS
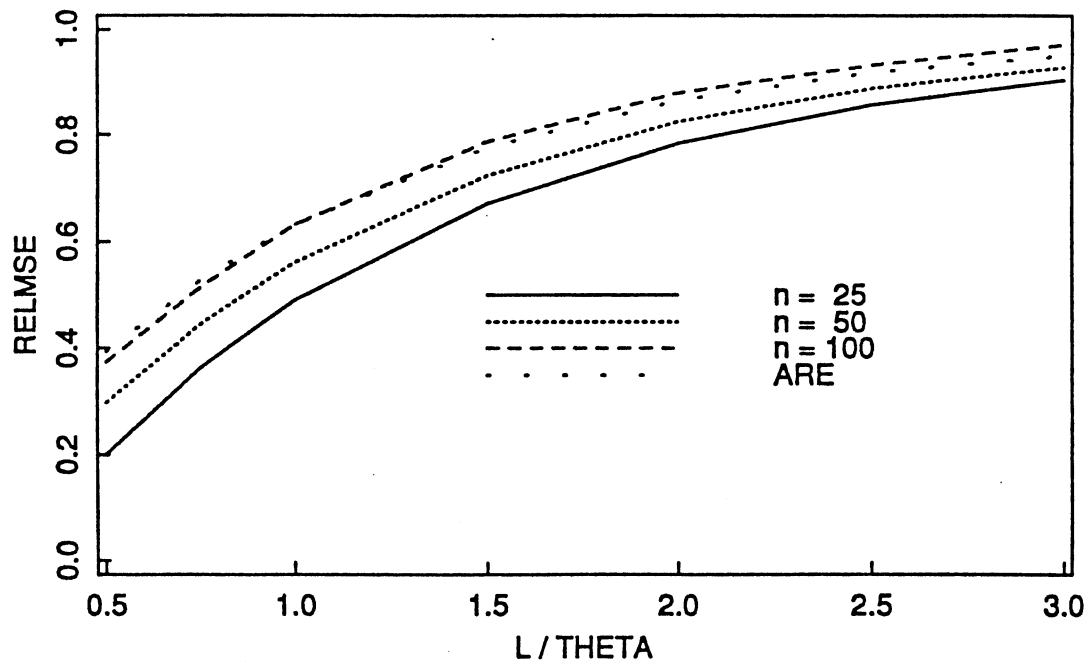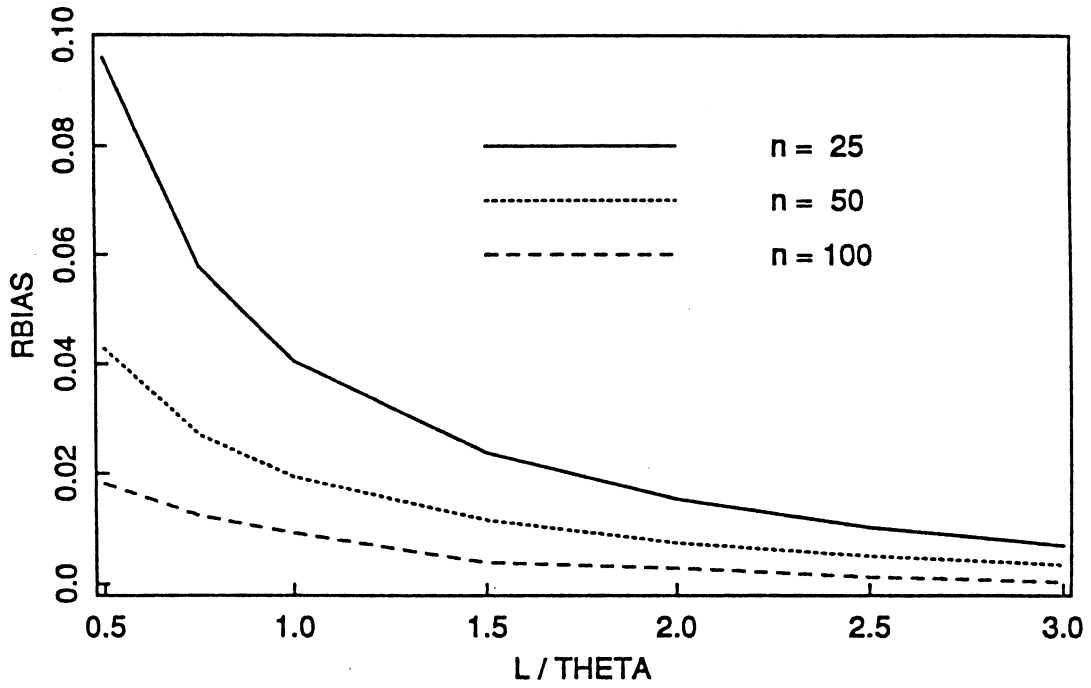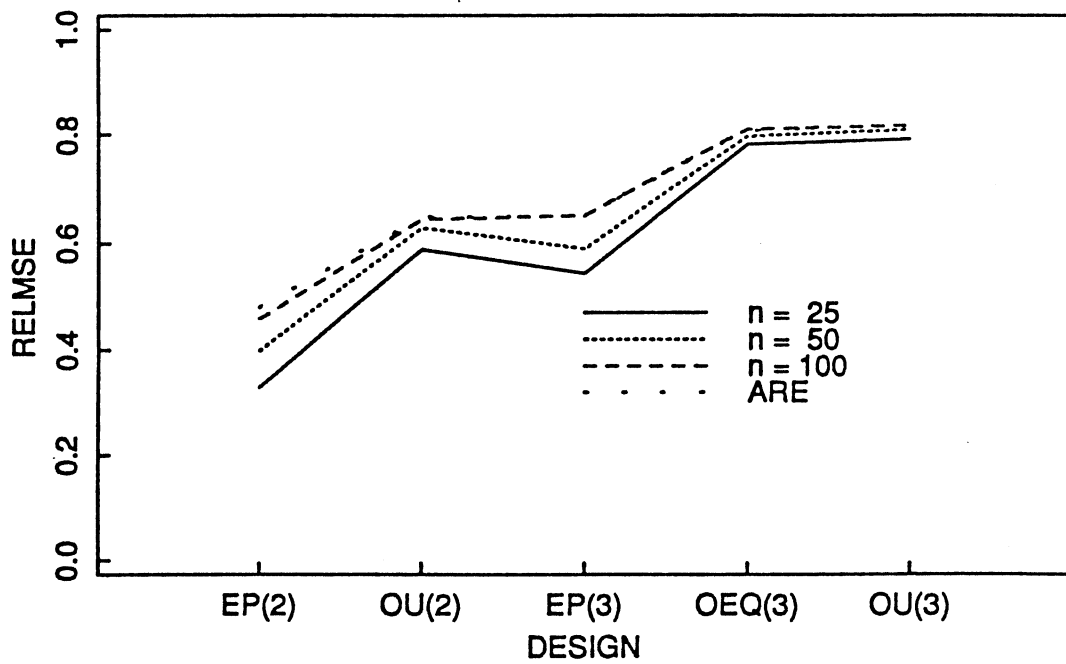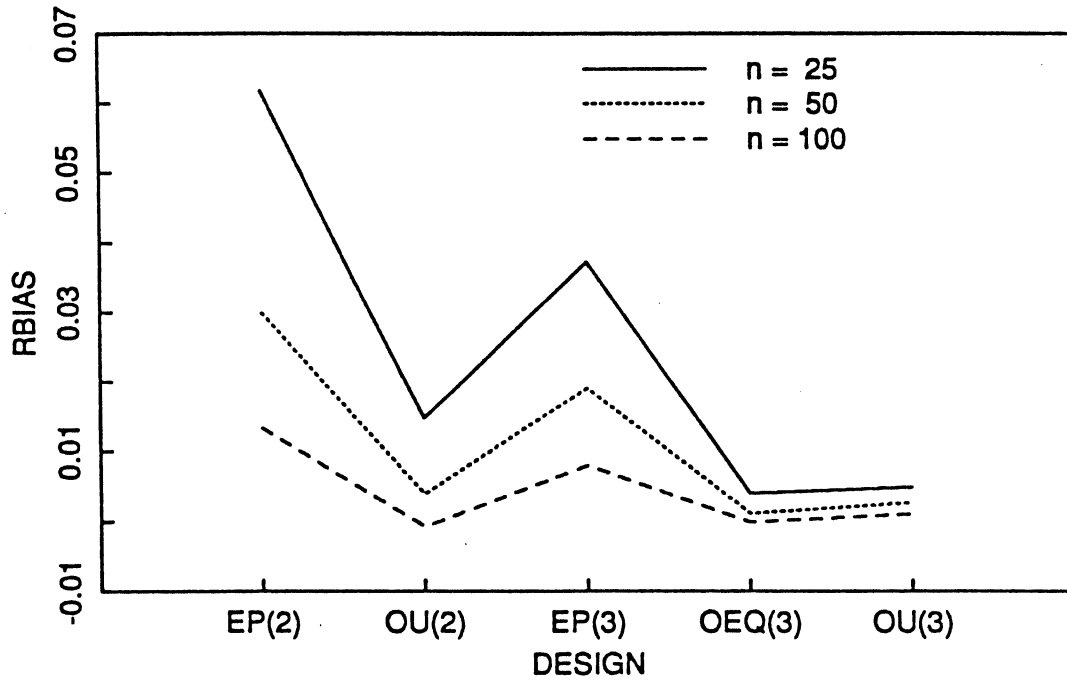
FIGURES 4.2 ROUNDED DATA SIMULATIONS
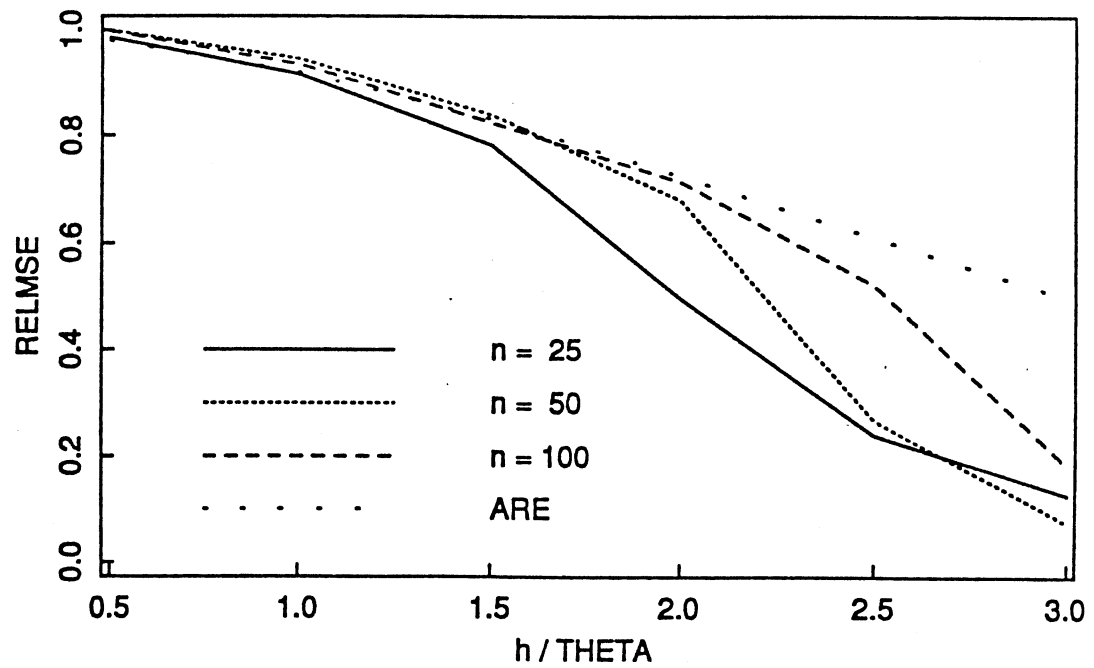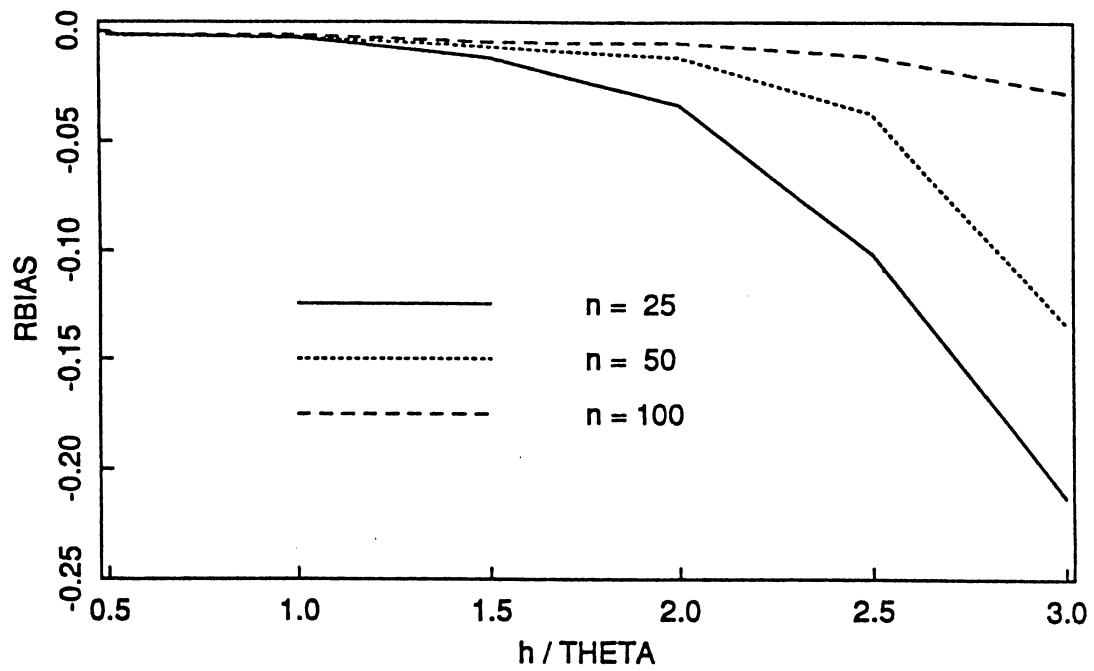
FIGURES 4.3 ROUNDED-CENSORED DATA SIMULATIONS

(L/THETA=2)

FIGURES 4.4 GROUPED DATA SIMULATIONS

FIGURES 4.5 GROUPED DATA SIMULATIONS (n=25)