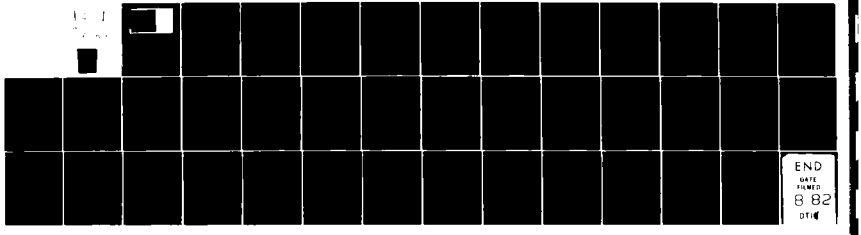
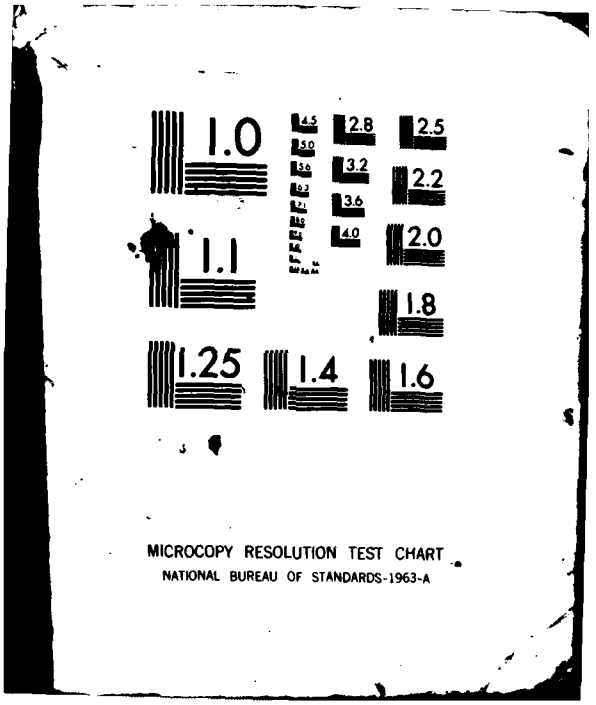
AD-A116 189 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
A CASE STUDY OF THE ROBUSTNESS OF BAYESIAN METHODS OF INFERENCE--ETC(U)
MAY 82 D B RUBIN DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2375 NL



END
DATE FILMED
8 82
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MRC Technical Summary Report #2375

A CASE STUDY OF THE ROBUSTNESS OF BAYESIAN METHODS OF INFERENCE: ESTIMATING THE TOTAL IN A FINITE POPULATION USING TRANSFORMATIONS TO NORMALITY

Donald B. Rubin

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

May 1982

(Received October 16, 1981)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

A CASE STUDY OF THE ROBUSTNESS OF BAYESIAN METHODS OF INFERENCE: ESTIMATING THE TOTAL IN A FINITE POPULATION USING TRANSFORMATIONS TO NORMALITY

Donald B. Rubin

Technical Summary Report #2375
May 1982

ABSTRACT

Bayesian methods of inference are the appropriate statistical tools for providing interval estimates in practice. The example presented here illustrates the relative ease with which Bayesian models can be implemented using simulation techniques to approximate posterior distributions but also shows that these techniques cannot be automatically applied to arrive at sound inferences. In particular, the example dramatizes three important messages. The first two messages are concrete and easily stated:

(1) Although the log normal model is often used to estimate the total on the raw scale (e.g., estimate total oil reserves assuming the logarithm of the values are normally distributed), the log normal model may not provide realistic inferences even when it appears to fit fairly well as judged from probability plots.

(2) Extending the log normal family to a larger family, such as the Box-Cox family of power transformations, and selecting a better fitting model by likelihood criteria or probability plots, may lead to less realistic inferences for the population total, even when probability plots indicate an adequate fit.

The third message is more philosophical, is not easy to state precisely, but is well-illustrated by the example.

(3) In general, inferences are sensitive to features of the underlying distribution of values in the population that cannot be addressed by the data. Consequently, for good statistical answers we need: (a) models that allow observed data to dominate prior restrictions, and either (b) flexibility in these models to allow specification of realistic underlying features of population values not adequately addressed by observed values, or (c) questions that are robust for the type of data collected in the sense that all relevant underlying features of population values are adequately addressed by the observed data.

AMS (MOS) Subject Classifications: 6207, 62A15, 62D05, 62E25, 62F15

Key Words: sample surveys, Box-Cox transformations, simulation, sensitivity

Work Unit Number 4 (Statistics and Probability)

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

A CASE STUDY OF THE ROBUSTNESS OF BAYESIAN METHODS OF INFERENCE: ESTIMATING THE TOTAL IN A FINITE POPULATION USING TRANSFORMATIONS TO NORMALITY

Donald B. Rubin

1. PROLOGUE-THE PRACTICAL INTERPRETATION OF INTERVAL ESTIMATES AS BAYES INTERVALS

Bayesian methods of inference will be, I believe, the primary statistical tools used to analyze data in the future, at least in those cases in which the purpose of statistical analysis is to provide a range of likely values for an unknown quantity, such as the total in a finite population or the relative effect of a treatment in an experiment. One reason for this belief is the inherent flexibility of Bayesian models with their multiple levels of randomness; such methods naturally lead to smoothed estimates in complicated data structures and consequently possess the ability to obtain better real world answers.

Another reason for this belief that Bayesian methods will constitute the standard tools for providing interval estimates is more psychological, and involves the relationship between the statistician and the client who is the consumer of the statistician's work. In nearly all practical cases, clients will interpret intervals provided by statisticians as Bayesian intervals, that is, as probability statements about the likely values of unknown quantities conditional on the evidence in the data. Such direct probability statements require prior probability specifications for unknown quantities, and thus the kinds of answers clients will assume are being provided by

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

statisticians, Bayesian answers, require prior probability assumptions. If the Bayesian answers vary dramatically for different reasonable assumptions unassailable by the data, then the resultant range of Bayesian answers must be entertained as legitimate, and I believe that the statistician has the responsibility to make the client aware of this fact.

Of course, there are assumptionless confidence intervals, but these are not generally useful inferentially. For an extreme example, consider the following 95% confidence interval: regardless of the values of the data, 95% of the time the interval is $(-\infty,\infty)$ and 5% of the time the interval is $[0,0]$. Confidence intervals are generally useful and fair summaries of data only when they can be interpreted as approximate (or, in some circumstances, conservative) Bayesian intervals.

In brief, interval estimates will be interpreted by clients as Bayesian (or approximately Bayesian) intervals and therefore statisticians have an obligation to try to provide interval estimates that can legitimately be interpreted as such, or at least to offer guidance as to when the intervals that are provided can be safely interpreted in this manner.

2. THE ROBUSTNESS OF BAYESIAN METHODS

The potential application of statistical methods is often demonstrated either (a) theoretically, (b) from artificial data generated following some convenient analytic form, or (c) from real data without a known correct answer. But quite generally, we understand tools through the consequences of their application, and these three kinds of demonstrations, although useful, provide somewhat limited evidence on how well the tools can be expected to work in practice. The case study presented here uses a small, real data set with a known value for the quantity to be estimated. It is surprising and instructive to see the care that may be needed to arrive at satisfactory inferences with real data.

The specific example concerns the estimation of the total population of the $N = 804$ municipalities in New York State from a simple random sample of $n = 100$ (source = Encyclopedia Britannica, 1960 census; New York City was represented by its five boroughs). Table 1 presents summary statistics for this population and two simple random samples. These two samples were the first and only ones chosen. With knowledge of the population, neither sample appears particularly atypical; sample 1 is very representative of the population, whereas sample 2 has a few too many large values. Consequently, it might at first glance seem straightforward to estimate the population total, perhaps overestimating the total from the second sample.

This example was originally studied to demonstrate the relative ease with which Bayesian models could be fit to such data using simulation techniques to approximate posterior distributions, and the example does illustrate this point. It does not, however, generate the message that these techniques can be automatically applied to arrive at sound inferences. Rather, it dramatizes three important messages.

The first two messages are concrete and address the accuracy of resultant inferences for covering the true population total.

(1) Although the log normal model is often used to estimate the total on the raw scale (e.g., estimate total pollutant, medical costs or oil reserves assuming the logarithm of the values are normally distributed), the log normal model may not provide accurate inferences for the total even when it appears to fit fairly well as judged from probability plots.

(2) Extending the log normal family to a larger family, such as the Box-Cox family of power transformations, and selecting a better fitting model by Bayesian/likelihood criteria or probability plots may lead to less realistic inferences for the population total, even when probability plots indicate an adequate fit.

TABLE 1: Summary Statistics for Populations of Municipalities in New York State; All 804 and Two Simple Random Samples of 100 (Source: Encyclopedia Brittanica - 1960 Census Figures)

	Population N = 804	Sample 1 N = 100	Sample 2 N = 100
Total	13,776,663	1,966,745	3,850,502
Mean	17,135	19,667	38,505
Std. Dev.	139,147	142,218	228,625
Low	19	164	162
5%	336	308.5	315
25%	800	891.5	863
Med.	1,668.5	2,081.5	1,740
75%	5,050	6,049.5	5,239
95%	30,295	25,130	41,718
High	2,627,319	1,424,815	1,809,578

These two points are not criticisms of the log transformation or the Box-Cox family of power transformations. Rather, they are warnings about the naive statement "better fits to data mean better models which in turn mean better real world answers". Statistical answers rely on prior assumptions as well as data, and better real world answers generally require models that incorporate more realistic prior assumptions as well as provide better fits to data. This comment naturally leads to the last message of this paper, which is a general one encompassing the first two.

(3) In general, inferences are sensitive to features of the underlying distribution of values in the population that cannot be addressed by the observed data.

Consequently, for good statistical answers we need

(a) models that allow observed data to dominate prior restrictions,

and either

(b) flexibility in these models to allow specification of realistic underlying features of population values not adequately addressed by observed values, such as behavior in the extreme tails of the distribution,

or

(c) questions that are robust for the type of data collected in the sense that all relevant underlying features of population values are adequately addressed by the observed values.

Finding models that satisfy 3a and 3b is a more general approach than finding questions that satisfy 3c because statisticians are often presented with hard questions that require answers of some sort, and do not have the luxury of posing easy (i.e. robust) questions in their place. For example, for environmental reasons it may be important to estimate the total amount of pollutant being emitted by a manufacturing plant using samples of the soil from the surrounding geographical area, or, for purposes of budgeting a health-care insurance program, it may be necessary to

estimate the total amount of medical expenses from a sample of patients. Such questions are inherently nonrobust in that their answers depend on the behavior in the extreme tails of the underlying distributions. Estimating more robust population characteristics, such as the median amount of pollutant in soil samples or the median medical expense for patients, does not address the essential questions in such examples.

At least from a Bayesian perspective, the more major effort in statistics currently seems to be focused on 3c rather than on 3a and 3b, that is on defining the estimand to be the midmean or the population analogue of some other robust estimator of location. Although such work is obviously important, it seems somewhat surprising that less effort is being devoted to the development of computationally attractive tools that are capable of addressing both easy and hard questions, especially since the current collection of statistical tools satisfying both criteria 3a and 3b seems to be rather limited.

This third point is not a criticism of any particular tool for inference, but it is a criticism of the claim that inferential tools, such as the jackknife (c.f. Miller, 1974) or bootstrap (Efron, 1980, Rubin, 1981) can be assumption free. We need to define conditions (i.e., prior assumptions, data, and questions) under which a particular statistical tool works well and those conditions under which it does not. Moreover, we must cautiously interpret statements like "normal looking samples automatically provide robust estimates of location" and "if it can't be estimated well, it won't affect inferences" as well as "if the data do not contradict the model, the model is satisfactory for drawing inferences". All statements are true under particular conditions but generally are false: in general, inferences depend on assumptions that the data at hand cannot address. Robustness of Bayesian inference is a joint property of data, prior knowledge, and questions under

consideration; the remainder of this article illustrates this general point in our example.

3. SAMPLE 1 -- INITIAL ANALYSIS

We begin the data analysis by trying to estimate the population total from Sample 1. The standard 95% interval for the finite population total is:

$$N\bar{y} \pm 2\, s\, N \sqrt{\frac{1}{n} - \frac{1}{N}} \,. \qquad (1)$$

For our problem $N = 804$, $n = 100$, and for Sample 1, the sample mean, $\bar{y}$, equals 19,667 and the sample standard deviation, s, equals 142,218. Hence, the observed value of interval (1) is approximately

$$(-5.6 \times 10^6,\ 37.2 \times 10^6) \,. \qquad (2)$$

Interval (2) can be justified under certain assumptions as a 95% interval from either the randomization theory perspective (c.f. Cochran, 1963) or the Bayesian perspective (c.f. Ericson, 1969; Rubin, 1978). From either perspective, the required assumptions are not well supported with a skew sample like Sample 1, but are supported with approximately normally distributed samples.

The practical man examining the standard 95% interval (2) might find the upper limit useful and simply replace the lower limit by the total in this sample, since the total in the population can be no less; this procedure would give a 95% interval estimate of $(2 \times 10^6,\ 37 \times 10^6)$ for the population total. We note that this does cover the true population total, 14×10^6.

Surely, modestly intelligent use of statistical models should produce a better answer because from Table 1, both the population and Sample 1 are very far from normal, and the standard interval is most appropriate with normal populations. Even before seeing any data, we know that sizes of municipalities are far more likely to look something like log normal than normal. Figures 1 and 2 show normal and log-normal probability plots for Sample 1.

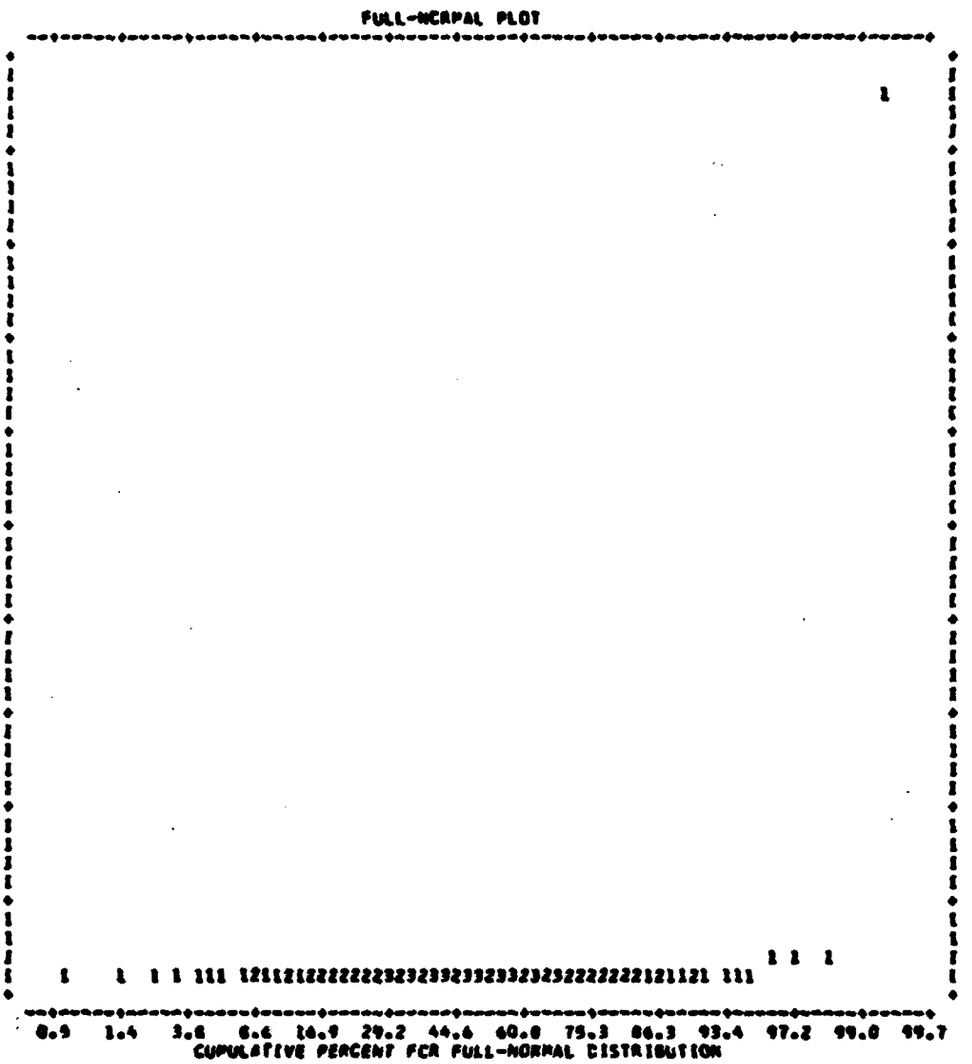


Figure 1: Normal Plot, Y_i, Sample 1.

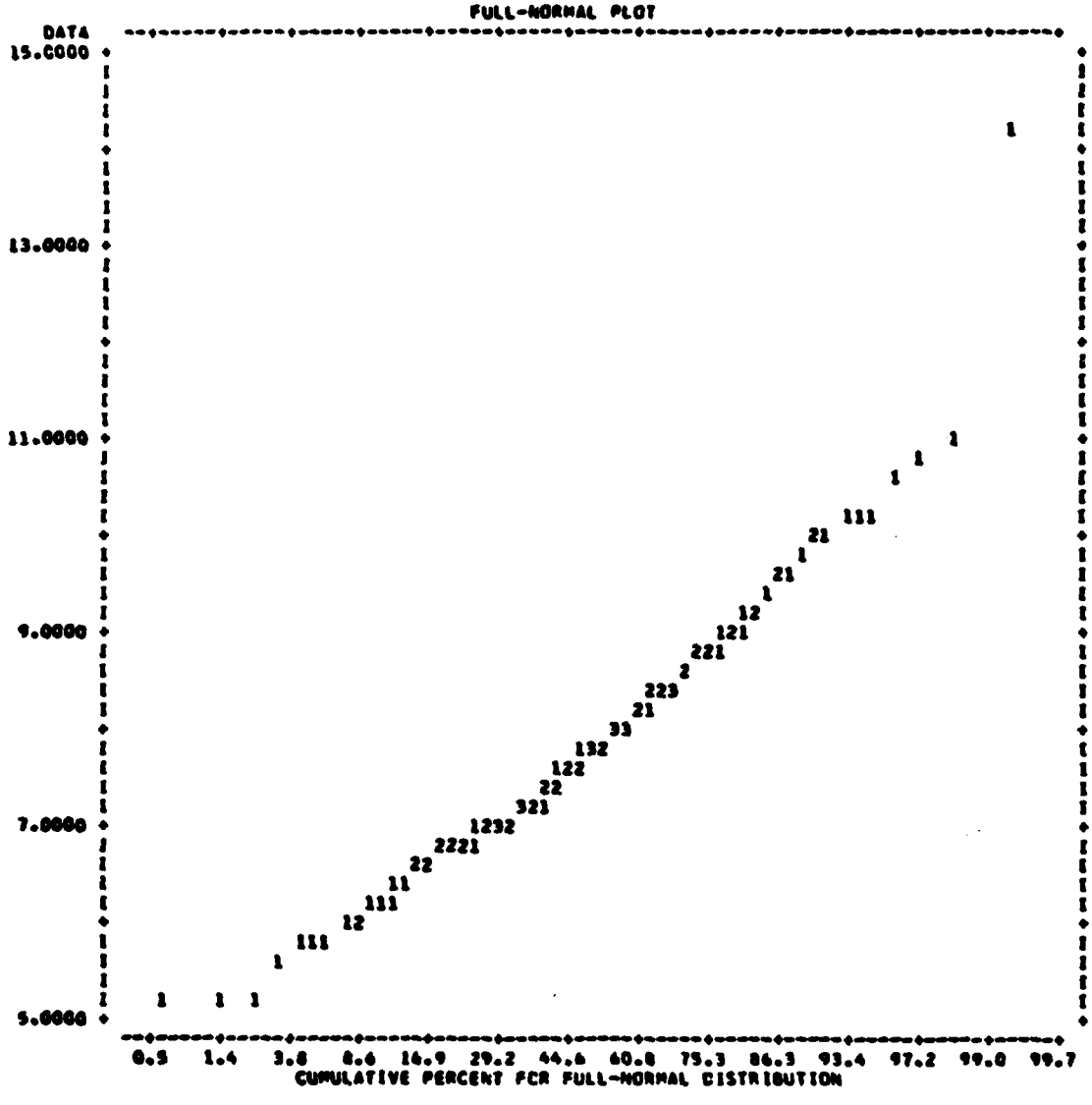


Figure 2: Normal Plot, $Log(Y_i)$, Sample 1.

Although the data do not appear to be exactly log normal (primarily because of one extreme value), they do appear to be so much closer to log normal than normal that an inference based on the log normal model should be superior to the standard inference, and thereby demonstrate that inferences based on more plausible models can easily dominate the standard inference.

Let Y_i, $i = 1,\ldots,804$ be the sizes of the 804 municipalities in New York State, and let $Z_i = \log(Y_i)$. Suppose the 804 values appear like an i.i.d. (independent and identically distributed) sample from a log normal distribution with mean μ and variance σ^2:

$$Z_i \overset{\text{i.i.d.}}{\sim} N(\mu,\sigma^2) \qquad i = 1,\ldots,804 .$$

Based on a random sample of 100 values of Z_i, we can easily obtain the joint posterior distribution of (μ,σ^2) corresponding to prior distribution $p(\mu,\sigma^2)$. Given this posterior distribution, we can find the posterior predictive distribution of the 704 unsampled values of Z_i in the population, and thus the posterior predictive distribution of the 704 unsampled Y_i, and hence the posterior predictive distribution of $Y_+ = \sum_1^{804} Y_i$. (We use the adjective "predictive" to emphasize the distribution of an observable quantity and the adjective "posterior" to mean conditionally given data and model specifications. Since the observable/unobservable distinction is usually obvious from context, we will henceforth drop the adjective "predictive").

Although this procedure is conceptually straightforward, because of the log transformation, the posterior distribution for Y_+ cannot be written in simple closed form. Consequently, we will approximate the posterior distribution of Y_+ using simple simulation techniques. The Appendix outlines the simulation procedure. With prior distribution $p(\mu,\sigma^2) \propto \sigma^{-1}$, the posterior distribution of μ given σ^2 is $N(\bar{Z},\sigma^2)$ and the posterior distribution

of σ^2 is s_z^2 times an inverted χ^2 on 99 d.f. Consequently, it is easy to draw (μ,σ^2) from its posterior distribution. Having drawn values of μ and σ^2, say μ_* and σ_*^2, it is easy to draw 804 values from the posterior distribution of Z_i, $i = 1,804$, given $\mu = \mu_*$ and $\sigma^2 = \sigma_*^2$: values of Z_i that are in the sample are fixed at their observed values and the 704 unsampled values of Z_i are drawn as i.i.d. $N(\mu_*,\sigma_*^2)$. Summing the 100 observed values of $Y_i = \exp(Z_i)$ and the 704 drawn values of $Y_i = \exp(Z_i)$ gives one value of Y_+ drawn from its posterior distribution. Note that any other feature of the population, such as the 95th percentile, can be calculated at this time. Drawing a second value of (μ,σ^2) and repeating the process yields a second value of Y_+.

We drew 100 values of Y_+ which are displayed in Stem-and-Leaf 1. Based on these 100 simulated values, we find that the posterior median of Y_+ is approximately 6.9×10^6, and the 95% interval based on the third and 97th of the 100 drawn values is $(5.4 \times 10^6, 9.9 \times 10^6)$. Although this interval is much narrower than the standard interval and at first glance its limits seem sensible, the interval fails to include the true Y_+, 13.8×10^6! Further, from Stem-and-Leaf 1, even the 99% interval based on all 100 simulated values of Y_+, $(5.2 \times 10^6, 11.8 \times 10^6)$, excludes the true value of Y_+ by a large amount as well as the estimate based on the sample mean, $N \times \bar{y} = 15.8 \times 10^6$. Of particular importance, this failure to include the population total occurs with a sample that from Table 1 appears quite representative of the population. For this sample, the inference for Y_+ based on the log normal specification is, at least for the practical man with hindsight, worse than the simple standard inference for Y_+.

A re-examination of Figure 2 suggests one possible reason for our excluding the right answer when using the log normal specification: although $\log(Y_i)$ is substantially more normal than Y_i, the 100 values of $\log(Y_i)$ are

STEM-and-LEAF 1: The posterior predictive distribution of Y_+ in Sample 1 based on a normal model for $\log(Y_i)$; 100 simulated values in units of 10^6.

```
 5.  124455778888999
 6.  000001112222222333445566667777778888999
 7.  00001112334444557778899
 8.  0112233555667
 9.  01234
10.  3
11.  38
```

STEM-and-LEAF 2: The posterior predictive distribution of Y_+ in Sample 1 based on a normal model for $Y_i^{-1/8}$; 100 simulated values in units of 10^6.

```
 5.  56899
 6.  023335566668999
 7.  0011236789
 8.  022334555888889
 9.  33455668999
10.  01123444466
11.  002356
12.  347
13.  457899
14.  26
15.  77
16.  23
17.  223
18.  0
```

High values 21.3, 21.1, 26.0, 27.1, 30.1, 31.8, 32.5, 53.5

still not really normally distributed. In particular, a straight line in the log transformation probability plot goes well below the largest observed value. As a consequence, values like the largest observed value will be generated less often by the log normal model than once in one-hundred, with the result that the total as estimated under the log normal specification will be relatively small. Perhaps another transformation that produced straighter probability plots would have led to better results.

Before considering other transformations, we note that the example illustrates the first point mentioned in the Section 2. Although the log normal seems to fit the data fairly well in a global sense as judged by the probability plot, the inference for the total seriously underestimates the actual total. Such behavior is not desirable when trying to estimate total amounts of pollutant, radiation, medical expenses or oil reserves, all examples which at times are handled by log normal specifications.

4. SAMPLE 1 -- EXTENDED ANALYSES

Box and Cox (1964) suggest that the following family of power transformations indexed by λ can be useful in Bayesian and likelihood data analyses:

$$Z_i = \begin{cases} Y_i^{\lambda} & \text{if } \lambda \neq 0 \\ \log(Y_i) & \text{if } \lambda = 0 \end{cases}$$

where the Z_i are then assumed to be i.i.d. $N(\mu,\sigma^2)$. With a particular choice of noninformative prior distribution on (λ,μ,σ^2), the posterior distribution of λ is proportional to

$$\text{Var}(Z_*)^{-(n-1)/2}$$

where

$$Z_{*i} = \begin{cases} (Y_i^{\lambda} - 1)/(\dot{y}^{\lambda-1}) & \text{if } \lambda \neq 0 \\ \dot{y}\ \log(Y_i) & \text{if } \lambda = 0 \end{cases}$$

$\dot{y}$ = geometric mean Y_i ,

and $\text{Var}(Z_*) = \sum (Z_{*i} - \overline{Z}_*)/(n - 1)$.

Table 2 presents values of $\text{Var}(Z_*)$ for twelve values of λ. Quite clearly, $\lambda = -1/8$ or even $\lambda = -1/4$ gives a substantially better fit to normality in Sample 1 than $\lambda = 0$. Figure 3 gives the normal probability plot of the sample values, $Y_i^{-1/8}$. Although it is not a straight line, the plot does seem somewhat straighter than the corresponding one for $\log(Y_i)$.

The same technique used to simulate the posterior distribution of Y_+ when $Z_i = \log(Y_i)$ was assumed normal, was used to simulate the posterior distribution of Y_+ when $Z_i = Y_i^{-1/8}$ was assumed normal: simply let $Z_i = Y_i^{-1/8}$ instead of $\log(Y_i)$ and let $Y_i = Z_i^{-8}$ instead of $\exp(Z_i)$. One problem that has to be addressed in principle, and possibly in practice, is that negative values of Z_i are possible because Z_i is assumed to be normally distributed, and negative Z_i values do not map properly into Y_i values. Formally, we will assume that Z_i is distributed as a truncated normal; thus, if a negative value of Z_i is generated, we will draw a new Z_i value; the Appendix provides details.

Based on the 100 simulated values displayed in Stem-and-Leaf 2, the posterior median of Y_+ is 9.6×10^6, and the 95% interval based on the 3rd and 97th values is $(5.8 \times 10^6, 31.8 \times 10^6)$. Note that the interval includes the true value, that the upper limit is similar to the upper limit of the standard interval but that the lower limit is closer to the true value.

Perhaps we have learned how to successfully apply likelihood/Bayesian methods with such data - use the Box-Cox family of power transformations as the basic model with

TABLE 2: Fit of Power Family:

$\mathrm{Var}(Z_*) \times 10^{-7}$

Power	Sample 1	Sample 2
1	2022.57	5226.94
1/2	14.06	30.84
1/4	2.58	4.55
1/8	1.59	2.43
1/16	1.37	1.95
1/32	1.29	1.78
log	1.23	1.65
-1/32	1.18	1.55
-1/16	1.15	1.48
-1/8	1.11	1.37
-1/4	1.13	1.32
-1/2	1.47	1.64

$$Z_* = \begin{cases} (y^\lambda - 1)/(\lambda \dot{y}^{\lambda-1}) & \lambda \neq 0 \\ \dot{y} \log(y) & \lambda = 0 \end{cases}$$ where $\dot{y}$ = geometric mean (y).

With noninformative prior, posterior proportional to $\mathrm{Var}(Z_*)^{-(n-1)/2}$

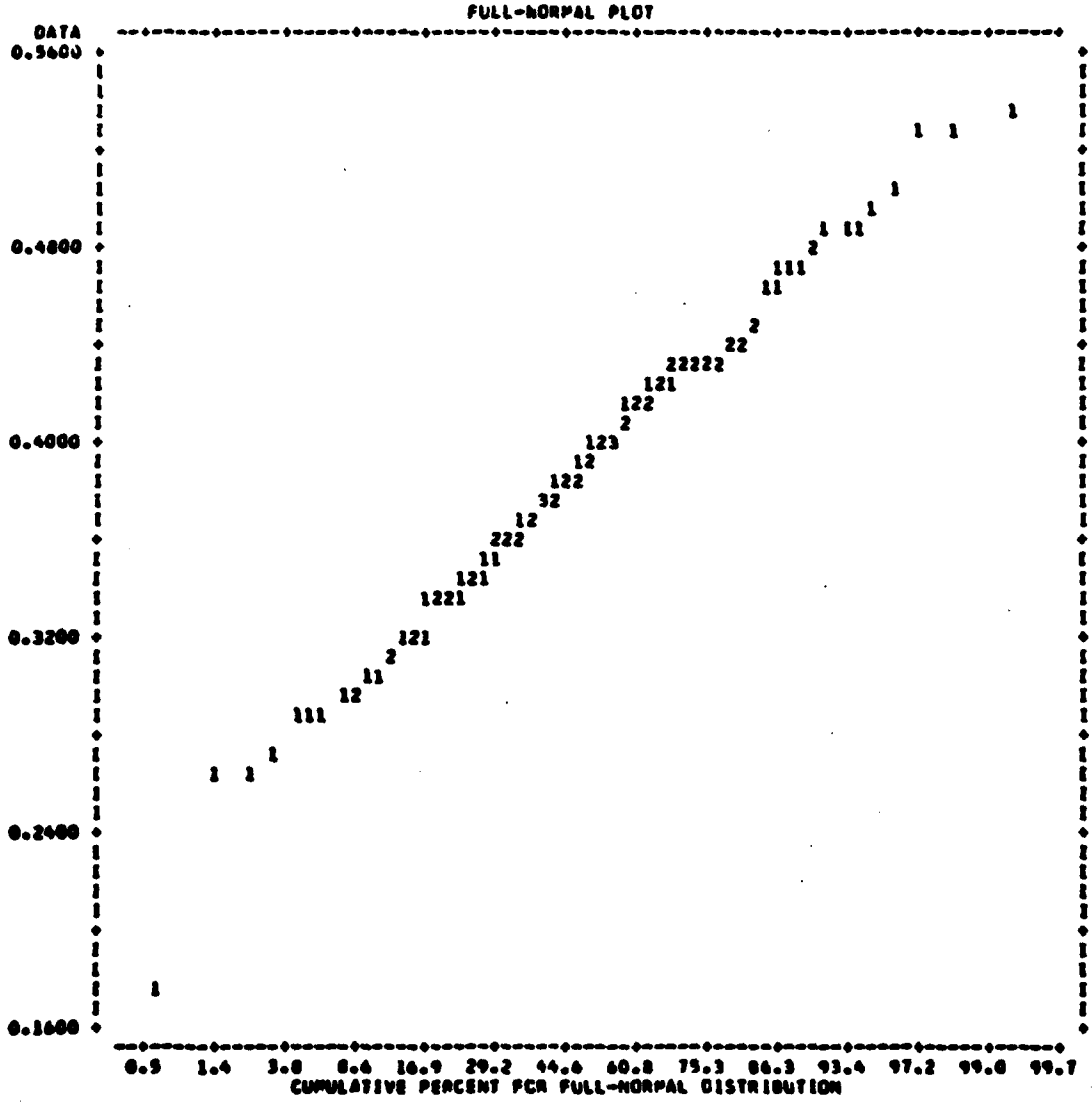


Figure 3: Normal Plot, $Y_i^{-1/8}$, Sample 1.

simulation techniques as the computational tool. But we did not conduct a very rigorous test of this conjecture. We started with the log transformation and obtained an inference that looked respectable but excluded the true value, a fact never known in practice; we then enlarged the family of transformations and found the best fitting transformation. This extended procedure seemed to work in the sense that the resultant 95% interval was plausible and covered the true value. To check on this extended procedure, we will try it on a second random sample of 100. This second sample was the only other one selected.

5. SAMPLE 2

The second sample of 100 cities and towns is summarized in Table 1. The standard inference for the population total from this sample is that $(-3.4 \times 10^6 \times 65.3 \times 10^6)$ is a 95% interval. Substituting the sample total for the lower limit gives $(3.9 \times 10^6, 65.3 \times 10^6)$, a large interval which includes the true value.

The Sample 2 data were first modelled as log normal, and 100 values were drawn from the posterior distribution of the total. The resultant posterior median is 10.6×10^6, and the 95% interval based on the third and 97th simulated values is $(8.2 \times 10^6, 19.6 \times 10^6)$; the 99% interval based on the lowest and highest simulated values is $(8.1 \times 10^6, 25.3 \times 10^6)$. The log normal inference is quite tight and covers the true value, although not the estimate based on the sample mean, $N\bar{y} \doteq 31 \times 10^6$. If we had drawn Sample 2 first, we might have concluded that the log normal model for this population was perfectly satisfactory. But based upon our experience with Sample 1, we should not trust the log normal interval and instead should consider the power family. Figure 4 shows that the log normal does not provide an entirely satisfactory fit to Sample 2 just as it did not to Sample 1. In fact, judging from the normal plots, the log normal fits more poorly in Sample 2 than in Sample 1 even though with pragmatic hindsight, the 95%

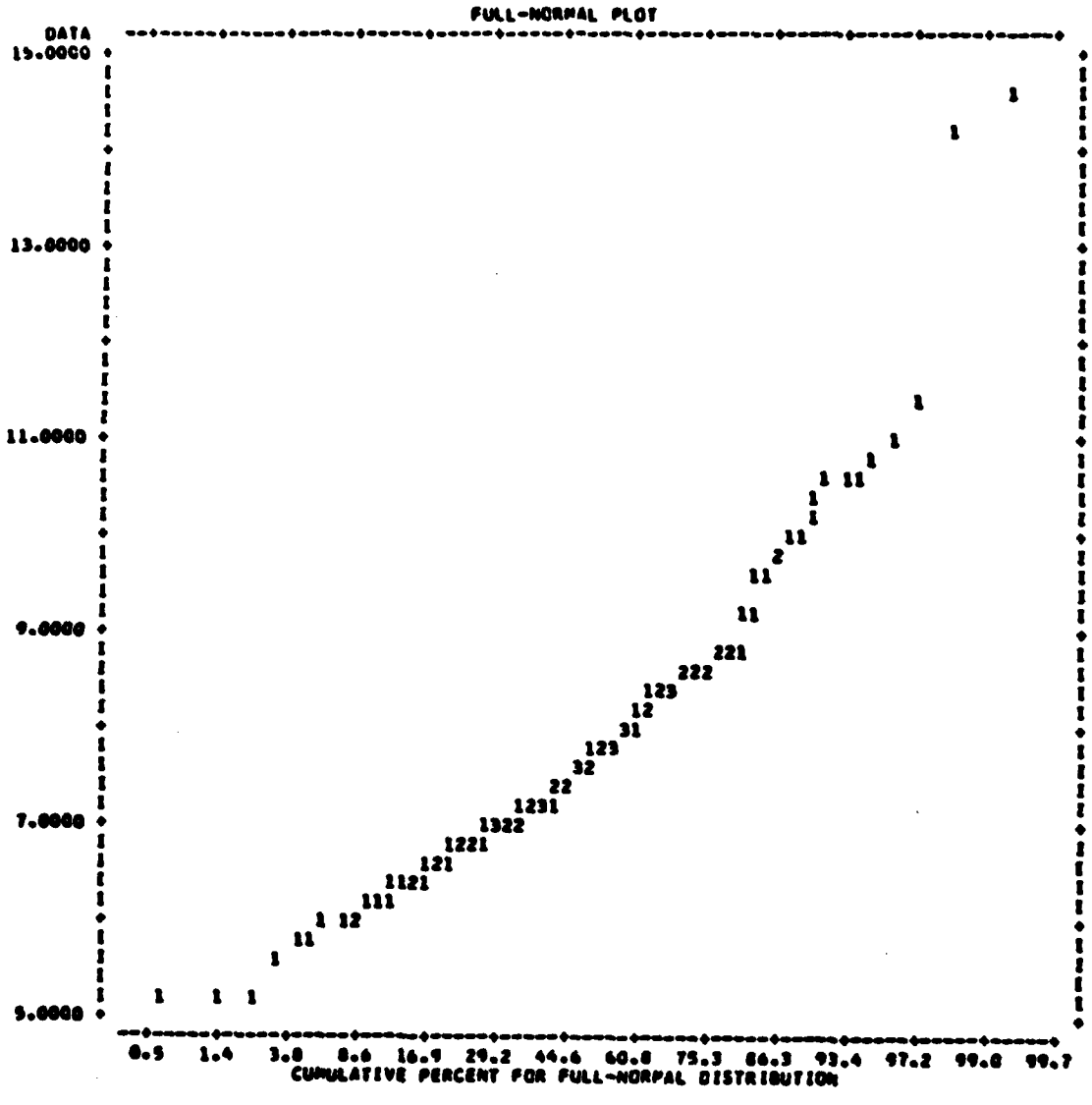


Figure 4: Normal Plot, $\text{Log}(Y_i)$, Sample 2.

interval for Y_+ in Sample 2 is more satisfactory than the 95% interval for Y_+ in Sample 1.

Values of $\mathrm{Var}(Z_*)$ for sample 2 are given in Table 2. As with sample 1, the log is not the best transformation; now, $\lambda = -1/4$ is best, slightly better than $\lambda = -1/8$. Figures 5 and 6 show the normal probability plots for $Z_i = Y_i^{-1/4}$ and $Z_i = Y_i^{-1/8}$ respectively for sample 2; both transformations appear better than the log transformation.

Even though the sampled values of $Y^{-1/4}$ appear to be rather normal, the inferences for the population total resulting from assuming that $Z_i = Y_i^{-1/4}$ follow a truncated normal distributed are, with pragmatic hindsight, atrocious: all 100 generated values of Y_+ are larger than the true value of Y_+ and most of them are much larger. In fact, the resulting 100 draws from the posterior distribution for Y_+ is so long-tailed that it is not well-summarized by a stem-and-leaf display: the minimum value generated is 14.1×10^6, the third lowest is 18×10^6, the median is 57×10^7, the 97th value is 14×10^{15} and the largest value generated is 12×10^{17}! The best value for λ yields entirely unsatisfactory inferences for Y_+: the 99% interval is extremely large and excludes the correct answer.

The inferences that result from using $\lambda = -1/8$ are, from a practical point of view, substantially better although still not very satisfying: the posterior median is 15.7×10^6 and the 95% interval based on the third and 97th values is $(8 \times 10^6, 200 \times 10^6)$. Although in Sample 2 both $Y^{-1/8}$ and $Y^{-1/4}$ are better transformations to normality than $\log(Y_i)$, at least judging by likelihood criteria and probability plots, the inferences for Y_+ under these models are far worse than the inferences for Y_+ under the log normal model, at least to the practical man who wants a tight interval that covers the true value. These results illustrate the second point in Section 2.

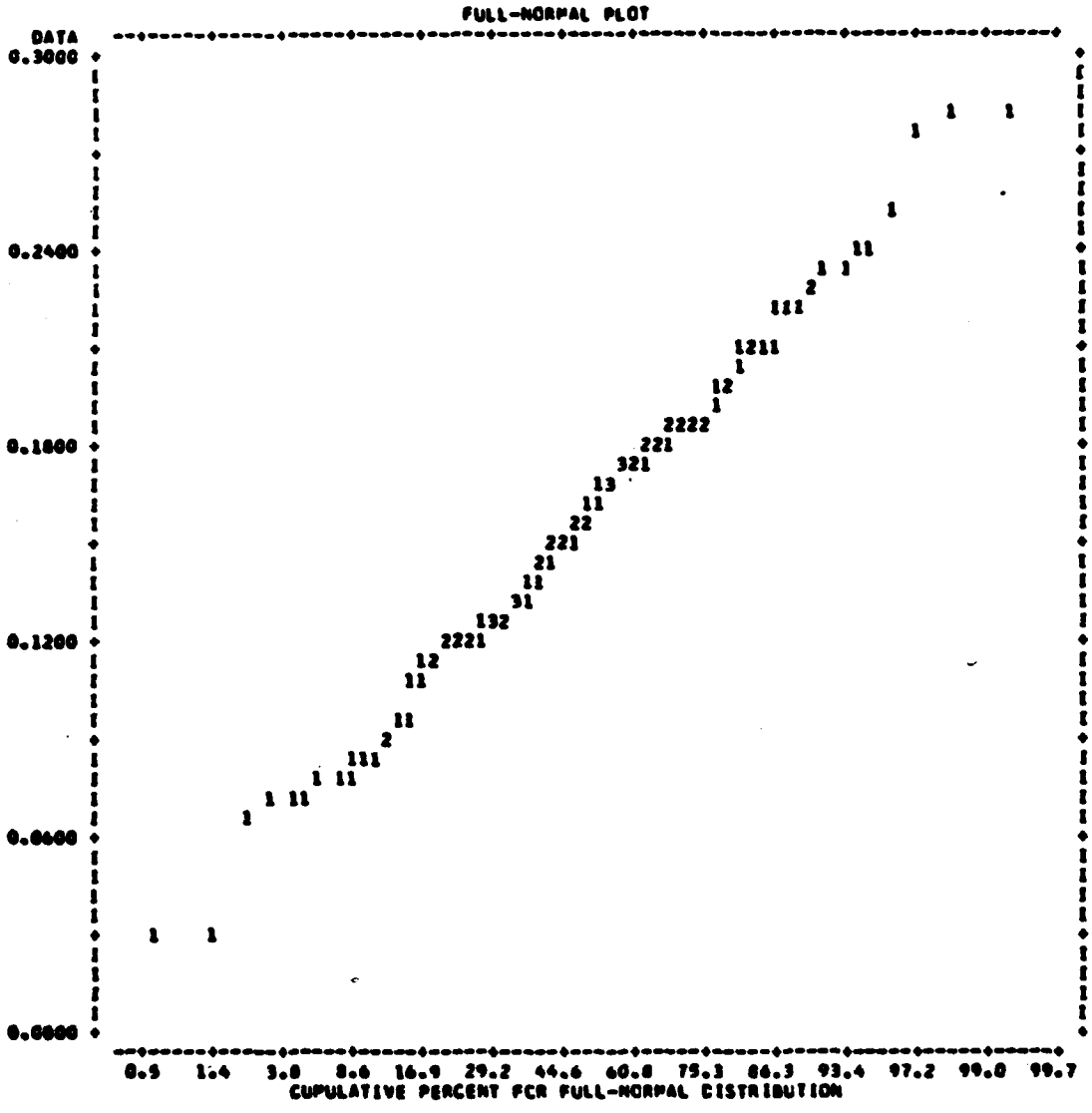


Figure 5: Normal Plot, $Y_i^{-1/4}$, Sample 2.

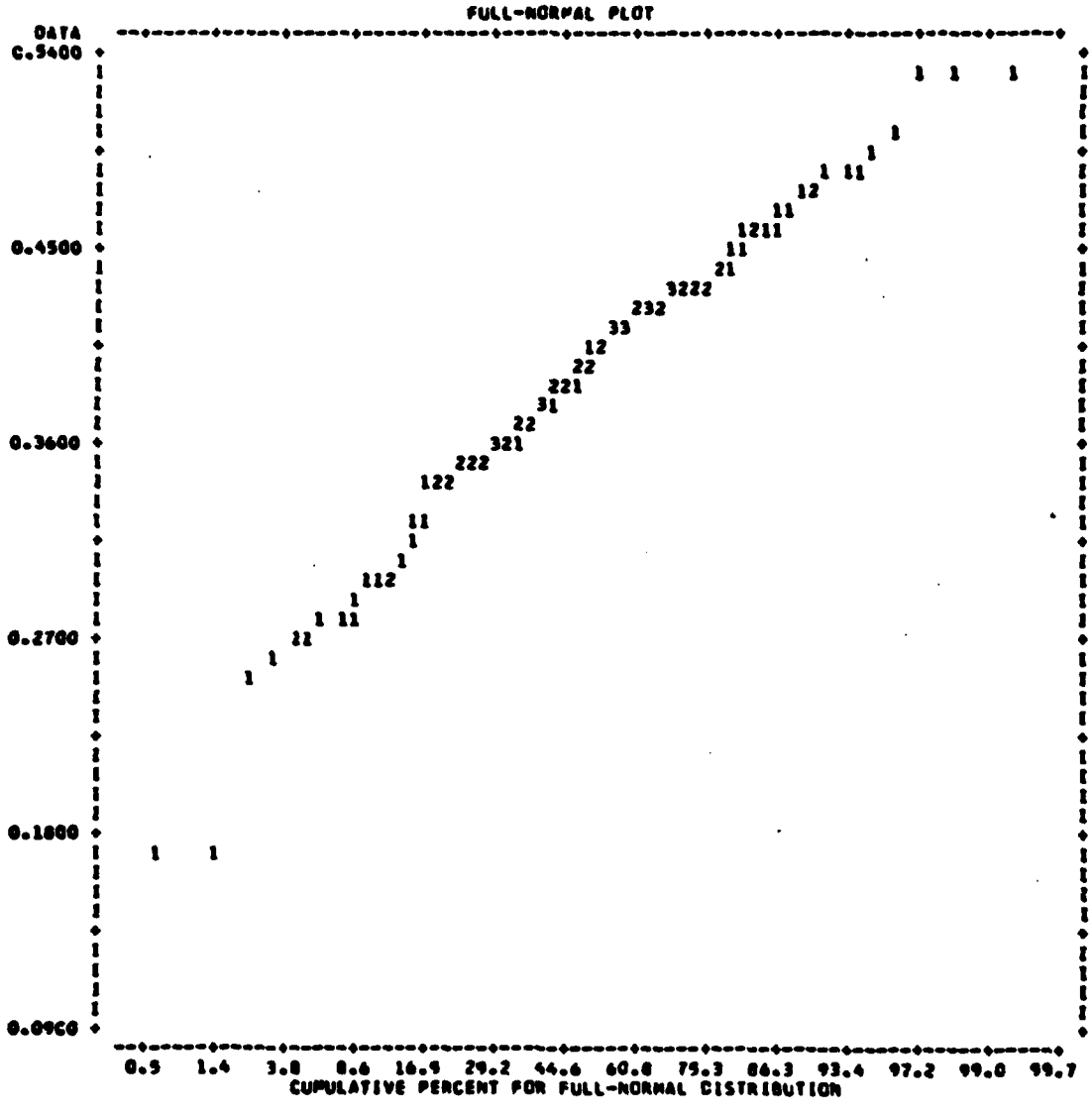


Figure 6: Normal Plot, $Y_i^{-1/8}$, Sample 2.

6. NEED TO SPECIFY CRITICAL PRIOR INFORMATION

What's going on? How can the inferences for the population total in Sample 2 be so much less realistic with better fitting models (e.g., with $Y_i^{-1/8}$ and $Y_i^{-1/4}$ distributed normally) than with worse fitting models (e.g., with $\log(Y_i)$ distributed normally)?

The problem with these inferences in this example is not an inability of the models to fit the data. A larger family of transformations to normality that could further straighten the normal probability plot is not what is needed. In fact, all monotone transformations that map the ith order statistic $Y_{(i)}$ into $\mu + \sigma\Phi^{-1}(\frac{i}{n+1})$ for any μ and σ yield essentially straight normal plots and identical likelihoods, yet these transformations can lead to drastically different inferences for Y_+ depending on their shape for values of Y between the order statistics and especially for values of Y greater than the largest order statistic, $Y_{(n)}$. There exists an infinity of such transformations and none can be contradicted by or selected by probability plots or likelihood criteria alone. The problem is that the question we are asking, "What is the total, Y_+, in the population?", does not have a stable answer from a simple random sample without information external to the observed data about the right tail of the distribution of sizes of municipalities. As we fit models like the power family, the right tail of these models, (especially beyond the upper 1/2 percentage point), is being wagged uncontrollably by the fit of the model to the body of the data (between the lower and upper 1/2 percentage points); behavior of the models in the extreme tails is not being addressed by the relative likelihoods of the models (or by the corresponding probability plots) because there are no data in the extreme tails. Yet the inference for Y_+ is critically dependent upon tail behavior beyond the percentile corresponding to the largest observed Y_i. In order to estimate the total, not only do we need a model that provides a reasonable fit to the observed data, we also

need a model that provides realistic extrapolations beyond the region of the data. For such extrapolations, we must rely on prior assumptions, such as specification of the largest possible size of a municipality.

More explicitly, for our two samples, the three parameters of the power family, λ, μ, σ^2, are basically enough to provide a reasonable fit to the observed data; $\lambda = -1/8$ in Sample 1 and $\lambda = -1/4$ in Sample 2 pretty much generate straight probability plots. But in order to obtain realistic inferences for the population of New York State from both samples, we need to constrain the distribution of large municipalities. Suppose that a priori we know that no city has population greater than 5×10^6. Then using the simulation techniques described in the Appendix, we can draw values from the posterior distribution of size of municipality truncated at 5×10^6. Stem-and-Leafs 3 and 4 display the resultant posterior distributions of Y_+ from Samples 1 and 2 using the best fitting power for each ($\lambda = -1/8$ and $\lambda = -1/4$ respectively) and truncating the size of municipality at 5×10^6. Although this method of providing prior information may seem somewhat clumsy, these Stem-and-Leaf displays yield quite reasonable inferences for the total population size; in both samples, the inferences for Y_+ are tighter than with the untruncated models and in Sample 2, the inference is realistic. In both samples, the 95% intervals cover the true value: the interval in Sample 1 is $(6 \times 10^6, 20 \times 10^6)$ and the interval in Sample 2 is $(10 \times 10^6, 34 \times 10^6)$.

The point is simple, and was stated in Section 2: if we ask a question and wish good statistical answers from the data at hand, we must in general provide models that (a) are flexible enough to let the data fit features it can (e.g., the power family of transformations to normality is nearly flexible enough to generate straight probability plots for our data), and (b) impose prior constraints on critical features of the underlying distribution that the data cannot

STEM-and-LEAF 3: The posterior predictive distribution of Y_+ in Sample 1 based on a truncated normal model for $Y_i^{-1/8}$, $Y_i < 5 \times 10^6$; 100 simulated values in units of 10^6.

```
 5.  56899
 6.  0233355666678999
 7.  0011234567789
 8.  0022233445556888889
 9.  33455668999
10.  011123444466
11.  0012356
12.  34
13.  24789
14.
15.  78
16.  24
17.  122
     High values:  20.0, 24.8, 26.0
```

STEM-and-LEAF 4: The posterior predictive distribution of Y_+ in Sample 2 based on a truncated normal model for $Y_i^{-1/4}$, $Y_i < 5 \times 10^6$; 100 simulated values in units of 10^7.

```
0.  88
1.  011
1.  2222223333333
1.  4444555555
1.  6666777777
1.  88888999999
2.  0000011111
2.  223333333333
2.  444444444445
2.  666777
2.  8
3.  000011
3.  3
3.  45
3.
3.  8
```

address (e.g., restrict all municipality sizes to be less than 5×10^6).

7. GOOD FITS AND SPECIFIED EXTREME VALUES ARE NOT ENOUGH WITH SUCH DATA

The results in the previous section might be seen as suggesting that in order to estimate the population total from such data, it is sufficient to (a) apply a transformation that produces a basically straight probability plot and (b) specify the smallest and largest possible values. This conclusion would be incorrect, however, because inferences for Y_+ are still sensitive to the particular shape of the implied distribution of Y between the order statistics, and once again the data cannot distinguish between the alternatives. Two rather ad hoc inferential techniques will be used to demonstrate this fact.

The first method applies an ad hoc transformation to the Y_i that produces an essentially straight normal probability plot. The method is similar to the use of power transformations in that a transformation is found that straightens the probability plot and then the transformation is regarded as known; it differs from the family of power transformations in that it fits, in some sense, $n - 1$ parameters rather than 1. The procedure for our data is as follows: map Y_i into $\phi^{-1}(\frac{i}{101})$ $i = 1,\ldots,100$; map $Y_{max} = 5 \times 10^6$ into 4 and $Y_{min} = 1$ into -4; linearly interpolate between these points and truncate at Y_{min} and Y_{max}. This procedure produces essentially straight probability plots and truncates at realistic values, yet the resulting inferences for Y_+ are quite different from the inferences for Y_+ based on the truncated $Y_i^{-1/8}$ transformation in Sample 1 or the truncated $Y_i^{-1/4}$ transformation in Sample 2, primarily because of the shape of the transformation between the large order statistics and between $Y_{(n)}$ and Y_{max}: the resultant 95% interval for Y_+ from Sample 1 is $(10 \times 10^6, 57 \times 10^6)$ and from Sample 2 is $(19 \times 10^6, 108 \times 10^6)$.

Relative to this ad hoc transformation, the power transformations smoothed the tails of the implied distributions for Y, and, in Sample 2, thereby discounted to some extent the fact that the two largest order statistics were similar and substantially larger than the other 98 values.

The second rather ad hoc method of inference for Y_+ used here is the Bayesian Bootstrap (Rubin, 1981), which places an improper Dirichlet prior distribution over all possible values, with the result that unobserved values have zero posterior probability and observed values are equally likely. Although not a transformation to normality, it implies a population distribution that perfectly reflects the sample distribution and so is like a transformation to normality with a straight normal probability plot. Note, however, the extreme form of the implied distribution of Y between the order statistics: all mass is concentrated at the order statistics, a vastly different assumption from the previous one which spread out the probability from $Y_{(i)}$ to $Y_{(i+1)}$ according to a linear interpolation rule. Applying the Bayesian Bootstrap to Sample 1 and Sample 2 yields simulated 95% intervals equal to $(4 \times 10^6, 49 \times 10^6)$ and $(7 \times 10^6, 81 \times 10^6)$ respectively. These intervals are respectable, although not particularly sharp, even though the prior specification on which they are based is absurd in that it leads to all posterior mass concentrated at the observed values.

The intervals based on the truncated power transformation, the ad hoc linear interpolation transformation, and the Bayesian Bootstrap are not extremely similar to each other. Consequently, having a model that provides a perfect fit to our data is not enough to draw robust inferences for the population total, even if supplemented with prior specification of extreme values. The inferences are still somewhat sensitive to the shape of the population distribution between the large order statistics implied by the specified transformation.

8. ROBUST QUESTIONS AND SAMPLES OBVIATE THE NEED FOR STRONG PRIOR INFORMATION

Of course, simulation techniques are not needed to estimate totals routinely in practice. Good survey practitioners know that a simple random sample is not a good survey design for estimating the total in a highly skewed population. If stratification variables were available (e.g., that categorized municipalities into villages, towns, cities, and boroughs of New York City), in order to estimate the population total from a sample of 100, oversampling the large municipalities would be highly desirable (e.g., sample all five boroughs of New York City, many cities, several towns, and a few villages).

It should not be overlooked, however, that the simple random samples we drew, although not ideal for estimating the population total, are quite satisfactory for answering many questions without imposing strong prior restrictions. Such questions are robust for our simple random samples in the sense that their answers are relatively stable over a broad range of plausible models. Robustness in this sense is a joint property of questions, data, and models that are not contradicted by observed data.

Table 3 illustrates the relative robustness of inference for interior percentiles from our data. Even with extreme interior percentiles and poorer fitting transformations, the resulting inferences are usually realistic. Better models tend to give better answers, but for questions such as these that are robust for the data at hand, the effect is rather weak: For these questions, prior constraints are not extremely critical and even relatively inflexible models can provide satisfactory answers. Of course, other robust questions would have been the value of the population mid-mean or some other population analogue of a robust-statistic.

The critical issue being illustrated is that robustness is not a property of data alone or questions alone, but particular combinations of data, questions and families of

TABLE 3: SIMULATED POSTERIOR DISTRIBUTIONS FOR PERCENTILES
Based on 100 Draws and Various Transformations to Normality

Population Percentile		Sample 1				Sample 2			
		Log	-1/8	-1/4	$-1/8^T$	Log	-1/8	-1/4	$-1/4^T$
5th ÷ 10^2 = 3.4	Low	1.1	1.9	2.3	1.8	1.0	1.1	1.6	1.6
	3rd	1.2	2.0	2.4	1.9	1.1	1.5	2.0	2.0
	Med^n	2.2	2.9	3.2	2.9	1.7	2.5	3.0	3.0
	97th	3.1	3.5	3.8	3.5	2.6	3.4	3.7	3.7
	High	3.1	3.8	4.0	3.8	2.7	3.6	3.9	3.9
25th ÷ 10^2 = 8.0	Low	5.9	6.1	6.1	6.1	5.2	4.8	5.1	5.1
	3rd	6.4	6.7	6.4	6.6	5.6	5.5	5.6	5.5
	Med^n	8.8	8.8	8.5	8.8	7.8	8.0	7.8	7.8
	97th	10.9	10.7	10.2	10.7	10.8	10.6	10.0	9.9
	High	11.0	11.1	11.1	11.5	11.8	11.0	10.3	10.2
Med^n ÷ 10^3 = 1.7	Low	1.7	1.4	1.3	1.4	1.5	1.3	1.2	1.2
	3rd	1.8	1.6	1.5	1.6	1.7	1.4	1.3	1.2
	Med^n	2.3	2.1	1.9	2.1	2.3	2.1	1.8	1.8
	97th	3.0	2.7	2.4	2.7	3.6	2.8	2.4	2.4
	High	3.6	2.8	2.5	2.8	4.0	3.0	2.6	2.6
75th ÷ 10^3 = 5.1	Low	4.4	4.3	3.9	4.3	4.9	4.0	3.5	3.4
	3rd	4.7	4.3	3.9	4.3	4.9	4.4	3.7	3.6
	Med^n	6.2	5.7	5.2	5.7	7.1	6.0	5.3	5.2
	97th	9.1	7.9	7.3	7.9	12.3	8.6	7.8	7.2
	High	9.9	8.9	8.2	8.8	14.6	9.4	8.2	8.0
95th ÷ 10^3 = 30.3	Low	19	19	22	19	23	22	23	23
	3rd	20	20	22	20	26	24	27	26
	Med^n	26	29	39	29	38	40	48	45
	97th	45	64	112	64	76	61	113	86
	High	47	77	133	74	128	75	127	93

T = truncated at 5×10^6

models. In many problems, statisticians may be able to define the questions being studied so as to have robust answers. We, in fact, did this by summarizing simulated posterior distributions by percentiles rather than moments. Often, however, the practical, important question is inescapably nonrobust. To repeat the central theme of this article: statisticians have an obligation to provide the kinds of answers clients will assume are being provided along with appraisals of the sensitivity of the inferences to assumptions unassailable by the data; we must face the fact that, in general, inferences rely on assumptions that the data at hand cannot address.

ACKNOWLEDGEMENTS

I wish to thank P. R. Rosenbaum for helpful comments on earlier drafts of this paper and D. T. Thayer for programming support. The work was partially supported by the Educational Testing Service and the Mathematics Research Center.

APPENDIX

1. Notation

Y_i $i = 1,\ldots,N$ are the N values of Y in the population.

A priori $\ell < Y_i < u$; e.g., $(0,\infty)$, $(2, 5 \times 10^6)$.

Y_i $i = 1,\ldots,n$ $n < N$ are the known values of Y in the sample.

$f(\cdot)$ is the normalizing transformation, $Z_i = f(Y_i)$.

$$Z_i = f(Y_i), \quad L < Z_i < U, \quad L = f(\ell), \quad U = f(u)$$

$$\bar{Z} = \sum_1^n Z_i/n$$

$$s_z^2 = \sum_1^n (Z_i - \bar{Z})/(n-1)$$

2. Distributions

Given parameters (μ,σ), we assume

$$Z_i \overset{\text{i.i.d}}{\sim} \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{Z_i - \mu}{\sigma}\right)^2\right]/k_L^U(\mu,\sigma) & \text{if } L < Z_i < U \\ 0 & \text{otherwise} \end{cases}$$

where $k_L^U(\mu,\sigma) = \int_L^U \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]dt$.

With prior distribution $p(\mu,\sigma)$ for (μ,σ), the posterior distribution of (μ,σ) is proportional to

$$\begin{cases} p(\mu,\sigma)k_L^U(\mu,\sigma)^{-n}\sigma^{-n}\exp\left[-\frac{1}{2}\sum_1^n\left(\frac{Z_i - \mu}{\sigma}\right)^2\right] & \text{if all } L < Z_i < U \\ 0 & \text{otherwise .} \end{cases}$$

Assuming

$$p(\mu,\sigma) \propto k_L^U(\mu,\sigma)^n/\sigma \ , \qquad \text{(A1)}$$

the posterior distribution of (μ,σ) is the same as with the usual "noninformative" prior distribution for (μ,σ) when $L = -\infty$ and $U = +\infty$.

For most values of L,U,μ and σ in the simulation presented here, $k_L^U(\mu,\sigma)^n \doteq 1$, so that usually the choice of the convenience prior distribution (A1) is not substantially different from the more standard choice proportional to σ^{-1}.

3. Simulation Loop

Each pass through the following three steps produces one draw from the posterior predictive distribution of population quantity.

Step 1 - Draw μ,σ from their posterior distribution

$$\sigma_*^2 = ns_z^2/\chi_{n-1}^2, \quad \chi_{n-1}^2 \text{ a } \chi^2 \text{ variate on } n-1 \text{ df.}$$

$$\mu_* = \bar{Z} + \sigma_* \times N(0,1)/\sqrt{n}, \quad N(0,1) \text{ a standard normal.}$$

Step 2 - Draw unobserved Y_i from posterior predictive distribution given $\mu = \mu_*$ and $\sigma = \sigma_*$.

For $i = n+1,\ldots,N$:

$$Z_i = \mu_* + \sigma_* \times N(0,1)$$

If $L < Z_i < U$, $Y_i = f^{-1}(Z_i)$; otherwise, redraw $N(0,1)$.

Step 3 - Calculate population quantity

E.g. population total = $\sum_1^N Y_i$

population median = median $\{Y_1,\ldots,Y_N\}$.

SURVEY SAMPLING NEW YORK DATA FOR TWO SAMPLES.

a = in neither sample
b = in sample 1
c = in sample 2
d = in both samples

2627319a	25000a	12784a	7184b	5094a
1809578c	24960a	12500a	7166a	5046a
1698281a	23948a	12254a	6992a	5020a
1424815d	23817b	12000a	6954a	5009b
532759a	23475b	11677a	6831a	5003c
318611a	23438a	11255a	6812a	4991a
221991a	23138c	11109a	6805a	4949a
216038a	22155a	11075a	6791a	4948a
190634a	21868a	11062a	6789b	4946a
172959a	21741a	10808a	6744a	4907a
129726a	21561a	10795a	6681a	4896a
102394a	21261a	10721a	6538a	4851a
100410a	20905a	10506a	6485a	4835a
81682a	20519a	10390a	6423a	4784a
76812c	20515b	10362a	6421d	4708a
76010a	20172c	10199b	6316b	4704a
75941a	20129d	10171a	6269a	4673a
70000a	19904a	9968a	6166a	4662a
67492a	19881a	9396a	6128a	4654a
65276a	19181a	9370a	6114b	4629a
65128a	19118a	9268a	6066a	4594a
51646a	18789a	9260b	6062a	4582c
50485a	18775b	9175b	5985d	4469a
50405d	18737a	9145d	5972a	4447a
46517a	18662a	9082a	5967a	4311d
46036a	18580a	9000a	5950c	4286a
45000a	18210a	8979a	5907a	4235c
44807d	18205a	8935a	5877a	4220d
41818a	17968a	8914a	5830a	4216a
38629d	17673a	8838a	5825a	4129a
38330a	17499a	8818a	5803a	4041d
35249c	17286c	8737a	5771a	4023d
34757a	17085a	8732a	5770a	4016a
34641c	16630d	8626c	5763a	4000a
34419a	16122a	8560a	5700a	3991b
34172a	15657a	8524a	5669d	3962a
33306a	15478a	8480a	5507a	3944a
32900a	15387a	8477a	5494a	3933c
30979a	14757a	8381a	5460c	3909b
30448a	14261a	8318a	5417a	3906a
30204a	14225d	8317a	5410a	3878a
30138a	14011a	8255a	5326d	3872a
29564a	13922a	8152a	5256c	3855b
29260a	13917a	7765a	5222c	3852a
28799a	13907a	7752a	5200a	3846a
28772c	13580a	7625a	5182a	3829a
27710a	13500a	7439a	5157a	3795a
26473a	13412b	7412a	5157c	3788a
26443b	12883a	7398b	5105a	3749a
26355a	12868d	7207b	5098a	3737a

3653a
3622a
3616a
3576a
3568a
3566a
3548a
3533a
3533a
3487a
3476a
3471a
3352c
3348d
3343a
3330a
3323a
3320c
3310a
3284a
3278a
3262a
3250a
3218a
3193a
3193b
3180a
3113a
3070a
3060a
3058a
3041a
3035a
2998b
2954c
2940a
2931a
2922a
2921b
2915b
2849a
2847a
2841a
2813a
2809d
2807a
2788a
2785a
2772a
2767a

2758a
2735a
2731a
2725a
2715a
2694a
2693b
2681a
2681a
2622c
2608a
2607a
2584a
2570a
2565a
2553a
2521a
2521d
2499a
2468a
2461b
2461a
2426a
2422a
2410a
2408a
2403d
2366a
2346a
2314c
2307a
2295a
2263a
2256a
2250a
2240a
2213d
2200a
2196a
2185a
2167a
2161a
2160a
2160d
2143d
2124a
2123a
2117a
2108a
2093a

2083b
2080d
2070a
2064a
2064a
2051a
2042a
2038a
2026a
2025a
2019a
2003a
1997a
1996a
1979a
1964a
1953a
1949a
1930a
1917a
1914a
1906a
1901a
1887c
1882a
1871a
1863b
1855a
1848a
1838a
1833a
1830a
1788a
1772a
1768a
1767a
1754a
1752a
1750a
1749d
1748a
1734a
1733a
1731a
1731c
1719a
1717a
1715a
1714b
1712a

1697a
1690d
1647d
1645a
1641a
1630a
1623a
1619a
1610a
1586a
1583a
1580a
1575a
1574a
1550d
1549a
1533b
1507a
1492a
1468a
1468a
1465a
1461a
1460a
1448a
1443a
1438a
1431a
1423a
1416a
1416a
1414a
1405a
1400a
1398a
1390b
1382d
1379a
1366a
1365a
1365c
1361c
1353a
1348a
1344a
1337a
1336a
1334c
1325a
1320a

1317a
1304a
1292a
1290a
1289a
1279a
1276a
1274a
1267a
1265a
1263a
1262a
1258a
1248a
1248a
1247b
1247a
1244b
1237a
1237a
1237a
1236a
1234a
1231a
1228a
1224a
1215a
1215a
1210a
1201c
1181c
1180d
1178a
1176a
1168a
1166a
1166b
1156c
1156a
1150a
1149d
1146a
1146c
1126a
1114a
1109b
1097a
1095a
1090a
1086a

1083a
1082a
1082a
1079d
1078a
1078a
1076a
1068a
1066a
1049a
1045a
1040b
1034a
1033a
1033a
1030a
1027b
1026a
1026a
1025c
1016a
1009c
1004a
1004a
1003a
988a
982a
976c
976a
975a
972a
964a
960a
956a
956d
956b
950b
950a
950a
950a
946a
935a
932c
931b
929a
925a
925a
921a
917c
913a

907a
905a
904a
903a
900a
900a
900a
900a
900d
900a
900a
900a
900a
898a
898a
896d
887b
886c
881a
876a
876a
876d
875a
873a
868a
868a
862a
850c
850a
850a
850a
842d
837a
834a
834b
833a
828a
827a
824a
821a
820a
818a
815a
810a
803a
800a
800a
800a
800b
800a

800a
800a
800a
800a
800a
800a
795a
789a
780a
779a
777a
773a
772a
770a
767a
764a
755a
754a
752a
750a
750a
750a
743d
739a
732a
732d
730a
729a
726a
726a
723a
723a
722a
720b
705a
701a
700d
700a
700a
700a
699b
697a
696a
696a
692a
690a
689c
686a
677a
675a

673a
668a
663a
658a
655a
650a
650a
649a
648a
645a
643a
640a
627a
625c
621a
618a
616a
612a
611a
602a
600a
600a
600a
600a
600a
600a
600a
594a
589a
585a
580a
578a
575a
573c
567a
567a
564a
564a
561a
556c
555a
553a
548a
543a
541a
539a
538a
533a
528c
526a
525a

525a
524a
523a
522a
522c
520d
516a
511a
507b
507a
500a
500a
500a
500a
500a
497a
493a
493a
490a
488a
483a
476a
471a
470a
465a
460a
457a
453a
450a
450a
450a
450a
450a
446b
443c
439c
437a
434a
434a
425b
422a
420d
400a
400a
399b
398c
396a
391a
379a
378a
375c
373b

373a
372a
372a
369a
365c
363a
363c
359a
355a
351a
350a
348a
345a
335a
335a
332b
330a
328a
327d
323a
321a
321a
314b
314a
309a
303d
300a
300a
300a
299a
295a
295a
295a
292a
291a
286a
277a
275a
273a
270a
253a
250d
250a
200a
180d
171b
164d
162c
125a
111a

85a
67a
28a
19a

REFERENCES

1. Box, G. E. P. and D. R. Cox (1964). An Analysis of Transformations. J. Roy. Statist. Soc. B, 26, 211.
2. Cochran, W. G. (1963). Sampling Techniques 2nd ed., New York, Wiley.
3. Efron, B. (1980). Bootstrap Methods: Another Look at the Jackknife. Ann. Statist., 7, 1-26.
4. Ericson, W. A. (1969). Subjective Bayesian Models in Sampling Finite Populations. J. Roy. Statist. Soc. B, 31, 195-224.
5. Miller, R. G. (1974). The Jackknife - a review. Biometrika 61, 1-15.
6. Rubin, D. B. (1978). The Phenomenological Bayesian Perspective in Sample Surveys from Finite Populations: Foundations. Imputation and Editing of Faulty or Missing Survey Data. U. S. Department of Commerce, pp. 10-18.
7. Rubin, D. B. (1981). The Bayesian Bootstrap. Ann. Statist., 9, 130-134.

DR/ed

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2375	2. GOVT ACCESSION NO. AD-A116189	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A CASE STUDY OF THE ROBUSTNESS OF BAYESIAN METHODS OF INFERENCE: ESTIMATING THE TOTAL IN A FINITE POPULATION USING TRANSFORMATIONS TO NORMALITY	5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Donald B. Rubin	8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709	12. REPORT DATE May 1982
	13. NUMBER OF PAGES 35
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

sample surveys, Box-Cox transformations, simulation, sensitivity

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Bayesian methods of inference are the appropriate statistical tools for providing interval estimates in practice. The example presented here illustrates the relative ease with which Bayesian models can be implemented using simulation techniques to approximate posterior distributions but also shows that these techniques cannot be automatically applied to arrive at sound inferences. In particular, the example dramatizes three important messages. The first two messages are concrete and easily stated:

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT (Continued)

(1) Although the log normal model is often used to estimate the total on the raw scale (e.g., estimate total oil reserves assuming the logarithm of the values are normally distributed), the log normal model may not provide realistic inferences even when it appears to fit fairly well as judged from probability plots.

(2) Extending the log normal family to a larger family, such as the Box-Cox family of power transformations, and selecting a better fitting model by likelihood criteria or probability plots, may lead to less realistic inferences for the population total, even when probability plots indicate an adequate fit.

The third message is more philosophical, is not easy to state precisely, but is well-illustrated by the example.

(3) In general, inferences are sensitive to features of the underlying distribution of values in the population that cannot be addressed by the data. Consequently, for good statistical answers we need: (a) models that allow observed data to dominate prior restrictions, and either (b) flexibility in these models to allow specification of realistic underlying features of population values not adequately addressed by observed values, or (c) questions that are robust for the type of data collected in the sense that all relevant underlying features of population values are adequately addressed by the observed data.


DATE
ILME