Ordered estimators for skewed populations

Martin Stephen Rosenzweig
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
Rosenzweig, Martin Stephen, "Ordered estimators for skewed populations " (1968). Retrospective Theses and Dissertations. 4627.
https://lib.dr.iastate.edu/rtd/4627

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
This dissertation has been microfilmed exactly as received 69-9890

ROSENZWEIG, Martin Stephen, 1935- ORDERED ESTIMATORS FOR SKEWED POPULATIONS.

Iowa State University, Ph.D., 1968 Statistics

University Microfilms, Inc., Ann Arbor, Michigan
ORDERED ESTIMATORS FOR SKewed POPULATIONS

by

Martin Stephen Rosenzweig

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of The Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Ames, Iowa
1968
TABLE OF CONTENTS

I. INTRODUCTION .......................... 1
II. REVIEW OF LITERATURE .................. 5
III. RELATIVE EFFICIENCY .................. 16
IV. TESTS ................................ 19
V. ESTIMATION ............................. 37
VI. SUMMARY AND CONCLUSIONS .......... 62
VII. BIBLIOGRAPHY ......................... 69
VIII. ACKNOWLEDGEMENTS .................. 72
I. INTRODUCTION

We propose in this thesis a possible solution to the problem of analysing sample survey data which contains "large" observations. By large, we mean that the list of numbers which constitutes the data contains values which are very much greater than the general body of data, and seemingly bear little resemblance in magnitude to the smaller numbers.

In the more controlled atmosphere of experimental design, the usual procedure is to reject by some method unusually large observations, saying in essence, that by the nature of the design such observations cannot reasonably occur unless someone has contaminated the experiment. There exists quite an array of tests for such "outlier" observations. Dixon (1950) presents an empirical comparison of a number of them. These tests are based on ranges, mid-ranges, and standardized ranges. They are constructed to be sensitive to changes in scale and location. Anscombe (1960, 1961) advances procedures based on the analysis of residuals for each experiment, adjusting the procedure to the individual experiment. However, Tukey (1962) suggests that perhaps it might be better to consider families of "longer-tailed" distributions in order to explain these outliers. He recommends research be done on long tailed symmetric distributions which he feels might be profitably applied in some instances. This approach finds more sympathy among samplers.
Unlike the experimental design situation, in sampling we generally have no reason to reject unusual data as spurious. The appearance of a millionaire in an income survey may be unusual but the value of the characteristic is certainly not a false one. However, the presence of these large observations in the population causes an unwanted increase in the variance of the sample mean. We desire then some adjustment procedure which, short of rejection, reduces the effect on the variance of the estimated mean by large observations.

One point of departure is to construct a means of deciding what is a large observation, or an observation which should be adjusted in some fashion, or what is a population where these procedures are needed. We shall consider two test statistics, one called F, the other T. And we shall use these as preliminary tests to determine the estimator to use. It is in the nature of these tests and the associated estimation procedure that we differ from classical sampling theory. Both testing and estimation will involve distributional assumptions.

Classical sampling theory is characterized by a reluctance to make distributional assumptions about populations of interest. This hesitancy is supported by at least two factors: (1) the extremely satisfactory performance of the sample mean as an estimator, and (2) the difficulty in justifying distributional assumptions. Even in the case of confidence intervals, the use of normality is regarded as an approximation.
resting on the Central Limit Theorem which we use only because we must.

However, the use of concomitant information does have an accepted place in classical sampling. If we locate a second characteristic related to the characteristic of interest, we may make use of a variety of methods for improving the precision of our estimate: regression, ratio estimation, and stratification are standard techniques, found, e.g., in Cochran (1963).

None of these methods employ distributional information, rather they involve sampling for one or more additional characteristics. We assume we have concomitant information in the form of distributional information. How, if at all, is this to be used?

We shall assume that above some level the population being sampled is well approximated by some member of the Weibull family of distributions. Our estimation procedure is based on the idea that we shall have to discriminate between Weibull distributions, namely, those with shape parameter one and those with shape parameter less than 1. The first is the exponential distribution for which the sample mean is the minimum variance, unbiased linear estimator.

The procedure, therefore, runs as follows: The sample is divided into "smaller" values, Group I and "larger" values, Group II. The estimator of the mean is the weighted sum of
the sample mean of Group I, plus the estimator for the mean of Group II. To find the estimator for Group II we need a test which discriminates between the two distributions above. An appropriate test is described below.

Conditional on the result of the test, the estimator for the mean of Group II will be either a sample mean, or a linear function of Weibull order statistics. If the test indicates an exponential distribution, then the mean will be used. If the Weibull with shape parameter less than 1 is indicated, a linear function of order statistics will be used.

The weights for two linear estimators based on Weibull order statistics will be presented. Also, we shall mention the use of a quadratic programming algorithm to produce non-negative weights, since in some instances negative weights may be objectionable.

Finally we shall present some discussion of the finite population problem.

As this thesis deals at length with the Weibull distribution which is a distribution not necessarily familiar to samplers, perhaps some brief comments are in order.

The distribution function of the Weibull is

\[ F(x) = 1 - \exp\left(\frac{x - \text{(location parameter)}}{\text{(scale parameter)}}\right) \]  

where the scale and location parameters are familiar, and the shape parameter has the following effect:
a) If the shape parameter is one, \( F(x) \) is an exponential distribution.

b) If the shape parameter is less than one, \( F(x) \) looks generally like an exponential, asymptotes at zero, but is more skewed than an exponential distribution.

c) If the shape parameter is greater than one, \( F(x) \) is a skewed, convex distribution reminiscent of the chi-square distribution.

In the following material, we shall consider two cases. The first is with the shape parameter known, and the second is with the location parameter known.

In the first case, we use the notation

\[
F(y) = 1 - e^{-(y-a)/(bA)^{1/d} + A} \quad y \geq ba^d - a
\]

where the scale, location and shape parameters are \( b \), \( a \) and \( \frac{1}{d} \). \( A \) is a truncation of the distribution which is justified below. Also, \( d \) is specialized to 3 on occasion.

In case two, we use the notation

\[
F(y) = 1 - e^{-(y/\lambda)^{\beta} + A} \quad y \geq \lambda A^{1/\beta}
\]

where \( \beta \) is the shape parameter and \( \lambda \) is the scale parameter.
II. REVIEW OF LITERATURE

The work published in the area of order statistics, non-parametric statistics or rank or serial statistics is voluminous. It may be of historical interest to observe that an article by Karl Pearson written in 1915 appears in Biometrika (10: 416) in which he referred to a 1907 paper of his own, both dealing with rank correlation. Another early paper is the Fisher and Tippett article in the Proceedings of the Cambridge Philosophical Society 24: 180 (1928) on the asymptotic distribution of the extreme order statistics. A series of articles by E. S. Gumbel appeared in several journals. He presented a general discussion of order statistics in the Annals of Mathematical Statistics 14: 163 (1942). He also wrote a well known treatise on extreme-value theory (Nat. Bur. Stand. Applied Math. Ser. 33, Washington (1954)).

In the area of estimation, Lloyd (1952) showed, that for distributions depending on scale and location parameters only, one can apply generalized least squares to ordered samples, and if the expectations and variances and covariances of the ordered observations satisfy the usual Gauss-Markov conditions, then the result will be minimum-variance, unbiased, linear estimates. Thus, if \( Y \) is a vector of order statistics with

\[
E(Y) = \mu + \sigma \xi
\]

\[
= \rho \theta, \text{ say}
\]
where \( p = (l, a) \) and \( \theta = (\mu, \sigma) \) and \( p \) is known, and

\[
E(Y - p\theta)(Y - p\theta)' = \sigma^2 w \text{ where } w \text{ is an } (n \times n) \text{ symmetric positive-definite matrix of known constants, then the minimum-variance linear unbiased estimator is }
\]

\[
\theta = (p'w^{-1}p)^{-1} p'w^{-1}Y
\]

and

\[
V(\theta) = (p'w^{-1}p)^{-1} \sigma^2.
\]

Lloyd also demonstrates that the sampling variance of the order estimator never exceeds that of the sample mean.

Blom (1956, 1958) considers the extension of minimum variance unbiased estimators to a class of approximations, which he calls "nearly best estimates" in the sense of asymptotic unbiasedness, or asymptotic minimum variance. A linear estimator of, say,

\[
\alpha = k_1 \mu + k_2
\]

from

\[
F\left[ \frac{X - \mu}{\sigma} \right] = F(z)
\]

may be written as

\[
\alpha = \Sigma W_i Z(i) = \Sigma W_i (\mu + \sigma x(i))
\]

He notes that if

\[
\theta_1 = f(\lambda_1)
\]

where \( \lambda_1 \) is a fractile of \( x \) which has a continuous distribution function \( F \), and a density function \( f \), then setting

\[
Y(1) = \theta_{i+1} x(i+1) - \theta_i x(i)
\]
yields a statistic whose approximate variance and covariance are independent of $F$.

Blom's approximation is considered in some detail in the section on tests.

The search in life-testing for distributions to help explain the failure rate of manufactured items, led to some of the work on the exponential, Weibull and extreme-value distributions. For example, the Epstein and Sobel articles (1953, 1954) considered the exponential distribution,

$$f(x) = \frac{1}{\sigma} e^{-x/\sigma} \quad x > 0, \sigma > 0,$$

where they have the first $r$ (out at $n$) ordered observations, a reasonable life-testing situation. The maximum likelihood estimate of $\sigma$ is

$$\sigma_{rn} = \left[ \frac{1}{r} \sum_{i=1}^{r} x(i) + (n-r) x(r) \right] / r.$$

Further, Epstein and Sobel show

$$\frac{2r\sigma^2_{rn}}{\sigma} \text{ distributed as } \chi^2_{(2r)}.$$

In addition to considering censored data in the one-parameter case, they show for the two parameter exponential,

$$f(x) = \frac{1}{\sigma} \exp \left[ \frac{1}{\sigma} (x-\theta) \right] \quad \theta > 0, \sigma > 0, x \geq \theta,$$

that the best linear unbiased estimators are

$$\theta = \frac{[(nr-1)x(1) - x(2) - \cdots - x(r-1) - (n-r+1)x(r)]}{n(r-1)}.$$
and
\[ \sigma = \frac{-(n-1)x(1) + x(2) + \ldots + x(r-1) + (n-r+1)x(r)}{(r-1)} \]

Lieblein and Zelen (1956) consider the question of linear estimation for the extreme value distribution

\[ H(y; \beta, \gamma) = 1 - \exp[-\exp(y^\gamma / \gamma)] - \infty < y < \infty \]

or

\[ H(W) = 1 - \exp[-\exp(W)] \]

setting
\[ y = \beta + \gamma W \]

to standardize. Then the fractiles are defined by
\[ y_p = \beta + \gamma W_p \]

Using the order statistics from the extreme value distribution, they define the estimator

\[ T = \sum_{i=1}^{n} g_i Y(i) \]

where the weights are chosen to minimize the variance of \( T \) subject to unbiasedness. They tabled these weights for samples to size 6. They also consider censored data.

Downton (1966) presents a survey of various estimation schemes for the parameters \( \mu \) and \( \sigma \) from

\[ F_n(x) = \exp[-e(x-\mu)/\sigma] - \infty < x < \infty \]

or

\[ F_1(x) = 1 - \exp[-e(x-\mu)/\sigma] - \infty < x < \infty \]
where $F_n$ is d.f. of largest values and $F_1$ is the d.f. of the smallest values from ordered samples. He suggests if $x_i$ are the order statistics from a sample of size $n$, and if we let

$$V = \sum_{i=1}^{n} X_i \quad \text{and} \quad W = \sum_{i=1}^{n} iX_i$$

be two random variables with realizations $v,w$, that the following unbiased estimators have a reasonable efficiency:

$$\mu = \frac{(n-1) \ln 2 - (n+1)\gamma}{n(n-1) \ln 2} V + \frac{2\gamma}{n(n-1) \ln 2} W$$

$$\sigma = -\frac{n+1}{n(n-1) \ln 2} V + \frac{2}{n(n-1) \ln 2} W$$

The author warns, however, that questions of robustness have not been answered, nor the question of convergence to normality.

The difficulty associated with the estimation of parameters from the Weibull distribution arises from the number of parameters, three, and the form of the distribution function. The Weibull is a generalization of the exponential distribution constructed by raising the exponent of the exponential d.f. to a power, e.g.,

$$F(x) = 1 - e^{-(\frac{x-a}{b})^c} \quad b > 0, \ c > 0, \ x \geq a$$

If it is possible to transform the Weibull to an extreme value distribution, then we can estimate the parameters of the transformed distribution. This also presents some difficulties of its own, and is discussed by Downton (1966). If the
location parameter of the Weibull distribution is known, then
the transformation
\[ y = \ln x \]
\[ \beta = \ln b \]
and
\[ \gamma = c^{-1} \]
yields the extreme-value distribution. Linear combinations of
the log order statistics have been found to be very satis­
factory estimators of the Weibull parameters, as Downton
indicates.

Most commonly, one of the three parameters is assumed
known and the other two parameters estimated. Clearly if \( c \) is
known, we transform to an exponential and use one of the
procedures given in the references above. A number of authors
have followed the practice of estimating the shape parameter
by assuming the location parameter known, e.g.,

\[ F(x) = 1 - e^{-\left(\frac{x}{b}\right)^c} \quad b > 0, \ c > 0, \ x \geq 0 \]

and using a variance (Menon, 1963) or a difference scheme
(Dubey, 1967) to eliminate the scale parameter. Menon uses

\[ d = \left(\frac{6}{n^2}\right) \sum_{i=1}^{n} (\ln X_i)^2 - \left( \frac{\sum_{i=1}^{n} \ln X_i}{n} \right)^2 / (n-1)^{\frac{1}{2}} \]

where

\[ E(d) = d + d0(\frac{1}{n}) \]

and the asymptotic efficiency of \( d \) is 66% compared to the
maximum likelihood estimator. Then
\( c = \frac{1}{d}. \)

He uses \( c \) to get an estimator of \( b \),

\[
\ln b = \left( \sum_{i=1}^{n} \ln X_i \right)/n - \lambda_1/c
\]

and

\[
b = e^{\ln b}
\]

\( \ln(b) \) is asymptotically 95\% efficient.

Dubey bases his procedure on the percentiles of the distribution, i.e., the values \( x_p \) such that

\[
F(x_p) = 1 - \exp[-(x_p/b)^c] = p
\]

He defines the sample 100\% percentile as

\[
y_p = \begin{cases} 
  x(np) & \text{if } np \text{ is an integer} \\
  x([np] + 1) & \text{otherwise}
\end{cases}
\]

and uses the estimator

\[
c = \frac{\ln[-\ln(1-p_1)] - \ln[-\ln(1-p_2)]}{\ln y_{p_1} - \ln y_{p_2}}
\]

where \( 0 < p_1 < p_2 < 1 \). He uses simulation techniques to find the optimum \( p_1 \) and \( p_2 \), and for that case he obtains 82\% asymptotic efficiency. He suggests a number of percentile estimators for \( b \). He also suggests an estimator of \( c \) based on the sum of successive differences of the type indicated above. The optimum fractiles in this case are not determined, though Dubey indicates this problem is found in Blom (1958).
One might observe that it is possible to transform from the Weibull to the exponential distribution, and e.g. this is how Menon proceeds. Also, the Weibull can be transformed into the extreme-value (smallest) distribution, and some work has been done employing this technique, as mentioned above.

Both Bershad (1961) and Cavallini (1963) consider the large observation problem for finite populations in the context of standard sampling techniques. That is, neither author chooses to make distributional assumptions.

Bershad (1961) considers the construction of a minimum mean-square error (MSE) estimator for simple random sampling without replacement. He orders the population, at least conceptually, as

\[ x_1 \leq x_2 \leq \ldots \leq x_N \]

He selects a value, \( x_M \), which divides the universe into "small" and "large" values. His estimator of the total is then

\[ x' = \frac{N}{n} \sum_{i} x_i + W \sum_{ii} x_i \]

where the first summation is over the population values less than \( x_M \) which appear in the sample, and the second over the remaining population values which appear in the sample. He then derives the \( x_M \) and \( W \) which minimize the MSE. He finds

\[ x_M^* = 2 \sum_{ii} x_i^2 \left( \sum_{i,ii} x_i + \frac{N(N-1)}{N-n} \sum_{i} x_i \right) / \left( \sum_{i,ii} x_i \right)^2 + \frac{2N(N-1)}{N-n} \sum_{ii} x_i^2 \]
and
\[ W^* = \frac{N}{n} \left[ \frac{\left( \sum_{I} x_i \right) \left( \sum_{II} x_i \right) - N \left( \sum x_i^2 \right)}{\left[ N \sum x_i^2 - (1 - \frac{N^2}{K}) \left( \sum x_i^2 \right) \right]} \right] \]

where
\[ \Sigma = \text{sum over all population values less than } x_M, \]
\[ \Sigma = \text{sum over all population values greater than or equal to } x_M, \]
\[ \Sigma = \text{sum over all population values}, \]
\[ \Sigma = \text{sum over all population values greater than } x_M, \]
and \[ K = \frac{N^2}{N-1} \left( \frac{N-n}{n} \right). \]

He notes that \( x_M \) can be expected to be an increasing function of the sampling rate.

Cavallini (1963) employs ratio and regression procedures to find proper weights for the large observations. Using the framework of contingency tables, he adjusts his data on the basis of marginal totals which he accepts as correct. Each cell, in the one- or two-way classifications, is split into large and small observations. If we observe \( y_{ijk} \) in each cell, let
\[ R_j = \left[ \Sigma_{i} \left( \frac{N}{n} \right) \Sigma_{k} y_{ijk} \right] / \left[ \Sigma_{i} \left( \frac{N}{n} \right) \Sigma_{k} x_{ijk} \right] \]
where the sum is over all sample members, and
\[ x_{ijk} = \begin{cases} y_{ijk} & \text{if } y_{ijk} < K_j, \text{ say} \\ K_j & y_{ijk} \geq K_j \end{cases} \]
His first estimate of the total for classification \( j \) is

\[
T_{ij}^{(1)} = R_j \frac{N_{ij}}{n_{ij}} x_{ijk},
\]

with approximate variance

\[
V(T_{ij}^{(1)}) = \frac{1}{t} \left[ \left( \sum_k \frac{N_{ij}}{n_{ij}} y_{ijk} \right)^2 + R_j^{(t-1)} \left( \sum_k \frac{N_{ij}}{n_{ij}} x_{ijk} \right)^2 \right]
\]

He also proposes the estimator

\[
Y_{ij} = Y_{i..} - Y_{j.} + Y_{..}
\]

under the assumption of the additive model

\[
Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}
\]

Finally, he examines estimation under a multiplicative model, considering the estimator

\[
Y_{ij}^* = \frac{\left( \sum_i Y_{ij} \right)\left( \sum_j Y_{ij} \right)}{\left( \sum_i \sum_j Y_{ij} \right)}
\]

where

\[
Y_{ij} = \frac{N_{ij}}{n_{ij}} \sum_k y_{ijk}
\]

Searls (1966) suggests the following estimator,

\[
y_t = \frac{\sum_{j=1}^r y_j + (n-r)t}{n} \quad r = 0, 1, \ldots, n, \quad y_j \leq t
\]

i.e. all values of \( y > t \) are replaced by \( t \). Letting \( \mu_t \) and \( \sigma_t^2 \) be the mean and variance of the population with d.f. \( F \), truncated at \( t \), he finds

\[
V(y_t) = \frac{F(t)}{n} \sigma_t^2 + [1-F(t)](t-\mu_t)^2
\]

Also, if \( \mu_t \) is the mean of the remainder of the population,
Bias \( \bar{y}_t \) = -(1-F(t))(\mu_t' - t).

And

\[
\text{MSE}(\bar{y}_t) = V(\bar{y}_t) + (\text{Bias})^2
\]
\[
= \frac{F(t)}{n} \left[ \sigma_t^2 + (1-F(t))(t-\mu_t)^2 \right] + (1-F(t))^2(\mu_t' - t)^2.
\]

Lastly, he observes that by equating \( \text{MSE}(\bar{y}_t) \) and \( V(\bar{y}) \) we can find the region where

\[ \text{MSE}(\bar{y}_t) < V(\bar{y}). \]
III. RELATIVE EFFICIENCY

To indicate the gain in precision over the sample mean which is possible in highly skewed populations, we shall find the asymptotic relative efficiency of the maximum likelihood estimator with respect to the sample mean using

\[ f(y) = \beta \lambda^{-\beta} y^{\beta-1} e^{-\left(y/\lambda\right)^{\beta}} \quad \lambda, \beta, y > 0 \]  

(1.1)

for \( \lambda = 1 \), and a range of values of \( \beta \).

These results will serve to motivate the study of more elaborate estimation schemes for families of distributions related to those in Equation 1.1. Further, values of the parameters most justifying more elaborate procedures will be suggested.

From Equation 1.1, we see

\[ \mu_y = E(y) = \int_{-\infty}^{\infty} \lambda^{-\beta} y^{\beta} e^{-\left(y/\lambda\right)^{\beta}} dy. \]

Using the transformation

\[ x = y^{\beta} \lambda^{-\beta} \]

we find

\[ \mu_y = \int_{0}^{\infty} x^{\beta} e^{-x} dx \]

\[ = \lambda \Gamma(\beta^{-1} + 1) \]

(1.2)

Also

\[ E(y^2) = \int_{0}^{\infty} \beta \lambda^{-\beta+1} y^{\beta+1} e^{-\left(y/\lambda\right)^{\beta}} dy \]

\[ = \lambda^2 \Gamma(2\beta^{-1} + 1). \]

(1.3)
So that

\[ V(\bar{y}) = \frac{1}{m} \lambda^2 \left( \Gamma(2\beta^{-1} + 1) - [\Gamma(\beta^{-1} + 1)]^2 \right) \]  

(1.4)

Since the asymptotic covariance matrix for the maximum likelihood estimators \( \lambda \) and \( \beta \) of the parameters in Equation 1.1 is

\[
\frac{1}{m} \begin{bmatrix}
1.10866 & \lambda^2 \beta^{-2} \\
\text{sym} & 0.25702 \lambda \\
0.60793 \beta^2 & 0.60793 \beta^2
\end{bmatrix}
\]

We can find the asymptotic variance of \( \mu y \) since it is expressible (approximately) as a linear function of \( \lambda \) and \( \beta \), say

\( (\lambda, \beta) \), by a Taylor expansion

\[ \mu y = \mu y + \lambda (\lambda, \beta)(x - \lambda) + \beta (\lambda, \beta)(\beta - \beta) \]

So

\[ V(\mu y) = \lambda (\lambda, \beta) V(\lambda) + 2 \lambda (\lambda, \beta) \beta (\lambda, \beta) \text{cov}(\lambda, \beta) \\
+ \beta (\lambda, \beta)^2 V(\beta) \]

(1.5)

The results are set out in Table 1.

\( H \) would appear that the sample mean is satisfactory for \( \beta \leq 1/2 \). For \( \beta \) less than 1/3, there is a sizable gain in efficiency for the maximum likelihood estimator over the sample mean.
Table 1. Asymptotic relative efficiency $\lambda = 1$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$mV(\mu_y)$</th>
<th>$mV(\bar{y})$</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.21416</td>
<td>0.21460</td>
<td>0.99795</td>
</tr>
<tr>
<td>1</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>1/2</td>
<td>18.43173</td>
<td>20.00000</td>
<td>0.92159</td>
</tr>
<tr>
<td>1/3</td>
<td>460.78527</td>
<td>684.00000</td>
<td>0.67366</td>
</tr>
<tr>
<td>1/4</td>
<td>15,791.41606</td>
<td>39,744.00000</td>
<td>0.39733</td>
</tr>
<tr>
<td>1/5</td>
<td>720,443.96337</td>
<td>3,614,400.00000</td>
<td>0.19933</td>
</tr>
<tr>
<td>1/6</td>
<td>$4.25163 \times 10^7$</td>
<td>$4.78483 \times 10^8$</td>
<td>0.08887</td>
</tr>
<tr>
<td>1/10</td>
<td>$4.29555 \times 10^{15}$</td>
<td>$2.43290 \times 10^{19}$</td>
<td>0.00018</td>
</tr>
</tbody>
</table>
IV. TESTS

In this section, we shall develop a test for skewness based on a statistic which has some intuitive appeal. We wish to distinguish between the exponential distribution, for which the sample mean is a satisfactory estimator, and the Weibull distribution with shape parameter less than one, as an example of a highly skewed distribution. We shall consider some of the distributional characteristics of the test statistic, and then compare it with another statistic of a more complex nature.

It is well-known that, on the average, the order statistics of a sample of size $n$ from a continuous distribution, divide that distribution into $n+1$ areas of equal probability. These areas are the sample coverages. If we think of a rectangle of height $dF(x_{(i)})$ and width $x_{(i+1)}-x_{(i)}$, i.e., the distance between the $i$th and $(i+1)$st order statistics, or a sample spacing, it is a Darboux approximation of the probability integral in that interval. The expected value of the rectangle is $\frac{1}{n+1}$. Thus in a distribution with a monotonically decreasing tail, we conclude that the sample spacings are expected to increase as the index $i$ increases.

Further, in more skewed distributions, one expects larger values of sample spacings in the tail of the distribution. So that dividing the sample spacings into two groups on the basis of index, and forming the ratio of the high index group
to the low index group, should yield a statistic sensitive to skewness.

If the null distribution is

$$f(x) = e^{-x} \quad x > 0$$  \hspace{1cm} (2.1)$$

then it can be shown (see, e.g. Renyi, 1953) that if

$$0 < x(1) < x(2) \cdots < x(n)$$

is the order statistic from Equation 2.1, that the random variables

$$z_1 = nx(1)$$

$$z_i = (n-1+i)(x(1) - x(i-1)) \quad i = 2, \ldots, n$$

are independently and identically distributed (i.i.d.) with density function given by Equation 2.1.

Now, we form the ratio of weighted sample spacings $z_i$, by dividing them into two groups by index and placing the higher indexed $z_i$'s on top, as

$$F = \frac{z_{m+1} + \cdots + z_n}{z_1 + \cdots + z_m}$$

Besides having intuitive appeal, this ratio has a null distribution whose properties are well known.

From the distribution of the $z_i$, we have

$$z_1 + z_2 + \cdots + z_m$$

is distributed as $\chi^2_{2m}$

and the test ratio

$$F = \frac{z_{m+1} + z_{m+2} + \cdots + z_n}{z_1 + z_2 + \cdots + z_m}$$

is distributed as $F_{2(n-m+1), 2m}$. 

We shall only discuss dividing the spacings equally into two groups. However, the independence of the \( z_i \)'s makes it an easy matter to consider other possibilities.

If the distribution of \( y \) under the alternative hypothesis is

\[
f(y) = (db)^{-1}(\frac{y-a}{b})^{\frac{1}{d}-1} \exp[-(\frac{y-a}{b})^{1/d} + c] \quad y \geq bc^d + a
\]

\[
b > 0 \\
c > 0 \\
d \in \mathbb{I}^+
\]  

then in order to find the power of the test, we must find the distribution of \( F \) under this distribution, at least approximately.

We proceed as follows: if we let

\[
x = (\frac{y-a}{b})^{1/d} - c
\]

then

\[
f(x) = e^{-x} \quad x > 0
\]

i.e., \( x \) is distributed exponentially with parameter 1. So that

\[
y = b(x+c)^d + a
\]

expresses a Weibull random variable as a polynomial in exponential random variables. Since the transformation Equation 2.4 is monotone in the region of interest, it follows order is preserved and

\[
y(i) = b(x(i) + c)^d + a
\]
where \( y(i) \) and \( x(i) \) are the order statistics from their respective distributions.

If we write \( x(i) = x_1 \) for simplicity, then a sample spacing from the Weibull distribution is given by

\[
y_{i+1} - y_1 = b(x_{i+1} + c)^3 + a - [b(x_1 + c)^3 + a] \\
= b[(x_{i+1}^3 - x_1^3) + 3c(x_{i+1}^2 - x_1^2) + 3c^2(x_{i+1} - x_1)] \quad (2.6)
\]

As it stands, the distributional problem appears intractable. We shall proceed to an approximate solution.

To do this let us recapitulate some of Blom’s work. The essence of his method is to use the transformation

\[
x = F^{-1}(u)
\]

He renames \( F^{-1} \) as \( G \). Thus, \( G(u) \) expresses any random variable as a function of a uniform random variable. Then using a Taylor expansion about the order parameter \( p_1 \), he approximates the order statistics from any distribution by

\[
x_1 = G(p_1) + G'(p_1)(u_1 - p_1) + \eta_1
\]

where \( u_1 \) is a \( U(0,1) \) random variable and \( \eta_1 \) is the remainder.

Next he considers, the random variable

\[
y_1 = \theta_1 x_{i+1} - \theta_1 x_1
\]

where \( \theta_1 = [G'(p_1)]^{-1} \). So that

\[
y_1 = m_1 + (u_{i+1} - u_1 - \frac{1}{n+1})
\]

where \( m_1 = \theta_{i+1} G(p_{i+1}) - \theta_1 G(p_1) \) and the expected value of the remainder is in absolute value less than \( M/n^2 \). Where \( M \) is
a constant, provided the following conditions hold:

(a) \( G(u) \) is a bounded transformation in the sense defined by Blom.

(b) Given a set of points in \( 0 \leq u \leq 1 \), the first four derivatives of \( G(u) \) are bounded and continuous, and \( G'(u) \neq 0 \) at any of the points in the set.

(c) and for each \( x_i \), \( \left( \frac{1}{n} \right) \) const. as \( (n \to) \), i.e., the \( x_i \) are not end rank statistics.

The \( y_1 \)'s have relatively simple approximate variances and covariances which are independent of the distribution of the order statistics \( x_1 \).

We proceed in much the same fashion, however we modify his method to specifically suit the Weibull distribution. Rather than consider the transformation of a general random variable to a uniform one, we transform to an exponential random variable. Thus, in this instance, our transformation is

\[
x_i = \left( \frac{y_i}{b} \right)^{1/d} - c \quad \text{for } x_i \geq 0
\]

and its inverse is

\[
y_1 = b(x_1 + c)^d + a
\]

So that we would ordinarily have

\[
\beta_1 = [G'(p_1)]^{-1}
\]

\[
= [db(p_1 + c)^{d-1}]^{-1}
\]
However, let us make the further adjustment of expanding the Taylor series about

$$a_1 = \frac{1}{2}[\text{E}(x_1) + \text{E}(x_{1+1})]$$

where $x_1$ is the order-statistic from an exponential distribution. We find that

$$y_{1+1} - y_1 = \theta_1^{-1}[(x_{1+1}-x_1) + \frac{\text{remainder}}{\theta_1}]$$  \hspace{1cm} (2.7)

$$= \theta_1^{-1}(x_{1+1}-x_1)$$

where $\theta_1^{-1} = \text{db}(\alpha_1+c)^{d-1}$

Further

$$y_{1+1} - y_1 = \theta_1^{-1}(x_{1+1}-x_1)$$

$$= \frac{\theta_1^{-1}}{n-1} [(n-1)(x_{1+1}-x_1)]$$

$$= \left(\frac{W_1}{n-1}\right) z_{i+1}\text{ say, i=1, \ldots, n-1.}$$  \hspace{1cm} (2.8)

where $W_1 = \theta_1^{-1}$ and the $z_i$ iid exp(1).

Since, our remainder (Equation 2.7) is the same form as Blom's except for expanding about $a_1$, we might hope that our approximation was also satisfactory. Empirical evidence from Monte Carlo runs indicates that this is so. However, the number of runs is too small for sound judgement.

Thus under the alternative
\[ F = \frac{(m+1)(y_{m+2} - y_{m+1}) + \ldots + (y_n - y_{n-1})}{(n-1)(y_2 - y_1) + \ldots + (m)(y_{m+1} - y_m)} \]

\[ = \frac{W_{m+1} Z_{m+1} + \ldots + Z_n}{W_z Z_2 + \ldots + W_m Z_m} \]

If we write

\[ F = \frac{N}{D} \]

then we can find the null distribution of \( F \) from the joint distribution of \( N \) and \( D \). By the independence of \( N \) and \( D \), their joint distribution is the product of the marginal distributions. Also, since the form of the numerator and the denominator is the same, i.e.

\[ \Sigma W_i Z_i \]

so is the form of their respective distributions.

We begin with \( n = 5 \), and observe

\[ F = \frac{W_3 Z_3 + W_4 Z_4}{W_1 Z_1 + W_2 Z_2} \]

may be written

\[ F = c \frac{Z_3 + W_4 Z_4}{Z_1 + W_2 Z_2} \]

where \( z_1 \) is used for convenience in numbering and is not \( nx(1) \) which is not used since in the case of a truncated distribution like Equation 2.2, the first order statistic has an inflated expectation. Since \( x(1) \) cannot take on all the values between 0 and \( x(2) \), but is restricted to the range \([c,x(2)]\),
where \( c > 0 \), its expected value is shifted to the right. For our purposes, 
\[
z_1 = (n-1)(x(2) - x(1)).
\]

To find \( f(N) \), we set 
\[
N = z_3 + w_4 z_4
\]
\[
v_3 = z_3
\]
and 
\[
g(N, v_3) = e^{-v_3} e^{-(N-v_3)a_4} a_4
\]
\[
= e^{-a_4 N} e^{-(1-a_4)v_3}
\]
where \( a_4 = w_4^{-1} \). 

Then 
\[
h(N) = \int_0^N g(N, v_3) dv_3
\]
\[
= a_4 e^{-a_4 N} \int_0^N e^{-v_3(1-a_4)} dv_3
\]
\[
= \frac{a_4}{1-a_4} \left[ e^{-a_4 N} - e^{-N} \right]
\]
Similarly, 
\[
h(D) = \frac{a_2}{1-a_2} \left[ e^{-a_2 D} - e^{-D} \right]
\]
And 
\[
f(N,D) = h(N)h(D).
\]

Using, \( F = N/D \), and \( S = D \),
\[ f(N,D) = f(FS,S)S \]
\[ = \frac{a_2}{1-a_2} \cdot \frac{a_4}{1-a_4} \left[ e^{-a_4 FS} - e^{-FS} \right] \left[ e^{-a_2 S} - e^{-S} \right] S \]

\[ 0 \leq S < \infty \]

So that

\[ f_1(F) = \frac{a_2}{1-a_2} \cdot \frac{a_4}{1-a_4} \left[ (a_4 F + a_2)^{-2} - (a_4 F + 1)^{-2} - (F + a_2)^{-2} + (F + 1)^{-2} \right] \]

(2.11)

We can proceed in the same fashion for

\[ N = z_4 + w_5 z_5 + w_6 z_6 \]
\[ D = z_1 + w_2 z_2 + w_3 z_3 \]

and find

\[ h(N) = \frac{a_5 a_6}{1-a_5} \left\{ \left[ \frac{1}{a_5 - a_6} - \frac{1}{1-a_6} \right] e^{-a_6 N} - \frac{e^{-a_5 N}}{a_5-a_6} + \frac{e^{-N}}{1-a_6} \right\} \]
\[ = \frac{a_5 a_6}{1-a_5} \left\{ A_1 e^{-a_6 N} - A_2 e^{-a_5 N} + A_3 e^{-N} \right\}, \text{ say,} \]

and

\[ h(D) = \frac{a_2 a_3}{1-a_2} \left\{ \left[ \frac{1}{a_2-a_3} - \frac{1}{1-a_3} \right] e^{-a_3 D} - \frac{e^{-a_2 D}}{a_2-a_3} + \frac{e^{-D}}{1-a_3} \right\} \]
\[ = \frac{a_2 a_3}{1-a_2} \left\{ B_1 e^{-a_3 D} - B_2 e^{-a_2 D} + B_3 e^{-D} \right\}, \text{ say.} \]

So, letting \( a_{1j} = A_j B_j \), we find
\[
f_1(F) = \frac{a_2a_3a_4a_5}{(1-a_2)(1-a_5)} \left[ \alpha_{11}(a_6 F + a_3)^{-2} - \alpha_{12}(a_6 F + a_2)^{-2} + \alpha_{13}(a_6 F + 1)^{-2} \right. \\
- \left. \alpha_{21}(a_5 F + a_3)^{-2} + \alpha_{22}(a_5 F + a_2)^{-2} + \alpha_{23}(a_5 F + 1)^{-2} \\
+ \alpha_{31}(F + a_3)^{-2} - \alpha_{32}(F + a_2)^{-2} + \alpha_{33}(F + 1)^{-2} \right]
\]

And so on, becoming more and more tedious.

However, by restricting consideration to \(m\) elements in the numerator and \(m\) elements in the denominator, we can set out the factors in the density function in matrix array, and conclude the general form of the density function is

\[
f_1(F) = \text{(const.)} \sum_{i=1}^{m} \sum_{j=1}^{m} (-1)^{i+j} \alpha_{ij}(a_{n-j+1} F + a_{m-i+1})^{-2}
\] (2.12)

This function (Equation 2.12) is inconvenient to use for \(n > 5\). Another approach is to find the distribution function (d.f.) of \(F\).

Now,

\[
P(f_0) = P[F \leq f_0] = P[N \leq f_0 D]
\]

\[
= P[z_4 + a_2 z_5 + a_6 z_6 \leq f_0 (z_1 + a_2 z_2 + a_3 z_3)] \text{ for } n = 7
\]

By straightforward integration, let us evaluate the following more general expression;

\[
P = P[b_4 x_4 + b_5 x_5 + b_6 x_6 < a_1 x_1 + a_2 x_2 + a_3 x_3]
\]

where \(x_i\) are exponential random variables.
We have

\[ P = \int_0^\infty \cdots \int_0^\infty b_4^{-1} \sum_{i=1}^3 a_1 x_i \int b_5^{-1} \left( \sum_{i=1}^3 a_1 x_i - b_4 x_4 \right) \]

\[ \int b_6^{-1} \left( \sum_{i=1}^5 a_1 x_i - \sum_{i=4}^6 b_1 x_i \right) - \sum_{i=1}^6 x_i \prod_{i=1}^6 dx_i \]

(2.13)

Eventually, one finds

\[ P = 1 - \left( \frac{b_4^2}{(b_4-b_5)(b_4-b_6)} \right)^3 \prod_{i=1}^3 \left( \frac{b_4}{b_4+a_1} \right) - \left( \frac{b_5^2}{(b_5-b_4)(b_5-b_6)} \right)^3 \prod_{i=1}^3 \left( \frac{b_5}{b_5+a_1} \right) \]

\[ - \left( \frac{b_6^2}{(b_6-b_5)(b_6-b_4)} \right)^3 \prod_{i=1}^3 \left( \frac{b_6}{b_6+a_1} \right) \]

(2.14)

Compare Likes (1967), for example. While this form is simpler for numerical computations, its singular advantage is that extension to larger sample sizes is quite clear, which is not true for the other form. That is, it doesn't seem necessary to re-integrate to find \( P \) for \( n=9 \), say.

To compute the power for \( n=7 \) and \( n=9 \), we must first calculate the appropriate weights from Equations 2.7 and 2.8. The expected value of the exponential order statistic is tabled. It may also be found directly from the expression (Blom, 1958)

\[ E X(1) = \frac{1}{n-j+1} \sum_{j=1}^n \frac{1}{n-j+1} \]

(2.15)
Using Equation 2.14 with the calculated weights, we find for \( n=7 \) the approximate power is \(.33\). And for \( n=9 \), the power of the size \(.05\) test is \(.46\).

Another avenue of approach is to develop a test for the shape parameter \( \beta \), say, of the Weibull distribution. In that event, the problem above is a test of \( \beta = 1 \) vs. \( \beta < 1 \). Let us consider the following development (Fuller, Comments on Weibull Estimation). If the hypothesized value of the shape parameter is \( \beta_0 \), we may write

\[
\beta_0 = \beta (1 + \Delta)
\]

(2.16)

where \( \beta \) is the true value of the parameter and \( \Delta \) is the deviation from the true value. Thus a test of \( \beta_0 = \beta \) is equivalent to the test \( \Delta = 0 \). The procedure is to find an estimator \( \hat{\Delta} \) of \( \Delta \) and use as a test statistic

\[
T = \frac{c_1 \hat{\Delta}}{1 - c_2 \hat{\Delta}}
\]

(2.17)

where \( c_1 \) and \( c_2 \) are constants. A more complete discussion of the derivation of \( \hat{\Delta} \) will be found in the next section.

As before, we can use Equation 2.14 with weights corresponding to this statistic to find the power of the \( T \) test. For \( n=7 \), the power is \(.79\). For \( n=9 \), the power is \(.86\).

In order to gather some empirical data on the behavior of the two tests discussed above, 1000 samples of size 7 and 1000
samples of size 9 were drawn from a Weibull population with d.f.
\[ F(x) = 1 - e^{-x^{1/3}} + 1 \quad x \geq 1 \]
which was generated by Monte Carlo methods. From the results, the power of both test statistics for test of size 0.05 was computed and is tabled below. Also, the approximate theoretical power was calculated and is also tabled.

Table 2. Power of F and T for n=7 and 9, and size = 0.05

<table>
<thead>
<tr>
<th>Calculation</th>
<th>7</th>
<th>9</th>
<th>7</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>.41</td>
<td>.33</td>
<td>.48</td>
<td>.46</td>
</tr>
<tr>
<td>T</td>
<td>.83</td>
<td>.79</td>
<td>.89</td>
<td>.86</td>
</tr>
</tbody>
</table>

At this point let us consider the derivation of the T statistic in some detail. The derivation was deferred until this section since we shall set the location parameter to zero.

We wish to estimate \( \lambda \) and \( \beta \), the parameters from
\[ f(y) = \beta \lambda^{-\beta} y^{\beta-1} e^{-(y/\lambda)^{\beta}} \quad y \geq 0 \quad (2.18) \]
We know that \( x \) obtained by the transformation
\[ x = (y/\lambda)^{\beta} \quad (2.19) \]
is distributed exponentially. Let us begin by developing some information about the exponential order statistics.

The distribution of the \( k^{th} \) order statistic from a sample of size \( n \) from an exp (1) distribution is
and by applying the binomial theorem

\[ f_{x(k)}(x) = \frac{n!}{(k-1)!(n-k)!} \sum_{i=0}^{k-1} \frac{(k-1)!}{i!(k-1-i)!} (-1)^i e^{-(i+n-k+1)x} \]

Recalling the discussion on the T statistic from Section III, we shall eventually want to discuss hypotheses about the relation

\[ \beta_0 = \beta(1 + \Delta) \]

where these terms are defined as before. To start, we see

\[ \int_0^\infty x^{1+\Delta} e^{-ax} dx = \int_0^\infty (a^{-1}y)^{1+\Delta} e^{-ya^{-1}} dy, \text{ say,} \]

\[ = a^{-(2+\Delta)} \Gamma(2+\Delta) \]

where \( y = ax \). So that using Equation 20, we see

\[ E(X^{1+\Delta}) = \int_0^\infty x^{1+\Delta} f_{x(k)}(x) \, dx \]

\[ = \frac{n!}{(k-1)!(n-k)!} \sum_{i=0}^{k-1} \frac{(k-1)!}{i!(k-1-i)!} (1+n-k+1)^{-(2+\Delta)} \Gamma(2+\Delta) \]

where we have set \( a = 1+n-k+1 \). For small \( \Delta \),

\[ a^{-(2+\Delta)} \Gamma(2+\Delta) \approx a^{-2} \Gamma(2) - \Gamma(2)a^{-2}(1na)\Delta + a^{-2} \Gamma'(2)\Delta \]

where \( \Gamma'(2) = \frac{d}{dx} \Gamma(x) \big|_{x=2} \)

Thus
\[ E(X_{(k)}^{1+\Delta}) = \frac{n!}{(k-1)!(n-k)!} \sum_{i=0}^{k-1} \frac{(k-1)!}{i!(k-1-i)!} (-1)^i (i+n-k+1)^{-2} \]

\[ + \sum_{i=0}^{k-1} \frac{(k-1)!}{i!(k-1-i)!} (-1)^i (i+n-k+1)^{-2} [-\ln(i+n-k+1)+0.423] \Delta \quad (2.21) \]

Observe that the first term on the right hand side of Equation 2.21 equals \( E(X_{(k)}) \), which also equals

\[ \sum_{i=0}^{k-1} \frac{1}{n-1} \]

hence the two expressions must be equal. If we set

\[ A_k = \sum_{i=0}^{k-1} \frac{1}{n-1} \]

and

\[ B_k = \frac{n!}{(n-k)!} \sum_{i=0}^{k-1} \frac{1}{i!(k-1-i)!} (-1)^i (i+n-k+1)^{-2} \ln(i+n-k+1), \]

then

\[ E(X_{(k)}^{1+\Delta}) = A_k + [0.423A_k - B_k] \Delta \quad (2.22) \]

It is well known that if \( X_{(1)} \) are the order statistics from an exponential distribution, then variables of the form

\[ (n-i+1)(X_{(1)}-X_{(i-1)}) \quad i = 1, \ldots, n, \ X_{(0)} = 0 \]

are i.i.d. exponentials.

We then remark that the variables

\[ V_k = (n-k+1)(X_{(k)}^{1+\Delta} - X_{(k-1)}^{1+\Delta}) \]

are approximately uncorrelated and, from Equation 2.22, have expectation
\[ E(V_k) = 1 + [0.423 - (n-k+1)(B_k - B_{k-1})] \Delta \]
\[ = 1 + [0.423 + C_k] \Delta, \text{ say.} \quad (2.23) \]

Now we are in a position to consider the variables

\[ Z_k = (n-k+1)(Y^{(k)} - Y^{(k-1)}) \]

where perhaps \( \beta_o \) is some consistent estimator from another procedure. Using Equations 2.19 and 2.23 and the fact \( \beta_o = \beta(1+\Delta) \)

we have immediately

\[ E(Z_k) = \lambda \beta_o [1 + (0.423 + C_k) \Delta] \]
\[ = \lambda \beta_o + (0.423 + C_k) \lambda \beta_o \Delta \quad (2.24) \]

Now, by the properties of the \( Z_k \), we can estimate \( \lambda^{1+\Delta} \) and \( \lambda^{1+\Delta} \Delta \) by the ordinary regression of \( Z_k \) on \( C_k \), i.e.,

\[ z_k = \bar{z} + b(c_k - \bar{c}) \]
\[ = \bar{z} - bc + bc_k \]
\[ = d_o + d_1 c_k, \text{ say,} \quad (2.25) \]

where \( b \) is the usual least-squares estimate of slope and \( \bar{c} = 0 \).

Then

\[ \hat{\Delta} = d_1 / d_o \]
\[ \lambda \hat{\beta}_o = d_o \]
\[ \hat{\beta} = \frac{\beta_o}{1 + \hat{\Delta}} \]

The statistic \( \hat{\Delta} \) is a ratio of linear combinations of random variables, which are independently distributed.
exponential variables if $\beta_0$ is the true value, i.e., if $\Delta = 0$.

From Equation 2.25,

$$d_1 = \frac{\Sigma c_k z_k}{\Sigma c_k^2}$$

and

$$d_0 = \frac{1}{n} \Sigma z_k - 0.423(\Sigma c_k z_k / \Sigma c_k^2),$$

we have that

$$\Delta = \frac{\Sigma c_k z_k}{\Sigma c_k^2[\Sigma (\frac{1}{n} - 0.423(\Sigma c_k^2)^{-1}) z_k]}$$

(2.26)

We now define the statistic $T$ by

$$T = \frac{\Sigma c_k z_k}{\Sigma z_k},$$

(2.27)

which can be shown to be a monotone function of $\hat{\Delta}$. The distribution of $T$ is like that of the serial correlation studied by R. L. Anderson [See Wilks (1962)]. The constants $C_k$ are related by

$$C_1 < C_2 < \ldots < C_n$$

and

$$P(T > r) = \sum_{i=1}^{m} \frac{(C_{n-i+1} - r)^{n-1}}{\Pi_{j=1}^{n} (C_{n-i+1} - C_j)_{j \neq n-i+1}}$$

(2.28)

So, Equation 2.28 enables us to find significance levels for $T$. By construction, we have

(a) $T$ and $\hat{\Delta}$ are independent of $\lambda$ (or $\beta_0$)

(b) $\hat{\Delta}$ diverges from zero as $\beta$ diverges from $\beta_0$,

therefore $\hat{\Delta}$ and $T$ both seem appealing statistics for testing
the hypothesis

\[ H_0 : \beta = \beta_0 \]

against

\[ H_A : \beta \neq \beta_0 \]

Under \( H_0 \), i.e., when \( \beta_0 \) is the true parameter, \( \Delta = 0 \), and this in turn implies, see Equation 2.26,

\[
E[ \sum_{k=1}^{n} c_k z_k ] = 0
\]

or

\[
E\left\{ \sum_{k=1}^{n} c_k [(n-k+1)(Y(\Delta) - Y(\Delta-1))] \right\} = 0 \tag{2.29}
\]

Now this suggests that a possible estimate of \( \beta \) is the \( \hat{\beta} \), say, which satisfies Equation 2.29. It has been shown (Fuller, Comments on Weibull Estimation) that \( \hat{\beta} \) is asymptotically fully efficient.
V. ESTIMATION

We desire to estimate the mean of a population which may be one of two possible types: the distribution is completely unspecified, although it is known that the tail of the distribution is skewed no more than an exponential, or the distribution is well approximated by a Weibull distribution with shape parameter \( \beta = 1/3 \), i.e.,

\[
F(x) = 1 - \exp\left[-\left(\frac{x-a}{b}\right)^{1/3}\right] + c \quad b > 0, \ c > 0, \ x \geq bc^3 + a.
\]  

Equation 3.1

This distribution was chosen because it is an example of a highly skewed distribution which is reasonably tractable mathematically. Weibull distributions with shape parameter less than one and \( c = 0 \) asymptote at \( a \) which is inappropriate for approximating the tail of a smoothly changing distribution, hence we truncate the distribution as indicated in Equation 3.1 with \( c > 0 \).

If we call the observations less than \( x_0 \) Group I, and the remainder of the sample Group II, then the proposed estimation scheme for the mean is to

1. compute the mean of Group I,
2. test Group II for skewness,
3. compute an estimator for Group II, and
4. combine the results of the computations on Group I and Group II.
We shall consider this scheme in some detail, and we begin with linear estimation for samples from a population given by Equation 3.1. In our case, we have

\[ x_o = bc^3 + a. \]

The test in Step 2 makes the estimation scheme non-linear, however the form of the estimator itself is linear. If we let

\[ y = \left( \frac{x-a}{b} \right)^{1/3} - c \]

as we did in Section III, then we know

\[ x = b(y+c)^3 + a \quad \text{for } y \geq 0 \quad (3.3) \]

expresses the Weibull variable as a polynomial in exponential variates. In the region of interest the transformation is monotone, so the same relation holds for the order statistics of both distributions, i.e.,

\[ x_r = b(y_r + c)^3 + a \]

where the subscript is used to mean the \( r \)th order statistic (or \( r \)th element of an order statistic of \( n \) elements). And we have

\[ E(x_r) = b[ E(y_r^3) + 3cE(y_r^2) + 3c^2E(y_r) + c^3 ] + a \quad (3.4) \]

Now we use the obvious fact that if we sum over the sample, the order of the observation is immaterial, i.e.

\[ \sum_{r=1}^{n} y_r = \sum_{i=1}^{n} y_1 \]
where $y_1$'s are unordered observations.

Then

$$E(x) = \frac{1}{n} \sum_{r=1}^{n} x_r$$

$$= b(6 + 6c + 3c^2 + c^3) + a$$  \hspace{1cm} (3.5)

which is the mean of Equation 3.1, and follows, e.g., from

$$\int_0^{\infty} y^3 e^{-y} dy = \Gamma(4)$$

For a specified covariance matrix $V$, we can find the minimum variance, linear unbiased estimator

$$\sum_{r=1}^{n} w_r x_r$$

by minimizing the quadratic form

$$W'VW$$

where $W$ is the vector of weights $w_r$ subject to the restrictions

1. $$\sum_{r=1}^{n} w_r = W'J, \text{ say,}$$

$$= 1$$

2. $$\sum_{r=1}^{n} w_r E(y_r) = W'Y_1, \text{ say,}$$

$$= 1$$

3. $$\sum_{r=1}^{n} w_r E(y_r^2) = W'Y_2, \text{ say,}$$

$$= 2$$

4. $$\sum_{r=1}^{n} w_r E(y_r^3) = W'Y_3, \text{ say,}$$

$$= 6$$
That is, we express the order statistics as
\[ X = \alpha_0 + \alpha_1 Y_1 + \alpha_2 Y_2 + \alpha_3 Y_3 + e \] (3.6)
where \( X \) is the vector of order statistics \( x_r \), \( e \) is the vector of errors with \( E(e) = 0 \)
\[ E(ee^\prime) = Vb^2 \]
and construct an estimator of
\[ \alpha_0 + \alpha_1 + 2\alpha_2 + 6\alpha_3 = (bc^3+a) + 3bc^2 + 6bc + 6b. \]
And we ignore the fact
\[ \alpha_2 = 3\alpha_1 \alpha_3 \]
We note the estimators constructed in this way are unbiased if the population order statistics can be expressed as in Equation 3.4. The estimator is also unbiased for any \( c \) independent of the value used in constructing \( V \), though efficiency is reduced for other \( c \)'s.

Since the coefficient of \( Y_3 \) is \( b \), and the covariance matrix, has the form \( Vb^2 \), where \( V \) is a matrix of constants, we can reduce the mean square error of our linear estimator by minimizing
\[ W'VW + (W'Y - 6)^2 \] (3.7)
subject to
\[
(1) \quad W'J = 1 \\
(11) \quad W'Y = 1 \\
(111) \quad W'Y_2 = 2
\]
which yields an estimator with mean square error
\[ \text{MSE} = b^2 \left[ W'VW + (W'Y_3 - 6)^2 \right] \]  

(3.8)

By examining Equation 3.4, we see that a linear estimator of \( b \) can be constructed which is minimum variance and unbiased by minimizing

\[ W'VW \]

subject to

(i) \( W'J = 0 \)
(ii) \( W'Y = 0 \)
(iii) \( W'Y_2 = 0 \)
(iv) \( W'Y_3 = 1 \)

Or, alternatively, an estimator of \( b \) with 'almost' minimum mean square error can be constructed by minimizing

\[ \overline{W}'\overline{W} + (\overline{W}'Y_3 - 1)^2 \]

subject to

(i) \( \overline{W}'J = 0 \)
(ii) \( \overline{W}'Y = 0 \)
(iii) \( \overline{W}'Y_2 = 0 \)

We may apply generalized least squares to Equation 3.6, and obtain a quadratic estimator of \( b \),

\[ \hat{b}^2 = (n-4)^{-1} (y-X\hat{a})' V^{-1} (y-X\hat{a}) \]  

(3.9)

where

\[
\hat{a} = \begin{bmatrix}
\hat{b}c^3 + a \\
\hat{b}c^2 \\
3bc^2 \\
6ba \\
\hat{b}
\end{bmatrix} = (X'V^{-1}X)^{-1}X'V^{-1}y
\]
and

\[ X = [J, Y_1, Y_2, Y_3] \]

The estimated variance is then

\[ V(x_2) = (X'V^{-1}X)^{-1} \hat{b}^2 \]

Now the weights derived above for the estimator of the mean, either unbiased or almost minimum mean square error, may be negative numbers which in some applications may be objectionable. To eliminate this consider the following, e.g.,

\[
\begin{align*}
\min \ W'VW \\
\text{subject to} \\
W'J = 1 \\
W'Y = 1 \\
W'Y_2 = 2 \\
W'Y_3 = 6 \\
W \geq 0
\end{align*}
\]

This is a quadratic programming problem, and may presumably be solved by the use of the appropriate techniques.

Column A (Table 3) gives the weights of the minimum variance unbiased estimator which is found by minimizing the equations below Equation 3.5. The weights in Column B are found by minimizing Equation 3.7.
The variances of these estimators are indicated below with obvious notation.

\[ V(x_A) = 205.8178 \]
\[ V(x_B) = 178.2060 \quad \text{MSE}(x_B) = 179.6620 \]
\[ V(\bar{x}) = 242.9909 \]

Table 3. Examples of weights for various order estimators n=7

<table>
<thead>
<tr>
<th>r</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.332</td>
<td>2.248</td>
</tr>
<tr>
<td>2</td>
<td>1.001</td>
<td>-2.576</td>
</tr>
<tr>
<td>3</td>
<td>-.062</td>
<td>-.068</td>
</tr>
<tr>
<td>4</td>
<td>-.079</td>
<td>.644</td>
</tr>
<tr>
<td>5</td>
<td>.112</td>
<td>.453</td>
</tr>
<tr>
<td>6</td>
<td>.229</td>
<td>.248</td>
</tr>
<tr>
<td>7</td>
<td>.132</td>
<td>.051</td>
</tr>
</tbody>
</table>

A. The Conditional Estimator

We are now in a position to define the estimator of the large observations, i.e., observations greater than \( x_0 \), as

\[
\tilde{x} = \begin{cases} 
\bar{x}_2 & \text{if } F \text{ (or } T) < \alpha, \text{ say} \\
\tilde{x}_2 & \text{if } F \text{ (or } T) \geq \alpha.
\end{cases}
\]  

(3.10)

where \( \bar{x}_2 \) is the simple mean of the observations greater than \( x_0 \), \( \tilde{x}_2 = \sum w_r x_r \), a linear function of order statistics, \( w_r \) are the weights derived above, and \( F \) and \( T \) are the test statistics derived in Section III.
We see
\[ \bar{E} = \bar{E}(\tilde{x}_2 | F < a)P(F < a) + \bar{E}(\tilde{x}_2 | F \geq a)P(F \geq a) \] (3.11)
where we use F as the test statistic for simplicity of exposition. Also
\[ \bar{E}^2 = \bar{E}(\tilde{x}_2^2 | F < a)P(F < a) + \bar{E}(\tilde{x}_2^2 | F \geq a)P(F \geq a) \]
So that
\[ V(\tilde{x}) = \bar{E}^2 - (\bar{E})^2 \]
\[ = \bar{E}(\tilde{x}_2^2 | F < a)P(F < a) + \bar{E}(\tilde{x}_2^2 | F \geq a)P(F \geq a) \]
\[ - [\bar{E}(\tilde{x}_2 | F < a)P(F < a) + \bar{E}(\tilde{x}_2 | F \geq a)P(F \geq a)]^2 \] (3.12)
\[ = P(F < a) \bar{E}(\tilde{x}_2^2 | F < a) - [\bar{E}(\tilde{x}_2 | F < a)]^2 \]
\[ + P(F \geq a) \bar{E}(\tilde{x}_2^2 | F \geq a) - [\bar{E}(\tilde{x}_2 | F \geq a)]^2 \]
\[ + P(F < a)[\bar{E}(\tilde{x}_2 | F < a)]^2 + P(F \geq a)[\bar{E}(\tilde{x}_2 | F \geq a)]^2 \]
\[ - [P(F < a)\bar{E}(\tilde{x}_2 | F < a) + P(F \geq a)\bar{E}(\tilde{x}_2 | F \geq a)]^2 \]
\[ = P(F < a)V(\tilde{x}_2 | F < a) + P(F \geq a)V(\tilde{x}_2 | F \geq a) \]
\[ + P(F < a) [\bar{E}(\tilde{x}_2 | F < a)]^2 - [P(F < a)\bar{E}(\tilde{x}_2 | F < a)]^2 \]
\[ + P(F \geq a)\bar{E}(\tilde{x}_2 | F \geq a)]^2 + P(F \geq a) [\bar{E}(\tilde{x}_2 | F \geq a)]^2 \]
\[ - [P(F < a)\bar{E}(\tilde{x}_2 | F < a) + P(F \geq a)\bar{E}(\tilde{x}_2 | F \geq a)]^2 \] (3.13)

In order to assess the effect of the preliminary test on the variance of the estimator, we shall want eventually to compare the conditional estimator with unconditional estimators.
when we know the population, exponential or Weibull, from which we are sampling. Therefore, we wish to be able to find the $V(x)$ numerically. To evaluate this expression, we recall first that

$$F = \frac{\sum_{i=m+1}^{n} Z_i}{m \sum_{i=1}^{m} Z_i} = \frac{S_1}{S_2}, \text{ say,}$$

where under the null hypothesis the $Z_1$'s are iid with finite non-zero variance. Then the Lindeberg-Levy Theorem guarantees the asymptotic normality of $S_1$ and $S_2$. We shall assume the distribution in Equation 3.1. On the basis of the work of Chernoff, Gastwirth and Johns (1967), it seems reasonable to conjecture that asymptotic normality obtains in the non-null case also. Let

$$S_3 = \Sigma \omega_r x_r$$

and for moderately large $n$, we may expect, at least approximately, that

$$(S_1, S_2, S_3) \sim N(\mu, \Sigma)$$

where $\Sigma = ((\sigma_{ij}))$ or

$$S^* = (S - \mu) \sim N(0, \Sigma) \quad (3.14)$$

To begin, let us find

$$E(\Sigma \omega_r x_r P \geq \alpha) = E(S_3 \frac{S_1}{S_2} > \alpha)$$

$$= E(S_3 \frac{S_1 - \alpha S_2}{S_2} > \alpha) \quad (3.15)$$
The covariance matrix \( V \) of \( S_1 - \alpha S_2 \), \( S_2 \) and \( S_3 \) is

\[
D \Sigma D'
\]

where \( D = \begin{bmatrix} 1 & -\alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \) \hspace{1cm} (3.16)

Hence

\[
V = \begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{22} & V_{23} & \text{sym} \\ \text{sym} & V_{33} \end{bmatrix} = \begin{bmatrix} \sigma_{11} - 2\alpha \sigma_{12} + \alpha^2 \sigma_{22} & \sigma_{12} - \alpha \sigma_{22} & \sigma_{13} - \alpha \sigma_{23} \\ \sigma_{12} - \alpha \sigma_{22} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} - \alpha \sigma_{23} & \sigma_{23} & \sigma_{33} \end{bmatrix}
\]

Let us find a non-singular transformation \( T \)

\[ TVT' = I \]

In this case, a suitable \( T \) is

\[
\begin{bmatrix} 1 & 0 & 0 \\ -b_{21} & 1 & 0 \\ -b_{31} & -b_{32} & 1 \end{bmatrix}
\]

where

\[
b_{21} = \frac{V_{21}}{V_{11}} \quad b_{31} = \frac{(V_{13}V_{22} - V_{12}V_{23})}{(V_{11}V_{22} - V_{12}^2)}
\]

and \( b_{32} = \frac{(V_{11}V_{23} - V_{21}V_{13})}{(V_{11}V_{22} - V_{12}^2)} \)

Applying Equations 3.16 and 3.17, we find

\[ DS^* = S', \text{ say,} \]

where

\[
S' = \begin{bmatrix} S'_1 \\ S'_2 \\ S'_3 \end{bmatrix} = \begin{bmatrix} S_{1*} - \alpha S_{2*} \\ S_{2*} \\ S_{3*} \end{bmatrix}
\]
and
\[ TS' = \begin{bmatrix} S_1' \\ S_2' - b_{21} S_1' \\ S_3' - b_{31} S_1' - b_{32} S_2' \end{bmatrix} \]
the elements of which are recognizable as deviations from regression. And if we set
\[ \tilde{d} = TS' \]
then
\[ V(d_1) = V_{11} = \gamma_1, \text{ say,} \]
\[ V(d_2) = V_{22} - (V_{12}^2/V_{11}) = \gamma_2, \text{ say} \]
\[ V(d_3) = V_{33} - (V_{11} V_{23}^2 - 2V_{12} V_{23} V_{31} + V_{22} V_{13}^2) / (V_{11} V_{22} - V_{12}^2) = \gamma_3, \text{ say} \]
It follows then
\[ \gamma_1 = d_1 / (\gamma_1)^{1/2} \]  \hspace{1cm} (3.18)
are jointly distributed \( N(0, I) \).

To express the original variables \( S_1 \) in terms of \( \gamma_1 \), we need \( (TD)^{-1} \), which is
\[ (TD)^{-1} = \begin{bmatrix} 1 + b_{21} & a & 0 \\ b_{21} & 1 & 0 \\ b_{31} + b_{21} b_{32} & b_{32} & 1 \end{bmatrix} \]
And
\[ S^* = (TD)^{-1} \tilde{d} = (TD)^{-1} (\gamma_1^{1/2}) \gamma_1 \]
Or

\[ S = (TD)^{-1} (Y^{1/2}) X + \mu \]

where

\[ \gamma^{1/2} = \begin{bmatrix} \gamma_1^{1/2} & 0 & 0 \\ 0 & \gamma_2^{1/2} & 0 \\ 0 & 0 & \gamma_3^{1/2} \end{bmatrix}, \quad X = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \]

Therefore,

\[ S_1 = \gamma_1^{1/2}(1+ab_{21})Y_1 + \gamma_2^{1/2}aY_2 + \mu_1 \]
\[ S_2 = \gamma_1^{1/2}b_{21}Y_1 + \mu_2 \]

and

\[ S_3 = \gamma_1^{1/2}(b_{31}+b_{21}b_{32})Y_1 + \gamma_2^{1/2}b_{32}Y_2 + \gamma_3^{1/2}Y_3 + \mu_3 \]
\[ = a_1Y_1 + a_2Y_2 + a_3Y_3 + \mu_3, \text{ say.} \]

Returning to Equation 3.15,

\[ E(S_3|S_1-aS_2>0) = E(S_3|Y_1>-(\mu_1-a\mu_2)/\gamma_1^{1/2}) \]
\[ = E(a_1Y_1+a_2Y_2+a_3Y_3+\mu_3|Y_1>-(\mu_1-a\mu_2)/\gamma_1^{1/2}) \]
\[ = E(a_1Y_1|Y_1>-(\mu_1-a\mu_2)/\gamma_1^{1/2}) + \mu_3 \]
\[ = E(a_1Y_1|Y_1>c) + \mu_3, \text{ say,} \]

where

\[ c = -\gamma_1^{-1/2}(\mu_1-a\mu_2) \]

To evaluate Equation 3.19 we write

\[ E(a_1Y_1|Y_1>c) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{c}^{\infty} \frac{a_1Y_1dF(y_1)dF(y_2)dF(y_3)}{P} \]

(3.20)
Where \( P = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_c^\infty dF(y_1) dF(y_2) dF(y_3) \)

\[ = \int_c^\infty \frac{c}{1 - \Phi(c)} \int_{-\infty}^{\infty} dF(y_1) \]

\[ = 1 - \Phi(c) \]

Hence

\[ E(a_1 y_1 | y_1 > c) = \frac{a_1}{1 - \Phi(c)} \int_c^\infty y_1 dF(y_1) \]

or

\[ E(S_3 | S_1 - aS_2 > 0) = \frac{a_1}{1 - \Phi(c)} \int_c^\infty y_1 dF(y_1) + \mu_3 \quad (3.21) \]

To evaluate the integral on the right hand side of Equation 3.21, let us make the following digression. If \( x \) is distributed \( N(0,1) \), then

\[ E(x | a < x < b) = \frac{1}{2\pi} \int_a^b \frac{xe^{-x^2/2}}{\Phi(b) - \Phi(a)} \, dx \]

where as usual \( \Phi \) is the standard normal d.f. If we let \( u = x^2 \),

\[ E(x | a < x < b) = \frac{1}{2\pi} \int_a^b \frac{2}{\Phi(b) - \Phi(a)} \, u^{1/2} e^{-u} \, du \]

where \( \Phi \) is the standard normal p.d.f. \( |a| < |b| \) \quad (3.22)

Also

\[ E(x^2 | a < x < b) = \frac{1}{2\pi} \int_a^b x^2 e^{-x^2/2} \, dx \]

\[ = \frac{1}{2\pi} \int_a^b \frac{2}{\Phi(b) - \Phi(a)} \, u^{1/2} e^{-u} \, du \quad \text{if } u = x^2/2 \]

and \( a \cdot b > 0 \)
We introduce the following notation due to Karl Pearson,
\[ I(u, p) = \int_{u^{p+1}}^{u^p e^{-v}} dv \]
\[ \Gamma(p+1) \]

In our case, \( p = 0.5 \) and \( u = \frac{1}{\sqrt{2\pi}} a \) = 0.4082 \( a^2 \) or 0.4082 \( b^2 \).

Therefore,
\[ E(x^2 | a < x < b) = \begin{cases} \frac{I(0.4082 b^2, 0.5) - I(0.4082 a^2, 0.5)}{2[\phi(b) - \phi(a)]} \text{ if } a \cdot b > 0 \\ \frac{I(0.4082 b^2, 0.5) + I(0.4082 a^2, 0.5)}{2[\phi(b) - \phi(a)]} \text{ if } a < 0, \frac{b}{b} > 0 \end{cases} \]

(3.23)

Returning to the original problem, we see by applying Equation 3.22 to Equation 3.21
\[ E(S_3 | S_2 > 0) = \mu_3 + \frac{a_1 e^{-c^2/2}}{\sqrt{2\pi} (1-\phi(c))} \text{ for } -\infty < c < \infty \]

(3.24)

and \( c \) is defined as before.

We proceed to find the variance by first observing that
\[ E(S_3^2 | S_1 - aS_2 > 0) = E[(a_1 Y_1 + a_2 Y_2 + a_3 Y_3 + \mu_3)^2 | Y_1 > c] \]
\[ = E[a_1^2 Y_1^2 + a_2^2 Y_2^2 + a_3^2 Y_3^2 + \mu_3^2 + 2a_1 \mu_3 Y_1 | Y_1 > c] \]
\[ = a_1^2 E(Y_1^2 | Y_1 > c) + 2a_1 \mu_3 E(Y_1 | Y_1 > c) + a_2^2 + a_3^2 + \mu_3^2 \]

(3.25)
From Equations 3.21, 3.23, 3.24 and 3.25, we see

\[
E(S_3^2|S_1-aS_2>0) = \begin{cases} 
\frac{a_1^2[1-I(0.4082 c^2, 0.5)]}{2(1-\phi(c))} + \frac{2a_1\mu_3 e^{-c^2/2}}{\sqrt{2\pi}(1-\phi(c))} + a_2^2 + a_3^2 + \mu_3^2 & \text{if } c > 0 \\
\frac{a_1^2[1+I(0.4082 c^2, 0.5)]}{2(1-\phi(c))} + \frac{2a_1\mu_3 e^{-c^2/2}}{\sqrt{2\pi}(1-\phi(c))} + a_2^2 + a_3^2 + \mu_3^2 & \text{if } c < 0 
\end{cases}
\]

Then from Equations 3.24 and 3.26, we find

\[
V(S_3|S_1-aS_2>0) = \begin{cases} 
\frac{a_1^2[1-I(0.4082 c^2, 0.5)]}{2(1-\phi(c))} + a_2^2 + a_3^2 - \frac{a_1 e^{-c^2}}{2\pi(1-\phi(c))^2} & \text{if } c > 0 \\
\frac{a_1^2[1+I(0.4082 c^2, 0.5)]}{2(1-\phi(c))} + a_2^2 + a_3^2 - \frac{a_1 e^{-c^2}}{2\pi(1-\phi(c))^2} & \text{if } c < 0 
\end{cases}
\]

By similar argument, we can show that

\[
E(S_3^2|S_1-aS_2<0) = \mu_3 - \frac{a_1 e^{-c^2/2}}{\sqrt{2\pi}\phi(c)} \quad -\infty < c < \infty
\]

and
\[ V(S_3 | S_1 - \alpha S_2 < 0) = \begin{cases} \frac{a_1^2[1+I(0.4082c^2, 0.5)]}{2\pi(c)} + a_2 + a_3 - \frac{a_1^2e^{-c^2}}{2\pi(\phi(c))^2} & \text{if } c > 0 \\ \frac{a_1^2[1-I(0.4082c^2, 0.5)]}{2\pi(c)} + a_2 + a_3 - \frac{a_1^2e^{-c^2}}{2\pi(\phi(c))^2} & \text{if } c < 0 \end{cases} \] 

Now referring to the equations, if we write
\[ E(S_3 | S_1 - \alpha S_2 > 0) = M_1 \]
\[ E(S_3 | S_1 - \alpha S_2 < 0) = M'_1 \]
\[ V(S_3 | S_1 - \alpha S_2 > 0) = V_3 \]
\[ V(S_3 | S_1 - \alpha S_2 < 0) = V'_3 \]

then
\[ E(\tilde{x}) = P(F < \alpha)M'_1 + P(F \geq \alpha)M_1 \]
\[ = \mu_3 + \frac{a_1e^{-c^2/2}}{\sqrt{2\pi}} \left[ \frac{P(F > \alpha)}{1-\phi(c)} - \frac{P(F < \alpha)}{\phi(c)} \right] \]
\[ = \mu_3 + \frac{a_1e^{-c^2/2}}{\sqrt{2\pi}} \left[ \frac{\phi(c) - P(F < \alpha)}{\phi(c)(1-\phi(c))} \right] \]

and Equation 3.13 becomes
\[ V(\tilde{x}) = P(F < \alpha)V'_3 + P(F > \alpha)V_3 + P(F < \alpha)[(M'_1)^2 - \mu_2^2] \]
\[ + P(F \geq \alpha)[M_1^2 - \mu_2^2] \]

where \( \mu_2 = E(\tilde{x}) \) is defined by Equation 3.11 and Equation 3.31.
Let us now compare the variances of the minimum variance unbiased estimator and minimum mean square error estimator described on pages 41 to 43 with the appropriate conditional estimator and with the sample mean for \( n = 9 \).

For the exponential distribution with parameter 1, and the truncated Weibull with \( \lambda = 1, \beta = 3, A = 1 \) (see Equation 1.3), the comparison is made in Table 4.

Table 4. Variances of estimators \( n = 9 \)

<table>
<thead>
<tr>
<th>Estimation procedure</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exponential ( \alpha = 1 )</td>
</tr>
<tr>
<td>Always use Estimator A</td>
<td>0.1668</td>
</tr>
<tr>
<td>Always use Estimator B</td>
<td>0.3317</td>
</tr>
<tr>
<td>Conditional Estimator A</td>
<td>0.1161</td>
</tr>
<tr>
<td>Conditional Estimator B</td>
<td>0.1249</td>
</tr>
<tr>
<td>(F Test)</td>
<td></td>
</tr>
<tr>
<td>Sample mean</td>
<td>0.1111</td>
</tr>
</tbody>
</table>

Estimator A = minimum variance unbiased estimator

Estimator B = minimum mean square error estimator
B. The Over-All Estimator

Now let us consider the over-all estimator $\hat{x}$. Recalling the model specified in Equation 3.1, we denote the proportion of the population less than $x_o$ by $P$, and the proportion greater than $x_o$ by $Q = 1-P$.

If the sample size is
\[ n = n_1 + n_2 \]
where $n_1$ = No. of observations less $x_o$ and
\[ n_2 = \text{No. of observations greater than, or equal to, } x_o, \]
then the random variables
\[ p = n_1/n \quad \text{and} \quad q = n_2/n \]
unbiasedly estimate $P$ and $Q$ respectively.

We define
\[ \hat{x} = p\bar{x}_1 + q\bar{x}_2 \]
where $\bar{x}_1$ is the mean of the sample observations less than $x_o$, and $\bar{x}_2$ is some appropriate estimator based on the remaining observations, e.g., one of the linear functions of order statistics derived above.

Let us find
\[ E(\hat{x}) = \mathbb{E}(\hat{x}|p) \]
\[ = \mathbb{E}(p\bar{x}_1 + q\bar{x}_2 | p). \quad (3.33) \]

Let $\mu_1$ and $\mu_2$ be the population means for the part of the population below and above $x_o$ respectively. If we assume
\[ E(x_2 | p) = \mu_2, \sigma_p, \quad (3.34) \]

and if
\[ E(x_1 | p) = \mu_1, \sigma_p \]
then, clearly,
\[
E(x) = E(p \mu_1 + q \mu_2) \\
= p \mu_1 + q \mu_2 \\
= \mu
\]
(3.35)

where \( \mu \) is the mean of the entire population. Equation 3.34 holds only for unbiased estimators. Thus, e.g., the minimum mean square estimator above fails to meet this condition.

Also
\[
V(x) = E[V(x | p)] + V[E(x | p)] \quad (3.36)
\]

First,
\[
V(x | p) = p^2 V(\bar{x}_1 | p) + q^2 V(\bar{x}_2 | p)
\]

Using the fact
\[ p = n_1 / n \]
and
\[ V(\bar{x}_1 | p) = \sigma_1^2 / n_1, \text{ say,} \]
we find
\[
V(x | p) = (p \sigma_1^2) / n + q^2 V(\bar{x}_2 | p)
\]

So that
\[
E[V(x | p)] = p(\frac{1}{n}) + E[q^2 V(\bar{x}_2 | p)] \quad (3.37)
\]

Also, for conditionally unbiased estimators we have
\[ E(x|p) = p\mu_1 + q\mu_2 \]

So that

\[ V(E(x|p)) = \mu_1^2v(p) + \mu_2^2v(q) + 2\mu_1\mu_2\text{cov}(p,q) \]

\[ = (\mu_1 - \mu_2)^2 v(p) \] (3.38)

since \( q = 1 - p \). Hence

\[ V(\hat{x}) = \frac{1}{n}[P\sigma_1^2 + nE[q^2v(x_2|p)] + (\mu_1 - \mu_2)^2PQ] \] (3.39)

If we specialize, we can proceed somewhat further in evaluating Equation 3.39. If \( \bar{x}_2 \) is the maximum-likelihood estimator of \( \mu_2 \), then we know the form of its asymptotic variance,

\[ V(\bar{x}_2|p) = V(\bar{x}_2|n_2) \]

\[ = \frac{k(b,c)}{n_2} \] (3.40)

From Equation 3.40,

\[ E[q^2v(\bar{x}_2|p)] = \frac{E[k(b,c)/n]}{P} \]

\[ = Qk(b,c)/n \] (3.41)

And we find

\[ V(\hat{x}) = \frac{1}{n}[P\sigma_1^2 + Qk(b,c) + (\mu_1 - \mu_2)^2PQ] \] (3.42)

A consistent estimator of Equation 3.39 is

\[ \hat{V}(\hat{x}) = \frac{1}{n}[pv(\bar{x}_1) + qv(x_2) + (\bar{x}_1 - \bar{x}_2)^2pq] \] (3.43)

where

\[ v(\bar{x}_1) = \frac{1}{n_1(n_1-1)} \sum \frac{1}{n_1} (x_i - \bar{x}_1)^2 \]

and \( v(\bar{x}_2) \) is defined below Equation 3.9.
Alternatively, if we write
\[ \hat{x} = p\bar{x}_1 + q\bar{x}_2 \]
\[ = px_1 - p\bar{x}_2 + \bar{x}_2 \]
then
\[ V(\hat{x}) = V(px_1) + V(p\bar{x}_2) + V(\bar{x}_2) - 2\text{COV}(p\bar{x}_1, p\bar{x}_2) \]
\[ - \text{COV}(p\bar{x}_1, \bar{x}_2) + \text{COV}(p\bar{x}_2, \bar{x}_2) \] (3.44)

Set
\[ \Delta p = p - E_p \]
\[ = p - P \]
i.e., it is the deviation from the mean which we assume is \(O(\sqrt{n})\) for all the variables in Equation 3.44. Then expanding each term on the right-hand side of Equation 3.44 in the following manner:
\[ \text{COV}(p\bar{x}_1, p\bar{x}_2) = E[(P+\Delta p)(\mu_1+\Delta \bar{x}_1)-P\mu_1][P+\Delta p)(\mu_2+\Delta \bar{x}_2)-P\mu_2] \]
\[ = E[P\Delta \bar{x}_1 + \mu_1 \Delta p + \Delta p \Delta \bar{x}_1][P\Delta \bar{x}_1 + \mu_2 \Delta p + \Delta p \Delta \bar{x}_2] \]
\[ = E[P^2 \Delta \bar{x}_1 \Delta \bar{x}_2 + \mu_2 \Delta p \Delta \bar{x}_1 + P\Delta \bar{x}_1 \Delta \bar{x}_2 \Delta p] \]
\[ + \Delta \mu_1 \Delta p \Delta \bar{x}_2 + \mu_1 \mu_2 (\Delta p)^2 + \mu_1 (\Delta p)^2 \Delta \bar{x}_2 \]
\[ + P\Delta p \Delta \bar{x}_1 \Delta \bar{x}_2 + \mu_2 (\Delta p)^2 \Delta \bar{x}_1 + (\Delta p)^2 \Delta \bar{x}_1 \Delta \bar{x}_2 \]
\[ = \mu_1 \mu_2 V(p) \] (3.45)

neglecting terms of \(O(n^{3/2})\).

If we proceed in this fashion we find
\[ \text{COV}(p\bar{x}_2, \bar{x}_2) = PV(\bar{x}_2) \] (3.46)
\[ \text{COV}(p\bar{x}_1, \bar{x}_2) = 0 \] (3.47)
\[ V(p \bar{x}_1) = P^2 V(\bar{x}_1) + \mu_1^2 V(p) \quad (3.48) \]

and
\[ V(p \bar{x}_2) = P^2 V(\bar{x}_2) + \mu_2^2 V(p) \quad (3.49) \]

So, that from Equations 3.45-3.49,
\[ V(x) = P^2 V(\bar{x}_1) + Q^2 V(\bar{x}_2) + (\mu_1 - \mu_2)^2 PQ/n \quad (3.50) \]
\[ = P \sigma_1^2 / n + Q^2 V(\bar{x}_2) + (\mu_1 - \mu_2)^2 PQ/n \]

which to the order of approximation is the same form as before.

C. Finite Populations

Of particular practical interest is the question of estimation in finite populations. We shall appeal to a "super" population argument in which we shall assume that our sample of size \( n \) is drawn from a population of size \( N \) which in turn is drawn from an infinite "super" population with mean \( \mu \) and variance \( \sigma^2 \).

Consider the estimator
\[ x^* = \left( \frac{n}{N} \right) \bar{x}_n + \left( 1 - \frac{n}{N} \right) \hat{\mu} \quad (3.51) \]
of the finite population mean where \( \bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \)
and \( \hat{\mu} \) is some appropriate estimator of the infinite population mean \( \mu \) constructed from the sample of \( n \) observations. This is the form of the estimator suggested by Fisher (1942), and to which Brewer and Ferrier refer. In Fisher's case, the mean and the maximum likelihood estimator are combined in this fashion.
We shall define the following expectation operators:

\[ E_1 = E, \text{ the average overall samples of size } N \text{ from the infinite population,} \]
\[ E_2 = E_n, \text{ the average overall samples of size } n \text{ from a sample of size } N, \text{ and} \]
\[ E = E_1E_2, \text{ as the usual average over the infinite population} \]

Clearly,
\[ E_2(\overline{x}_n) = \overline{x}_N \]

and
\[ E_1(\overline{x}_N) = \mu, \text{ the infinite population mean.} \]

If \( \hat{\mu} \) unbiasedly estimates \( \mu \), then
\[ E(\hat{\mu}) = E_1E_2(\hat{\mu}) = \mu \]

although \( E_2(\hat{\mu}) \) need not be \( \overline{x}_N \).

We derive the average squared error of estimator which we shall call variance, i.e.,
\[ E_1E_2(x_* - \overline{x}_N)^2 = V(x^*). \]

We assume \( \hat{\mu} \) is unbiased for \( \mu \). Then
\[ V(x^*) = \left( \frac{n}{N} \right)^2 E_1[E_2(\overline{x}_n - \overline{x}_N)^2] + \left( 1 - \frac{n}{N} \right)^2 E_1[E_2(\hat{\mu} - \overline{x}_N)^2] \]
\[ + 2\left( \frac{n}{N} \right) \left( 1 - \frac{n}{N} \right) E_1[E_2(\overline{x}_n - \overline{x}_N)(\hat{\mu} - \overline{x}_N)] \quad (3.52) \]

Now
\[ E_1E_2(\overline{x}_n - \overline{x}_N)^2 = E_1 \left( \frac{N-n}{Nn} \right) S_x^2 \]
\[ = \frac{N-n}{Nn} \sigma_x^2 \quad (3.53) \]
where \( E_1(S_X^2) = \sigma_X^2 \), the infinite population variance of \( x_1 \).

Also,
\[
E_1E_2(\hat{\mu} - \bar{X}_N)^2 = E_1E_2[(\hat{\mu} - \mu) - (\bar{X}_N - \mu)]^2
\]
\[
= E_1E_2(\hat{\mu} - \mu)^2 + E_1(\bar{X}_N - \mu)^2 - 2E_1E_2(\hat{\mu} - \mu)(\bar{X}_N - \mu)
\]

\( (3.54) \)

Further,
\[
E_1E_2[(\bar{x}_n - \bar{X}_N)(\hat{\mu} - \bar{X}_N)] = E_1E_2[(\bar{x}_n - \mu) - (\bar{X}_N - \mu)][(\hat{\mu} - \mu) - (\bar{X}_N - \mu)]
\]
\[
= E_1E_2[\bar{x}_n(\hat{\mu} - \mu) - (\bar{x}_n - \mu)(\bar{X}_N - \mu) - (\bar{X}_N - \mu)(\hat{\mu} - \mu) + (\bar{X}_N - \mu)^2]
\]

But
\[
E_1E_2[(\bar{x}_n - \mu)(\bar{X}_N - \mu)] = E_1(\bar{X}_N - \mu)E_2(\bar{x}_n - \mu)
\]
\[
= E_1(\bar{X}_N - \mu)^2,
\]

so that
\[
E_1E_2[(\bar{x}_n - \bar{X}_N)(\hat{\mu} - \bar{X}_N)] = E_1E_2[(\bar{x}_n - \mu)(\hat{\mu} - \mu)] - E_1E_2[(\bar{X}_N - \mu)(\hat{\mu} - \mu)]
\]

In addition, with random sampling \( \hat{\mu} \) is independent of \( \bar{X}_{N-n} \) when averaging over all samples of size \( N \), that is, the sample of \( n \) and the sample of \( N-n \) are independent samples from the super population. This means
\[
E_1E_2[(\bar{x}_n - \bar{X}_N)(\hat{\mu} - \bar{X}_N)] = E_1E_2[(\bar{x}_n - \mu)(\hat{\mu} - \mu)]
\]
\[
- E_1E_2[(\frac{n}{N}\bar{x}_n + (\frac{N-n}{N}\bar{X}_{N-n} - \mu)(\hat{\mu} - \mu)]
\]
\[
= E_1E_2[(\bar{x}_n - \mu)(\hat{\mu} - \mu)] - (\frac{n}{N})E_1E_2[(\bar{x}_n - \mu)(\hat{\mu} - \mu)]
\]
\[
= (1 - \frac{n}{N})E_1E_2(\bar{x}_n - \mu)(\hat{\mu} - \mu)
\]

\( (3.55) \)
Using Equations 3.51-3.54,
\[ E_1E_2(x^*-E_1E_2x^*)^2 = V(x^*) \]
\[ = \left( \frac{n}{N} \right)^2 \left( \frac{N-n}{N} \right) \sigma_X^2 + \left( 1 - \frac{n}{N} \right)^2 E_1E_2(\hat{\mu}-\mu)^2 + E_1E_2(\bar{x}_N-\mu)^2 \]
\[ - 2E_1E_2[(\hat{\mu}-\mu)(\bar{x}_N-\mu)] \]
\[ + 2\left( \frac{n}{N} \right)(1-\frac{n}{N})E_1E_2[(\bar{x}_n-\mu)(\hat{\mu}-\mu)] \]

Using the reduction in Equation 3.54,
\[ V(x^*) = \frac{n}{N^2}(1-\frac{n}{N})\sigma_X^2 + (1-\frac{n}{N})^2 \left[ E_1E_2(\hat{\mu}-\mu)^2 + E_1E_2(\bar{x}_N-\mu)^2 \right] \]
\[ - 2\left( \frac{n}{N} \right)(1-\frac{n}{N})E_1E_2[(\bar{x}_n-\mu)(\hat{\mu}-\mu)] + 2\left( \frac{n}{N} \right)(1-\frac{n}{N})E_1E_2[(\bar{x}_n-\mu)(\hat{\mu}-\mu)] \]
\[ = \frac{n}{N^2}(1-\frac{n}{N})\sigma_X^2 + (1-\frac{n}{N})^2 \left[ E_1E_2(\hat{\mu}-\mu)^2 + E_1E_2(\bar{x}_N-\mu)^2 \right] \]
\[ = \frac{n}{N^2}(1-\frac{n}{N})\sigma_X^2 + (1-\frac{n}{N})^2 \frac{1}{N} \sigma_X^2 + (1-\frac{n}{N})^2 E_1E_2(\hat{\mu}-\mu)^2 \]
\[ = \frac{1}{N}(1-\frac{n}{N})\sigma_X^2 + (1-\frac{n}{N})^2 \sigma_{\hat{\mu}}^2 \quad (3.56) \]

where \( \sigma_{\hat{\mu}}^2 = E_1E_2(\hat{\mu}-\mu)^2 \).

Now if \( \hat{\mu} \) is an unbiased, or consistent, estimator of \( \sigma_{\hat{\mu}}^2 \), then
\[ V(x^*) = \frac{1}{N}(1-\frac{n}{N})s_X^2 + (1-\frac{n}{N})^2 \hat{\sigma}^2 \quad (3.57) \]

is an unbiased or consistent estimator of \( V(x^*) \).

In our particular case,
\[ \hat{\mu} = px_1 + q\bar{x}_2, \]
hence

\[ x^* = \left( \frac{n}{N} \right) \bar{x}_h + \left( 1 - \frac{n}{N} \right) (p \bar{x}_1 + q \bar{x}_2) \]

where \( \bar{x}_h \) is the mean of the whole sample, \( \bar{x}_1 \) is the mean of the sample values less than some value \( x_o \), and \( \bar{x}_2 \) is the conditional estimator based on the sample values greater than \( x_o \). \( p \) and \( q \) are sample proportions as before.

Or, using previously defined notation

\[ x^* = \left( \frac{n}{N} \right) \bar{x}_h + \left( 1 - \frac{n}{N} \right) \tilde{x}. \]

Thus from Equation 3.55,

\[ V(x^*) = \frac{1}{N} \left( 1 - \frac{n}{N} \right) s_X^2 + \left( 1 - \frac{n}{N} \right)^2 V(\tilde{x}) \]  \hspace{1cm} (3.58)

where \( V(\tilde{x}) \) is defined earlier in Section V.

From Equation 3.56, we also have

\[ v(x^*) = \frac{1}{N} \left( 1 - \frac{n}{N} \right) s_X^2 + \left( 1 - \frac{n}{N} \right)^2 v(\tilde{x}) \]  \hspace{1cm} (3.59)

where \( v(\tilde{x}) \) is defined in Section V.
VI. SUMMARY AND CONCLUSIONS

In Section III, it was shown that for some population models maximum likelihood estimation achieves greater precision than the sample mean. This motivates the investigation for alternative estimation procedures for highly skewed distributions.

The proposed procedure is as follows: we select some value \( x_0 \), say, perhaps on the basis of prior information, to divide the data. From the observations smaller than \( x_0 \), a mean is computed for each sample. The observations larger than \( x_0 \) are tested for skewness. We consider two test statistics based on the Weibull distribution

\[
F(x) = 1 - \exp\left[-\left(\frac{x}{\lambda}\right)^\beta + A\right] \quad \lambda > 0, \ \beta > 0, \ A > 0 \quad \text{if } x \geq \lambda A^{1/\beta} \tag{4.1}
\]

And we test

\[ H_0 : \ \beta = 1 \]

versus

\[ H_A : \ \beta < 1. \]

If the null hypothesis is not rejected, that is, if the 'tail' of the distribution appears to be exponential (or less skewed than an exponential), we calculate a sample mean for the observations greater than \( x_0 \) also. If, on the other hand, the null hypothesis is rejected on the basis of the sample data, we proceed on the assumption that the tail of the distribution
is well approximated by Equation 4.1. In this case, we construct a linear estimator based on the order statistics from Equation 4.1.

The first test statistic is

$$F = \frac{Z_{m+1} + \ldots + Z_n}{Z_1 + \ldots + Z_m}$$

(4.2)

where $Z_i = (n-i+1)(x_{(i)} - x_{(i-1)})$ for $i = 2, \ldots, n$ and $x_{(1)}$ are the order statistics of the large observations in the sample. Under the null hypothesis, $Z_i$ are iid exponential random variables hence $F$ is distributed as $\chi_2^2(n-m+1),2m$. For the particular case $\beta = 1/3$, $F$ has the approximate density

$$\text{(const)} \sum_{i=1}^{m} \sum_{j=1}^{m} (-1)^{i+j} a_{ij} (a_{n-j+1}^F + a_{m-i+1}^F)^{-2}$$

(4.3)

where $a_{ij}$, $a_{n-j+1}$ and $a_{m-j+1}$ are constants.

The second test criterion is

$$T = \frac{\sum_{i=1}^{r} c_k Z_k}{\sum_{i=1}^{n} Z_k}$$

(4.4)

where $c_k$ is a constant and $Z_k$ is defined below Equation 4.2.

Also, the $c_k$ are related by

$$c_1 < c_2 < \ldots < c_n$$

A result of R. L. Anderson states
\[ P(T-r) = \sum_{i=1}^{m} \frac{(c_{n-1}+1-r)^{n-1}}{n \prod_{j=1}^{n} (c_{n-1}+1-c_{j}) \text{ for } c_{n-m} \leq r \leq c_{n-m+1}} \] (4.5)

Equation 4.5 can be used to find significance levels for \( T \).

For example, the 5\% significance level for

\[ n = \begin{cases} 7 & \text{is } \cdot2385 \\ 9 & \text{is } \cdot2999 \end{cases} \]

Under the alternative hypothesis, defining \( Z_k \) as before

\[ Z_k = d_k y_k \]

where \( d_k \) is a constant and the \( y_k \) are iid exponentially.

Thus

\[ T = \frac{\Sigma c_k d_k y_k}{\Sigma d_k y_k} \]

The power of this test, as well as \( F \), can be computed from

\[ P[\Sigma c_k d_k y_k < r \Sigma d_k y_k] \]

For particular cases, this results in a simple expression depending on \( r, c_k \) and \( d_k \) only. For \( n=7 \) and \( 9 \), the power is given in Section IV.

The estimation procedure uses one of these statistics as a preliminary test. We define the estimator of the population mean to be

\[ x = px_1 + qx_2 \] (4.6)

where \( p \) and \( q \) are the proportions of the sample below and above \( x_c \), respectively, \( x_1 \) is the mean of the observations less than
\[ x_0, \text{ and} \]
\[
\tilde{x}_2 = \begin{cases} 
\bar{x}_2 & \text{if } F, \text{ or } T, < a \\
\Sigma w_r x_r & \text{if } F, \text{ or } T, \geq a 
\end{cases}
\]
where \( \bar{x}_2 \) is the mean of the observations greater than \( x_0 \), \( a \) is the significance level and \( \Sigma w_r x_r \) is a linear function of the order statistics of

\[
F(x) = 1 - \exp\left[-\left(\frac{x-a}{b}\right)^{1/3}\right] \quad b > 0 \quad c > 0, \quad (4.7)
\]
\[ x > bc^3 + a \]
i.e. a Weibull distribution with \( \beta = 1/3 \). The estimators \( \Sigma w_r x_r \) are found by minimizing the quadratic form

\[
w'Vw \quad (4.8)
\]
where \( w \) is the vector of weights and \( V \) is the covariance matrix which is assumed constant. Equation 4.8 is minimized subject to unbiasedness, or to 'almost' minimize mean-square error. For example for \( n=7 \) and \( 9 \), we find

\[
\tilde{x}_{21} = -0.332x_1 + 1.00x_2 - 0.62x_3 - 0.079x_4 + 0.12x_5 + 0.29x_6 + 0.32x_7
\]
and the almost minimum mean-square error estimator is

\[
\tilde{x}_{22} = 2.248x_1 - 2.576x_2 - 0.68x_3 + 0.64x_4 + 0.453x_5 + 0.248x_6 + 0.051x_7
\]
For \( n=9 \),

\[
\tilde{x}_{21} = -0.851x_1 + 1.436x_2 + 0.158x_3 - 0.147x_4 - 0.098x_5 + 0.036x_6 + 0.146x_7 + 0.185x_8 + 0.097x_9
\]
and
\[ \hat{x}_{22} = 2.906x_1 - 2.851x_2 - 0.593x_3 + 0.242x_4 + 0.451x_5 + 0.405x_6 + 0.275x_7 + 0.127x_8 + 0.037x_9 \]

The variance of the unbiased estimators is given by,

\[ V(\hat{x}_{21}) = \begin{cases} 271.86 \text{ for } n=7 \\ 184.08 \text{ for } n=9 \end{cases} \]

and

\[ \text{MSE}(\hat{x}_{22}) = \begin{cases} 179.66 \text{ for } n=7 \\ 119.80 \text{ for } n=9 \end{cases} \]

For the sample mean, we have

\[ V(\bar{x}) = \begin{cases} 273.43 \text{ for } n=7 \\ 197.20 \text{ for } n=9 \end{cases} \]

For the estimator in Equation 4.6, it is shown that

\[ V(x^*) = \frac{1}{n} \left[ P\sigma_1^2 + Qk(b,c) + (\mu_1 - \mu_2)^2 PQ \right] \quad (4.9) \]

where \( \sigma_1^2 \) is the variance of the population less than \( x_0 \),
\( \mu_1 \) is the mean of the population less than \( x_0 \),
\( P \) is the proportion of the population less than \( x_0 \),
\( Q = 1-P \),
\( \mu_2 \) is the mean of the remainder of the population, and
\( k(b,c) \) is the asymptotic variance of the estimator \( \bar{x}_2 \)
which is a function of the parameters \( b \) and \( c \), shown in Equation 4.7.

Under the assumption that
\[ S_1 = \bar{x}_2, \]
\[ S_2 = \Sigma w_r x_r, \]
and \( S_3 = F \) or \( T \)

are jointly normally distributed asymptotically, the approximate variance of \( \bar{x}_2 \) is shown to be

\[
V(\bar{x}_2) = P(F < \alpha)[M_2' - (M_1')^2] + P(F \geq \alpha)[M_2 - M_1'^2] \\
+ P(F < \alpha)[(M_1')^2 - \mu_2^2] + P(F \geq \alpha)[M_1^2 - \mu_2^2]
\]  

(4.10)

where \( \bar{x}_2 \) is the sample mean of the observations larger than \( x_0 \),

\( E\hat{w}_r x_r \) is a linear function of order statistics,

\( F \) and \( T \) are the test statistics derived in Section IV,

\( \mu_2 = E\bar{x}_2 \),

\( \alpha \) is the significant value of the statistic, and

\( M_1', M_1, M_2 \) and \( M_2' \) are derived in Section V.

Finally, we consider the finite population case. We use a "super" population argument to derive expressions for the variance of the estimators. We assume the infinite population model described below Equation 4.1. From this population, we draw a sample of size \( N \), the finite population. From the latter, we draw a sample of size \( n \).

We consider the estimator

\[
x^* = \left( \frac{n}{N} \right) \bar{x}_n + \left( 1 - \frac{n}{N} \right) \hat{\mu}
\]

(4.11)

where \( \bar{x}_n \) is the sample mean and \( \hat{\mu} \) is some estimator of the infinite population mean. We find

\[
V(x^*) = \frac{1}{N}(1 - \frac{n}{N})\sigma_x^2 + (1 - \frac{n}{N})\frac{\sigma_\mu^2}{\hat{\mu}}
\]

(4.12)
where $\sigma_x^2$ and $\sigma_{\hat{\mu}}^2$ are infinite population variances. The estimator of Equation 4.12 is

$$v(x^*) = \frac{1}{N}(1 - \frac{N}{N})s_x^2 + (1 - \frac{N}{N})v(\hat{\mu})$$

(4.13)

where $s_x^2$ is the sample mean-square and $v(\hat{\mu})$ is assumed to be an unbiased or consistent estimator of $\sigma_{\hat{\mu}}^2$.

Since $\hat{\mu}$ was general, we may apply these results to the special case $\hat{\mu} = \bar{x}_2$, and Equations 4.12 and 4.13 also hold.
VII. BIBLIOGRAPHY


Downton, Frank. 1966. Linear estimates of parameters in the extreme value distribution. Technometrics 8: 3-17.


VIII. ACKNOWLEDGEMENTS

The author wishes to thank his major professor, Wayne Fuller, for his guidance and assistance throughout the research and preparation of this thesis. It was a rewarding period. Also, the author wishes to thank his wife, Susan, for unflagging encourage and her assistance in the calculations appearing in the thesis.