A FURTHER CONTRIBUTION TO THE THEORY OF SYSTEMATIC STATISTICS

by

Junjiro Ogawa
University of North Carolina

Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States government. Table 3.1 produced under sponsorship of the Office of Ordnance Research.

Institute of Statistics
Mimeograph Series No. 168
April, 1957

A FURTHER CONTRIBUTION TO THE THEORY OF SYSTEMATIC STATISTICS[1]

by

Junjiro Ogawa, Institute of Statistics, University of North Carolina

Introduction. Until 1945 the main interest in the field of statistical estimation seems to have been in the so-called "efficient" estimators. But from the point of view of economy in practical use, it seems reasonable to inquire whether the output of information is worth the input measured in money, man-hours, or otherwise. Thus we may ask whether comparable results could have been obtained by a smaller expenditure. From this standpoint F. Mosteller proposed [2][2] in 1946 the use of order statistics for such purposes on the ground that, however large the sample size, the observations can easily be ordered in magnitude with the help of punch-card equipment. He considered the problem of estimation of the mean and standard deviation of a univariate normal population, and that of estimation of the correlation coefficient of a bivariate normal distribution. Then in 1951 J. Ogawa [4] considered more systematically the problem of estimation of the location and the scale parameter of a population whose density function depends only on these two parameters, and obtained the optimum solutions in some cases for the univariate normal population.

There are many cases in which the samples are by their very nature

---

[1]Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States government. Table 3.1 produced under sponsorship of the Office of Ordnance Research.

[2]The numbers in square brackets refer to the bibliography listed at the end.

ordered in magnitude as for example in a life test of electric lamps or a fatigue test of a certain material. Usually in such cases the population probability density functions are assumed to be exponential. So at least for the exponential distribution estimation and testing of an hypothesis based upon systematic statistics are of great importance, from the point of view of application.

The first two sections of this paper will be devoted to the general theory of systematic statistics from the population whose density function depends only upon the location and scale parameters for sufficiently large values of sample size. Then in § 3 the application of the general theory to the exponential distribution will be given and optimum spacing of the selected sample quantiles and the corresponding best estimators will be tabulated. Finally, in § 4, discussion on testing an hypothesis will be presented.

1. Relative Efficiencies. The fundamental probability distribution of the large sample theory of systematic statistics from the population whose probability density function depends only on the location parameter m and the scale parameter σ is as follows. The limiting frequency function is [4]:

$$h(x_{(n_1)}, x_{(n_2)}, \ldots, x_{(n_k)})$$

$$= C_{n,k} \exp\left[-\frac{n}{2\sigma^2}\left\{ \sum_{i=1}^{k} \frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2 (x_{(n_i)}-m-\sigma u_i)^2 \right.\right.$$

$$\left.\left. - 2\sum_{i=2}^{k} \frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}} (x_{(n_i)}-m-\sigma u_i)(x_{(n_{i-1})}-m-\sigma u_{i-1}) \right\}\right] , \qquad (1.1)$$

where $f(u)$ denotes the standardized frequency function of the population and u_i stands for the λ_i-quantile of the population, i.e.,

$$\lambda_i = \int_{-\infty}^{u_i} f(t)dt, \quad i = 1,2,\dots,k \tag{1.2}$$

and

$$f_i = f(u_i), \quad i = 1,2,\dots,k \ . \tag{1.3}$$

Finally, we put

$$C_{n,k} = (2\pi\sigma^2)^{-k/2} f_1 f_2 \dots f_k [\lambda_1(\lambda_2-\lambda_1)\dots(\lambda_k-\lambda_{k-1})(1-\lambda_k)]^{-1/2} n^{k/2} , \tag{1.4}$$

$$\lambda_0 = 0, \quad \lambda_{k+1} = 1. \tag{1.5}$$

In the first place, we define the relative efficiency of the systematic statistics with respect to parameter to be the ratio of the amount of information in Fisher's sense derived from (1.1) to that derived from the original whole sample.

We shall consider the following two cases separately:

Case I. The case in which the scale parameter σ is known and the location parameter m is unknown. For the sake of convenience, put

$$S = \sum_{i=1}^{k} \frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2 (x_{(n_i)}-m-\sigma u_i)^2$$

$$- 2 \sum_{i=2}^{k} \frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}} (x_{(n_i)}-m-\sigma u_i)(x_{(n_{i-1})}-m-\sigma u_{i-1}), \tag{1.6}$$

where the right-hand side is the same as the expression in the exponent in the density function $h(x_{(n_1)}, x_{(n_2)}, \ldots, x_{(n_k)})$ except for the constant factor $-n/2\sigma^2$, then we have

$$\log h(x_{(n_1)}, \ldots, x_{(n_k)}) = -\frac{n}{2\sigma^2} S + \text{term independent of } m. \qquad (1.7)$$

Thus it follows by differentiation with respect to m that

$$-\frac{\partial^2 \log h}{\partial m^2} = \frac{n}{2\sigma^2}\frac{\partial^2 S}{\partial m^2} = \frac{n}{\sigma^2} K_1 , \qquad (1.8)$$

where

$$K_1 = \sum_{i=1}^{k+1} \frac{(f_i - f_{i-1})^2}{\lambda_i - \lambda_{i-1}} \qquad (1.9)$$

and $f_{k+1} = 0$ and f_0 should be equal to 1 or 0 as the case may be. Hence the amount of information $I_s(m)$ of the systematic statistics with respect to the location parameter m is asymptotically equal to

$$I_s(m) = \mathcal{E}\left(\frac{\partial \log h}{\partial m}\right)^2 = -\mathcal{E}\left(\frac{\partial^2 \log h}{\partial m^2}\right) = \frac{n}{\sigma^2} K_1 . \qquad (1.10)$$

The likelihood function of the original whole sample, considered as a random sample of size n, is

$$L = \sigma^{-n} \prod_{i=1}^{n} f\left(\frac{x_i - m}{\sigma}\right) , \qquad (1.11)$$

hence we have

$$\log L = \sum_{i=1}^{n} \log f\left(\frac{x_i - m}{\sigma}\right) - n \log \sigma , \qquad (1.12)$$

and consequently the amount of information $I_0(m)$ of the original whole sample with respect to the location parameter m is equal to

$$I_0(m) = \mathcal{E}\left(\frac{\partial \log L}{\partial m}\right)^2 = \frac{1}{\sigma^2} \mathcal{E}\left(\sum_{i=1}^{n} \frac{f'(U_i)}{f(U_i)}\right)^2 \qquad (1.13)$$

where $U_i = (X_i - m)/\sigma$, i.e. U stands for the standardized variate.

If $f(t) = (2\pi)^{-1/2} \exp\left\{- t^2/2\right\}$, then

$$I_0(m) = \frac{n}{\sigma^2} , \qquad (1.14)$$

and if $f(t) = e^{-t}$ for $t > 0$, then

$$I_0(m) = \frac{n^2}{\sigma^2} . \qquad (1.15)$$

Thus the relative efficiency of the systematic statistics with respect to the location parameter m is defined as:

$$\eta(m) = \frac{I_s(m)}{I_0(m)} = nK_1 / \mathcal{E}\left(\sum_{i=1}^{n} \frac{f'(U_i)}{f(U_i)}\right)^2 . \qquad (1.16)$$

In particular, for the normal distribution we have

$$\eta(m) = K_1 , \qquad (1.17)$$

and for the exponential distribution we get

$$\eta_e(m) = \frac{1}{n}K_1 = \frac{1}{n} \sim 0 \quad , \tag{1.18}$$

In the case of the one-sided exponential distribution defined by

$$g(x) = \frac{1}{\sigma} e^{-\frac{x-m}{\sigma}} \qquad \text{for } x > m, \tag{1.19}$$

$$= \quad 0 \qquad \text{otherwise,}$$

we know already that $\min\limits_{1 \leq i \leq n} x_i$ i.e. $x_{(1)}$ is the maximum likelihood estimator, and further that this is the uniformly minimum variance estimator of m for all values of n. Thus, the estimator based upon the frequency function (1.1) becomes meaningless.

Case 2. The case in which the location parameter m is known and the scale parameter σ is unknown. In this case

$$\log h = - k \log \sigma - \frac{n}{2\sigma^2} S + \text{a term independent of } \sigma. \tag{1.20}$$

Therefore we get

$$\frac{\partial^2 \log h}{2\sigma^2} = \frac{k}{\sigma^2} - \frac{3n}{\sigma^4} S + \frac{2n}{3\sigma^3}\frac{\partial S}{\partial \sigma} - \frac{n}{2\sigma^2}\frac{\partial^2 S}{\partial \sigma^2} \tag{1.21}$$

and since it can easily be seen that

$$\mathcal{E}\left(\frac{n}{\sigma^2} S\right) = k, \quad \mathcal{E}\left(\frac{\partial S}{\partial \sigma}\right) = 0 \ , \qquad \frac{\partial^2 S}{\partial \sigma^2} = 2K_2 \tag{1.22}$$

where
$$K_2 = \sum_{i=1}^{k+1} \frac{(f_i u_i - f_{i-1} u_{i-1})^2}{\lambda_i - \lambda_{i-1}}, \tag{1.23}$$

we obtain finally the amount of information of the systematic statistics with respect to the scale parameter σ, i.e.

$$I_s(\sigma) = \mathcal{E}\left(\frac{\partial \log h}{\partial \sigma}\right)^2 = \mathcal{E}\left(-\frac{\partial^2 \log h}{\partial \sigma^2}\right) = \frac{2k}{\sigma^2} + \frac{n}{\sigma^2} K_2 , \tag{1.24}$$

or

$$I_s(\sigma) = \frac{n}{\sigma^2}\left(K_2 + \frac{2k}{n}\right) \sim \frac{n}{\sigma^2} K_2 \tag{1.25}$$

On the other hand we have

$$\frac{\partial \log h}{\partial \sigma} = -\frac{n}{\sigma} - \frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{(x_i - m) f'\left(\frac{x_i - m}{\sigma}\right)}{f\left(\frac{x_i - m}{\sigma}\right)} \tag{1.26}$$

Hence by making use of the general relation

$$\mathcal{E}\left(\frac{Uf'(U)}{f(U)}\right) = 1, \tag{1.27}$$

we obtain

$$I_0(\sigma) = E\left(\frac{\partial \log L}{\partial \sigma}\right)^2 = E\left[\frac{1}{\sigma^2}\left(n + \sum_{i=1}^{n} U_i f'(U_i)/f(U_i)\right)^2\right]$$

$$= \frac{n}{\sigma^2}\left\{E(Uf'(U)/f(U))^2 - 1\right\} . \tag{1.28}$$

In particular, for the normal distribution we have

$$I_0(\sigma) = \frac{2n}{\sigma^2} \tag{1.29}$$

and for the exponential distribution we have

$$I_0(\sigma) = \frac{n}{\sigma^2} \tag{1.30}$$

Thus we obtain the relative efficiency of the systematic statistics with respect to the scale parameter σ as follows:

$$\eta(\sigma) = \frac{I_s(\sigma)}{I_s(\sigma)} = \frac{\frac{2k}{n} + K_2}{E\left(\frac{Uf'(U)}{f(U)}\right)^2 - 1} \sim \frac{K_2}{E\left(\frac{Uf'(U)}{f(U)}\right)^2 - 1}, \tag{1.31}$$

In particular, for the normal distribution we have

$$\eta(\sigma) = \frac{1}{2} \quad K_2 \tag{1.32}$$

and for the exponential distribution we have

$$\eta(\sigma) = K_2 \ . \tag{1.33}$$

2. <u>The Best Linear Unbiased Estimators of the Unknown Parameters Based upon the Selected Sample Quantiles for Sufficiently Large Sample Size n.</u>

In the circumstance now under consideration, the basic distribution is given by (1.1), and the unknown parameters are m and σ. Since the distribution given in (1.1) is a k-dimensional normal distriubtion, we can apply the extended Gauss-Markov theorem on least squares [3] in order to find the best linear unbiased estimators which are asymptotically efficient estimators. Hence, as far as the large sample problems are concerned there are no other estimators which are more efficient than the best linear unbiased ones.

Case I. The case in which the scale parameter σ is known and the location parameter m is unknown.

In this case, we can find the best linear unbiased estimator $\hat{m}$ of the location parameter m by solving the single normal equation

$$\left.\frac{\partial S}{\partial m}\right|_{m=\hat{m}} = 0, \tag{2.1}$$

which turns out to be

$$K_1\hat{m} = X - K_3\sigma . \tag{2.2}$$

where

$$X = \sum_{i=1}^{k+1} \frac{(f_i - f_{i-1})(f_i x_{(n_i)} - f_{i-1} x_{(n_{i-1})})}{\lambda_i - \lambda_{i-1}} \tag{2.3}$$

and

$$K_3 = \sum_{i=1}^{k+1} \frac{(f_i - f_{i-1})(f_i u_i - f_{i-1} u_{i-1})}{\lambda_i - \lambda_{i-1}} \tag{2.4}$$

Hence the best linear unbiased estimator $\hat{m}$ of m is

$$\hat{m} = \frac{1}{K_1} X - \sigma \cdot \frac{K_3}{K_1} , \tag{2.5}$$

and it can easily be seen after small calculation that

$$\mathrm{Var}(\hat{m}) = \frac{\sigma^2}{n} \frac{1}{K_1} \quad . \tag{2.6}$$

From (2.6), we can see that the best linear unbiased estimator $\hat{m}$ is an "efficient estimator" (so to speak) from the point of view of the relative efficiency given in the preceding section.

If in particular the frequency function f(t) of the population is symmetric with respect to the origin, for example the normal distribution, and the spacing of the selected sample quantiles $x_{(n_1)}, x_{(n_2)}, \ldots, x_{(n_k)}$ is also symmetric, i.e.,

$$n_i + n_{k-i+1} = n; \quad i = 1,2,\ldots,k \tag{2.7}$$

or in terms of λ_i

$$\lambda_i + \lambda_{k-i+1} = 1; \quad i = 1,2,\ldots,k, \tag{2.8}$$

then we have

$$u_i + u_{k-i+1} = 0; \quad i = 1,2,\ldots,k \tag{2.9}$$

and

$$f_i = f_{k-i+1}; \quad i = 1,2,\ldots,k. \tag{2.10}$$

In such a case, it follows clearly that

$$K_3 = 0. \tag{2.11}$$

Hence we have

$$\hat{m} = \frac{1}{K_1} X, \tag{2.12}$$

and

$$\text{Var}(\hat{m}) = \frac{\sigma^2}{n} \frac{1}{K_1} \quad . \tag{2.13}$$

<u>Case 2</u>. <u>The case in which the location parameter m is known and the scale parameter σ is unknown</u>.

In a quite similar manner, we can find the best linear unbiased estimator $\hat{\sigma}$ of the scale parameter σ by solving the normal equation

$$\left.\frac{\partial S}{\partial \sigma}\right|_{\sigma = \hat{\sigma}} = 0 \ . \tag{2.14}$$

This comes to be

$$K_2 \hat{\sigma} = Y - K_3 m \ , \tag{2.15}$$

where

$$Y = \sum_{i=1}^{k+1} \frac{(f_i u_i - f_{i-1} u_{i-1})(f_i x_{(n_i)} - f_{i-1} x_{(n_{i-1})})}{\lambda_i - \lambda_{i-1}} \quad . \tag{2.16}$$

Hence we get

$$\hat{\sigma} = \frac{1}{K_2} Y - m \frac{K_3}{K_2} \tag{2.17}$$

and

$$\mathrm{Var}(\hat{\sigma}) = \frac{\sigma^2}{n} \; \frac{1}{K_2} \quad . \tag{2.18}$$

In the particular case in which f(t) and the spacing of the selected sample quantiles are symmetric we get as before

$$\hat{\sigma} = \frac{1}{K_2} Y \tag{2.19}$$

and the variance of $\hat{\sigma}$ is the same as before.

3. Determination of the Optimum Spacings for the Estimation of the Scale Parameter of the One-Parameter Exponential Distribution Based on the Selected Sample Quantiles for Sufficiently Large Sample Size n. [5]

In [4], the author developed the general theory of estimation of the location and the scale parameters based upon the selected sample quantiles for sufficiently large sample size n and determined the optimum spacings in the case of the normal population.

Owing to its importance in practical applications, optimum spacings are calculated herein for the estimation of the scale parameter σ of the exponential distribution whose density function is given as follows:

$$g(x) = \begin{cases} \frac{1}{\sigma} e^{-\frac{x}{\sigma}}, & \text{if } x > 0 \\ 0 & \text{, otherwise} \end{cases} \tag{3.1}$$

Suppose we are given an ordered sample of size n, where n is sufficiently large. In other words, the sample size n is large enough so that the conclusions drawn making use of the limit distribution are valid with enough accuracy. For given k real numbers such that

$$0 < \lambda_1 < \lambda_2 < \ldots < \lambda_k < 1, \qquad (3.2)$$

we select k sample λ_i-quantiles for $i = 1,2,\ldots,k$, i.e., k order statistics

$$x_{(n_1)}, x_{(n_2)}, \ldots, x_{(n_k)},$$

where

$$n_i = [n\lambda_i] + 1, \; i = 1,\ldots,k \qquad (3.3)$$

and the symbol $[n\lambda_i]$ stands for the greatest integer not exceeding $n\lambda_i$.

Furthermore, let the λ_i-quantile of the standardized exponential distribution, whose density is given by

$$f(x) = \begin{cases} e^{-x}, & \text{if } x > 0 \\ 0\;, & \text{otherwise} \end{cases} \qquad (3.4)$$

be u_i, and that of (3.1) be x_i, then it is clear that

$$x_i = u_i \sigma\;, \; i = 1,2,\ldots,k. \qquad (3.5)$$

The limiting joint distribution of the k sample quantiles $x_{(n_1)},\ldots,x_{(n_k)}$ has the density function

$$h(x_{(n_1)}, x_{(n_2)}, \ldots, x_{(n_k)})$$

$$= C_{n,k} \exp\left[-\frac{n}{2\sigma^2} \left\{ \sum_{i=1}^{k} \frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2 (x_{(n_i)}-u_i\sigma)^2 \right.\right.$$

$$\left.\left. - 2 \sum_{i=2}^{k} \frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}} (x_{(n_i)}-u_i\sigma)(x_{(n_{i-1})}-u_{i-1}\sigma) \right\} \right], \quad (3.6)$$

where

$$f_i = e^{-u_i}, \quad i = 1,2,\ldots,k \quad (3.7)$$

$$\lambda_i = 1-e^{-u_i}, \quad i = 1,2,\ldots,k . \quad (3.7)$$

Hence we get

$$\log h = -\frac{k}{2} \log \sigma^2 - \frac{n}{2\sigma^2} S + \text{term which is independent of } \sigma \quad (3.8)$$

where

$$S = \sum_{i=1}^{k} \frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2 (x_{(n_i)}-u_i\sigma)^2$$

$$- 2 \sum_{i=2}^{k} \frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}} (x_{(n_i)}-u_i\sigma)(x_{(n_{i-1})}-u_{i-1}\sigma), \quad (3.9)$$

Thus we get the amount of information of the systematic statistics with respect to the scale parameter σ in this case

$$I_s(\sigma) = E(-\frac{\partial^2 \log h}{\partial \sigma^2}) = \frac{2k}{\sigma^2} + \frac{n}{\sigma^2} K_2 , \tag{3.10}$$

where

$$K_2 = \sum_{i=1}^{k+1} \frac{(f_i u_i - f_{i-1} u_{i-1})^2}{\lambda_i - \lambda_{i-1}} = \sum_{i=1}^{k+1} \frac{(e^{-u_i} u_i - e^{-u_{i-1}} u_{i-1})^2}{e^{-u_{i-1}} - e^{-u_i}} . \tag{3.11}$$

Consequently we get the relative efficiency of the systematic statistics with respect to σ as

$$\eta(\sigma) = K_2 . \tag{3.12}$$

The best linear unbiased estimator $\hat{\sigma}$ of σ is

$$\hat{\sigma} = \frac{1}{K_2} Y, \tag{3.13}$$

where

$$Y = \sum_{i=1}^{k+1} \frac{(e^{-u_i} u_i - e^{-u_{i-1}} u_{i-1})(e^{-u_i} x_{(n_i)} - e^{-u_{i-1}} x_{(n_{i-1})})}{e^{-u_{i-1}} - e^{-u_i}} \tag{3.14}$$

and this can be written as

$$Y = \sum_{i=1}^{k} a_i x_{(n_i)} \tag{3.15}$$

where

$$a_i = e^{-u_i} \left\{ \frac{e^{-u_i}u_i - e^{-u_{i-1}}u_{i-1}}{e^{-u_{i-1}} - e^{-u_i}} - \frac{e^{-u_{i+1}}u_{i+1} - e^{-u_i}u_i}{e^{-u_i} - e^{-u_{i+1}}} \right\}, \quad (3.16)$$

The coefficients a_i for the optimum spacings are calculated in Table 3.1. From (3.13) and (3.15) we have

$$\hat{\sigma} = \sum_{i=1}^{k} a_i' x_{(n_i)} \quad (3.17)$$

where

$$a_i' = a_i / K_2 \ , \ i = 1,2,\ldots,k, \quad (3.18)$$

Although it may appear to be more convenient to tabulate a_i' than a_i, for $k = 10$, for instance, there are ten rounding errors in using (3.17), while the calculation of $\hat{\sigma}$ by the formula

$$\hat{\sigma} = \sum_{i=1}^{k} a_i x_{(n_i)} / K_2 \quad (3.19)$$

has the error obtained by only one rounding process.

The required optimum spacing is the set $(\lambda_1,\lambda_2,\ldots,\lambda_k)$, or equivalently the set of values $(u_1,u_2,\ldots,u_k)$ which gives the maximum value of the relative efficiency K_2. We can obtain the optimum spacings step by step as follows:

(i) The case where k = 1: in this case it is easily seen that

$$K_2 = \frac{u_1^2}{e^{u_1}-1} \quad . \tag{3.20}$$

which has only one maximum $K_2 = 0.6471$ at $u_1 = 1.59$, and the corresponding probability is $\lambda_1 = 1 - e^{-1.59} = 0.7961$.

(ii) The case where k = 2: put $u_2 = u_1 + x$, then we have

$$K_2 = e^{-u_1}\left(\frac{u_1^2}{e^{u_1}-1} + u_1^2 + \frac{x^2}{e^x-1}\right) \quad , \tag{3.21}$$

and this function of x and u_1 is maximized at the point

$$u_1 = 1.02, \qquad x = 1.59$$

i.e.,

$$u_1 = 1.02, \qquad u_2 = 2.61$$

and the maximum value of K_2 is

$$K_2 = 0.8203$$

(iii) The case where k = 3: put

$$u_2 = u_1 + x, \quad u_3 = u_2 + y = u_1 + x + y,$$

then it follows after some calculations that

$$K_2 = e^{-u_1} \left[\frac{u_1^2}{e^{u_1}-1} + u_1^2 + e^{-x} \left\{ \frac{x^2}{e^x-1} + x^2 + \frac{y^2}{e^y-1} \right\} \right] . \quad (3.22)$$

It is easily seen that this function of x,y and u_1 is maximized at the point

$$u_1 = 0.75,\ x = 1.02,\ y = 1.59.$$

Hence we get the optimum spacing in this case as follows:

$$u_1 = 0.75,\ u_2 = 1.77,\ u_3 = 3.36,$$

and the correspoinding maximum value of K_2 is

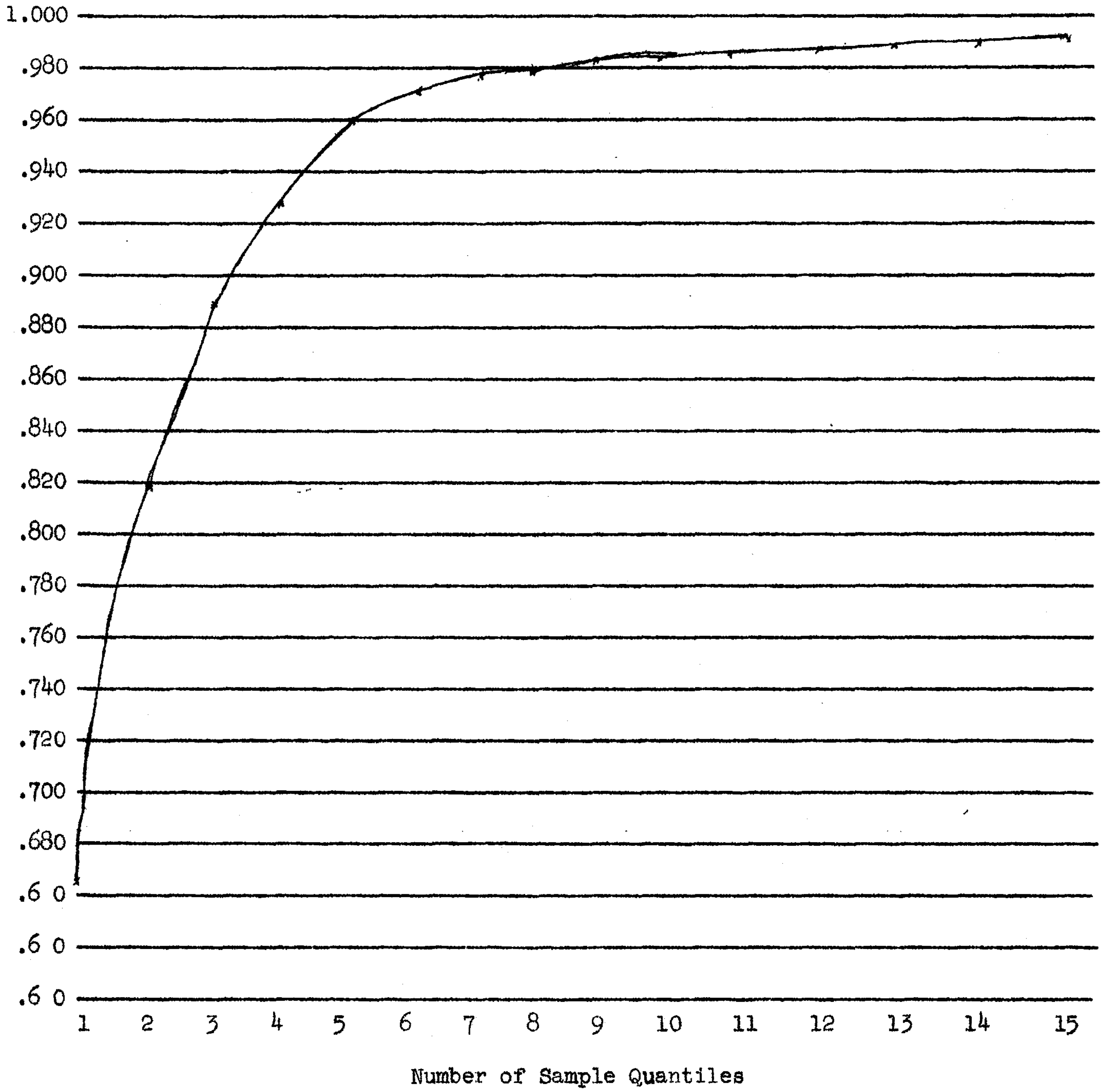
$$K_2 = 0.8905. \quad (3.23)$$

In a similar way, the following results in Table 3.1 have been calculated. In Table 3.1, u_i stands for the λ_i-quantile of the standardized population. The bottom row of the table which representing the relative efficiencies K_2 has been graphed as Fig. 1 to show the increasing rate of relative efficiencies against the number of sample quentiles which has been selected . It will be seen from Fig. 1 that after k = 10, the gain in relative efficiency is not appreciable.

Table 3.1 Optimum Spacings for Estimates of Relative Efficiencies and the Coefficients of Best Estimates*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.59	1.02	0.75	0.61	0.50	0.43	0.37	0.33	0.30	0.27	0.25	0.23	0.21	0.20	0.19
1	.79607	.63941	.52763	.45665	.39347	.34949	.30927	.28108	.25918	.23662	.22120	.20547	.18942	.18127	.17304
1	.40731	.42835	.39974	.36098	.33051	.29900	.27352	.24989	.23224	.21644	.20181	.19003	.17766	.16759	.16096
2		2.61	1.77	1.36	1.11	0.93	0.80	0.70	0.63	0.57	0.52	0.48	0.44	0.41	0.39
2		.92647	.82967	.74334	.67044	.60545	.55067	.50341	.46741	.43447	.40548	.38122	.35596	.33635	.32294
2		.14687	.20233	.21719	.21896	.21499	.20654	.19671	.18514	.17729	.16860	.16032	.15409	.14552	.13861
3			3.36	2.38	1.86	1.54	1.30	1.13	1.00	0.90	0.82	0.75	0.69	0.64	0.60
3			.96527	.90745	.84433	.78562	.72747	.67697	.63212	.59343	.55956	.52762	.49842	.47271	.45119
3			.06938	.10994	.13173	.14242	.14850	.14845	.14569	.14135	.13803	.13398	.13001	.12609	.12031
4				3.97	2.88	2.29	1.91	1.63	1.43	1.27	1.15	1.05	0.96	0.89	0.83
4				.98113	.94387	.89873	.85192	.80407	.76069	.71917	.68336	.65006	.61711	.58934	.56395
4				.03770	.06668	.08571	.09839	.10677	.10999	.11122	.11012	.10971	.10856	.10644	.10430
5					4.47	3.31	2.66	2.24	1.93	1.70	1.52	1.38	1.26	1.16	1.08
5					.98855	.96348	.93005	.89354	.85485	.81732	.78129	.74842	.71635	.68651	.66040
5					.02286	.04394	.05919	.07074	.07910	.08396	.08661	.08748	.08887	.08890	.08803
6						4.90	3.68	2.99	2.54	2.20	1.95	1.75	1.59	1.46	1.35
6						.99255	.97478	.94971	.92113	.88920	.85773	.82623	.79607	.76776	.74076
6						.01487	.02996	.04255	.05239	.06037	.06539	.06880	.07093	.07281	.07350
7							5.27	4.01	3.29	2.81	2.45	2.18	1.96	1.79	1.65
7							.99486	.98187	.96275	.93980	.91371	.88696	.85914	.83304	.80795
7							.01027	.02154	.03152	.04000	.04702	.05195	.05578	.05802	.06019
8								5.60	4.31	3.56	3.06	2.68	2.39	2.16	1.98
8								.99630	.98657	.97156	.95311	.93144	.90837	.88467	.86193
8								.00739	.01596	.02407	.03115	.03736	.04213	.04570	.04800
9									5.90	4.58	3.81	3.29	2.89	2.59	2.35
9									.99726	.98975	.97785	.96275	.94442	.92498	.90463
9									.00547	.01218	.01874	.02475	.03028	.03447	.03778
10										6.17	4.83	4.04	3.50	3.09	2.78
10										.99791	.99201	.98240	.96980	.95450	.93796
10										.00418	.00948	.01489	.02006	.02479	.02851
11											6.42	5.06	4.25	3.70	3.28
11											.99837	.99365	.98574	.97528	.96237
11											.00325	.00754	.01207	.01643	.02051
12												6.65	5.22	4.45	3.89
12												.99871	.99486	.98832	.97956
12												.00258	.00611	.00989	.01358
13													6.86	5.47	4.64
13													.99895	.99579	.99034
13													.00210	.00500	.00817
14														7.06	5.66
14														.99914	.99652
14														.00172	.00414
15															7.25
15															.99929
15															.00142
	.64761	.82026	.89049	.92691	.94757	.96056	.96926	.97537	.97982	.98316	.98374	.98739	.98939	.99071	.99180

Fig. 1. The curve of relative efficiencies against the number of selected sample quantiles to be used



Example: As an illustration of the estimation procedure, we shall calculate the estimate of the standard deviation of the time intervals in days between explosions in mines, involving more than 10 men killed from 6th December 1875 to 29th May 1951. The data are from Maguire, Pearson and Wynn [1]. It should be noted that this example may not be the best to point out the value of this method because the sample size (n = 109) is not so large that the labor necessary for the classical estimation is almost comparable to that necessary for our estimate. This example should be regarded as an illustration of the calculating procedure itself. There are cases, which are important in applications such as in life testing of the electric lamps and in fatigue tests of certain material, where the samples are ordered in magnitudes in natural order. In such a case the procedure here may be of great help in getting quickly the estimate of σ, especially for a large bulk of data.

The Table 1 of B. A. Maguire, E. S. Pearson and A. H. A. Wynn is reproduced here with observations arranged in order of magnitude as follows:

Table 3.2. Time intervals in days between explosions in mines, involving more than 10 men killed from 6 Dec. 1875 to 29 May 1951. (B. A. Maguire, E. S. Pearson and A. H. A. Wynn)

Order	Observation	Order	Observation	Order	Observation
1	1	46	113	91	354
2	4	47	114	92	361
3	4	48	120	93	364
4	7	49	120	94	369
5	11	50	123	95	378
6	13	51	124	96	390
7	15	52	129	97	457
8	15	53	131	98	467
9	17	54	137	99	498
10	18	55	145	100	517
11	19	56	151	101	566
12	19	57	156	102	644
13	20	58	171	103	745
14	20	59	176	104	871
15	22	60	182	105	1205
16	23	61	188	106	1312
17	28	62	189	107	1357
18	29	63	195	108	1613
19	31	64	203	109	1630
20	32	65	208		
21	36	66	215		
22	37	67	217		
23	47	68	217		
24	48	69	217		
25	49	70	224		
26	50	71	228		
27	54	72	233		
28	54	73	255		
29	55	74	271		
30	58	75	275		
31	59	76	275		
32	59	77	275		
33	61	78	286		
34	61	79	286		
35	66	80	312		
36	72	81	312		
37	72	82	315		
38	75	83	326		
39	78	84	326		
40	78	85	329		
41	81	86	330		
42	93	87	336		
43	96	88	338		
44	99	89	345		
45	108	90	348		

For k = 10, we can see from Table 3.1 that

$$n_1 = [n\lambda_1] + 1 = [109 \times .23662] + 1 = 26$$

$$n_2 = [n\lambda_2] + 1 = [109 \times .43447] + 1 = 48$$

$$n_3 = [n\lambda_3] + 1 = [109 \times .59343] + 1 = 65$$

$$n_4 = [n\lambda_4] + 1 = [109 \times .71917] + 1 = 79$$

$$n_5 = [n\lambda_5] + 1 = [109 \times .81732] + 1 = 90$$

$$n_6 = [n\lambda_6] + 1 = [109 \times .88920] + 1 = 97$$

$$n_7 = [n\lambda_7] + 1 = [109 \times .93980] + 1 = 103$$

$$n_8 = [n\lambda_8] + 1 = [109 \times .97156] + 1 = 106$$

$$n_9 = [n\lambda_9] + 1 = [109 \times .98975] + 1 = 108$$

$$n_{10} = [n\lambda_{10}] + 1 = [109 \times .99791] + 1 = 109$$

Thus, we used the ordered observations as follows:

Order(n_i)	Observations($x_{(n_i)}$)	Coefficients(a_i)
26	50	.21646
48	120	.17729
65	208	.14135
79	291	.11122
90	348	.08396
97	459	.06037
103	745	.04000
106	1312	.02407
108	1613	.01218
109	1630	.00418

$$K_2 = .98316.$$

Thus we calculate the expression

$$\sum_{i=1}^{10} a_i x_{(n_i)} = 238.50937 ,$$

and consequently we have

$$\hat{\sigma} = \frac{1}{K_2} \sum_{i=1}^{10} a_i x_{(n_i)} = 242.59466 .$$

If we compare the estimate $\sigma = 241$ which was calculated using all of the sample by the classical method with the above, there is quite good agreement.

4. Testing Statistical Hypothesis. In order to test the hypothesis

$$H_0: \sigma = \sigma_0 , \tag{4.1}$$

if we start with the frequency function given by (1.1), then this can be done as follows: if the null-hypothesis H_0 is true, then

$$\frac{n}{\sigma_0^2} \left\{ \sum_{i=1}^{k} \frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2 (x_{(n_i)}-u_i\sigma_0)^2 \right.$$

$$\left. -2 \sum_{i=2}^{k} \frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}} (x_{(n_i)}-u_i\sigma_0)(x_{(n_{i-1})}-u_{i-1}\sigma_0) \right\} \tag{4.2}$$

is distributed according to the χ^2-distribution with degrees of freedom k. This comes out as

$$\frac{n}{\sigma_0^2}\left\{\sum_{i=1}^{k+1}\frac{(f_i x(n_i)-f_{i-1}x(n_{i-1}))^2}{\lambda_i-\lambda_{i-1}}+K_2(\sigma_0^2-2\sigma_0\hat{\sigma})\right\}\ . \quad (4.3)$$

The minumum value S_0 of S under the variation of σ is given by

$$S_0=\sum_{i=1}^{k+1}\frac{(f_i x(n_i)-f_{i-1}x(n_{i-1}))^2}{\lambda_i-\lambda_{i-1}}-K_2\hat{\sigma}^2 \quad (4.4)$$

and $\frac{n}{\sigma_0^2} S_0$ is distributed according to the χ^2-distribution with degrees of freedom k-1, provdded H_0 is true.

Hence

$$\frac{n}{\sigma^2}\ K_2(\hat{\sigma}-\sigma_0)^2 \quad (4.5)$$

is independent of S_0 and is distributed according to the χ^2-distribution with degree of freedom 1. Thus the statistic

$$t=\sqrt{k-1}\ \frac{\sqrt{K_2}(\hat{\sigma}-\sigma_0)}{\sqrt{S_0}} \quad (4.6)$$

follows the Student's t-distribution with degrees of freedom k-1 if the null-hypothesis H_0 is true.

If the null-hypothesis H_0 is not true, and an alternative hypothesis $H: \sigma = \sigma(\neq \sigma_0)$ is true, then the distribution of t in (4.6) follows the non-central t distribution with non-centrality parameter

$$\delta = \sqrt{K_2}\ (1 - \frac{\sigma_0}{\sigma})\ , \qquad (4.7)$$

and the power function of the t-test is an increasing function of the absolute value of δ. Hence it is reasonable to select sample quantiles the spacing of which makes K_2 maximum, i.e. optimum spacing for estimation purpose is also optimum for testing purpose. Thus we can use Table 3.1 also for testing purposes.

Finally, it should be noted that the confidence interval of σ with confidence coefficient $100(1-\alpha)$ per cent it given by

$$\hat{\sigma} - t_{k-1}(100\alpha)\sqrt{\frac{S_0}{(k-1)K_2}} < \sigma < \hat{\sigma} + t_{k-1}(100\alpha)\sqrt{\frac{S_0}{(k-1)K_2}} \qquad (4.8)$$

where $t_{k-1}(100\alpha)$ stands for the 100α per cent point of the t-distribution with degrees of freedom k-1.

<u>Example:</u> We shall calculate the 95% confidence interval for σ in the example of the preceding section, as an illustration of the procedure explained in this section. We consider case $k = 10$. Then we have

$$S_0 = \sum_{i=1}^{11} \frac{(f_i x_{(n_i)} - f_{i-1} x_{(n_{i-1})})^2}{\lambda_i - \lambda_{i-1}} - K_2 \hat{\delta}^2$$

$$= \sum_{i=1}^{11} \frac{(f_i x_{(n_i)} - f_{i-1} x_{(n_{i-1})})^2}{\lambda_i - \lambda_{i-1}} - \frac{(\sum_1^{10} a_i x_{(n_i)})^2}{K_2} , \qquad (4.9)$$

and calculation will be executed as shown in the following table. The 95°/o confidence interval calculated below seems to be somewhat wide, which may be due to the smallness of the sample size.

Calculating Scheme for S_0

	λ_0		f_0						
		$\lambda_1-\lambda_0$				$f_1x(n_1)$	$\frac{f_1x(n_1)}{\lambda_1}$	$\frac{f_1^2x(n_1^2)}{\lambda_1}$	
a_1	λ_1	.23662	f_1	$x(n_1)$	$f_1x(n_1)$	38.16900	16.13093		$a_1x(n_1)$
.21644	.23662	$\lambda_2-\lambda_1$	.76338	50	38.16900	$f_2x(n_2)-f_1x(n_1)$	$\frac{f_2x(n_2)-f_1x(n_1)}{\lambda_2-\lambda_1}$	$\frac{(f_2x(n_2)-f_1x(n_1))^2}{\lambda_2-\lambda_1}$	
a_2	λ_2	.19785	f_2	$x(n_2)$	$f_2x(n_2)$	29.69460	15.00864		$a_2x(n_2)$
.17729	.43447	$\lambda_3-\lambda_2$	.56553	120	67.86360	$f_3x(n_3)-f_2x(n_2)$	$\frac{f_3x(n_3)-f_2x(n_2)}{\lambda_3-\lambda_2}$	$(f_3x(n_3)-f_2x(n_2))^2$	
a_3	λ_3	.15896	f_3	$x(n_3)$	$f_3x(n_3)$	16.70296	10.50765		$a_3x(n_3)$
.14135	.59343	$\lambda_4-\lambda_3$	.40657	208	04.56656	$f_4x(n_4)-f_3x(n_3)$	$\frac{f_4x(n_4)-f_3x(n_3)}{\lambda_4-\lambda_3}$	$\frac{(f_4x(n_4)-f_3x(n_3))^2}{\lambda_4-\lambda_3}$	
a_4	λ_4	.12574	f_4	$x(n_4)$	$f_4x(n_4)$	-2.84503	-2.26263		$a_4x(n_4)$
.11122	.71917	$\lambda_5-\lambda_4$	.28083	291	81.72153	$f_5x(n_5)-f_4x(n_4)$	$\frac{f_5x(n_5)-f_4x(n_4)}{\lambda_5-\lambda_4}$	$\frac{(f_5x(n_5)-f_4x(n_4))^2}{\lambda_5-\lambda_4}$	
a_5	λ_5	.09815	f_5	$x(n_5)$	$f_5x(n_5)$	-18.14889	-184.90973		$a_5x(n_5)$
.08396	.81732	$\lambda_6-\lambda_5$	.18268	348	63.57264	$f_6x(n_6)-f_5x(n_5)$	$\frac{f_6x(n_6)-f_5x(n_5)}{\lambda_6-\lambda_5}$	$\frac{(f_6x(n_6)-f_5x(n_5))^2}{\lambda_6-\lambda_5}$	

Calculating Scheme for S_0 (Continued)

a_6	λ_6	.07188	f_6	$x(n_6)$	$f_6x(n_6)$	-12.93704	-179.98108		$a_6x(n_6)$
.06037	.88920	$\lambda_7-\lambda_6$	.11080	457	50.63560	$f_7x(n_7)-f_6x(n_6)$	$\frac{f_7x(n_7)-f_6x(n_6)}{\lambda_7-\lambda_6}$	$\frac{(f_7x(n_7)-f_6x(n_6))^2}{\lambda_7-\lambda_6}$	
a_7	λ_7	.05060	f_7	$x(n_7)$	$f_7x(n_7)$	-5.78660	-114.35969		$a_7x(n_7)$
.04000	.93980	$\lambda_8-\lambda_7$	.06020	745	44.84900	$f_8x(n_8)-f_7x(n_7)$	$\frac{f_8x(n_8)-f_7x(n_7)}{\lambda_8-\lambda_7}$	$\frac{(f_8x(n_8)-f_7x(n_7))^2}{\lambda_8-\lambda_7}$	
a_8	λ_8	.03176	f_8	$x(n_8)$	$f_8x(n_8)$	-7.16836	-225.70403		$a_8x(n_8)$
.02407	.97156	$\lambda_9-\lambda_8$	.02872	1312	37.68064	$f_9x(n_9)-f_8x(n_8)$	$\frac{f_9x(n_9)-f_8x(n_8)}{\lambda_9-\lambda_8}$	$\frac{(f_9x(n_9)-f_8x(n_8))^2}{\lambda_9-\lambda_8}$	
a_9	λ_9	.01819	f_9	$x(n_9)$	$f_9x(n_9)$	-21.14739	-1162.58329		$a_9x(n_9)$
.01218	.98975	$\lambda_{10}-\lambda_9$	.01025	1613	16.53325	$f_{10}x(n_{10})-f_9x(n_9)$	$\frac{f_{10}x(n_{10})-f_9x(n_9)}{\lambda_{10}-\lambda_9}$	$\frac{(f_{10}x(n_{10})-f_9x(n_9))^2}{\lambda_{10}-\lambda_9}$	
a_{10}	λ_{10}	.00816	f_{10}	$x(n_{10})$	$f_{10}x(n_{10})$	-13.12655	-1608.64563		$a_{10}x(n_{10})$
.00418	.99791	$\lambda_{11}-\lambda_{10}$	.00209	1630	3.40670	$-f_{10}x(n_{10})$	$\frac{-f_{10}x(n_{10})}{1-\lambda_{10}}$	$\frac{f_{10}^2x^2(n_{10})}{1-\lambda_{10}}$	
	λ_{11}	.00209				-3.40670	-1630.00000		
							Total	60461.82398	238.50937

$$S_0 = \sum_{i=1}^{11} \frac{(f_i x(n_i) - f_{i-1} x(n_{i-1}))^2}{\lambda_i - \lambda_{i-1}} - \frac{(\sum_{i=1}^{10} a_i x(n_i))^2}{K_2}$$

$$= 2600.72348$$

$$t_9(5) \cdot \sqrt{\frac{S_0}{(k-1)K_2}} = 2.262 \times \sqrt{\frac{2600.72348}{9 \times .98316}} = 38.77971$$

REFERENCES

[1] Maguire, B. A., Pearson, E. S. and Wynn, A. H. A., "The time intervals between industrial accidents," Biometrika, Vol. 39 (1957). pp. 168-180.

[2] Mosteller, F., "On some useful 'inefficient' statistics," Annals of Mathematical Statistics, Vol. 17(1948). pp. 377-408.

[3] Neyman, J and David, F. N., "Extension of the Markoff's theorem on least squares," Statistical Research Memoirs, Vol. 1(1936). pp. 105-116.

[4] Ogawa, J., "Contributions to the theory of systematic statistics, I," Osaka Mathematical Journal, Vol 3(1951). pp. 175-213.

[5] Ogawa, J., Determination of optimum spacings for the estimation of the scale parameter of the exponential distribution based on sample quantiles. Typewritten manuscript at, the Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina

ACKNOWLEDGEMENT

The author wishes to express his thanks to Professor B. G. Greenberg and Dr. A. E. Sarhan of the Department of Biostatistics, School of Public Health, University of North Carolina for their kind encouragements, essential suggestions and discussions given to him while this paper was prepared. The author's thanks are also due to Professor Harold Hotelling and Dr. David B. Duncan of the University of North Carolina for their valuable comments on this paper.