Estimators of σ from a Normal Population

S. Maghsoodloo and Dilcu B. Helvaci

Abstract

This article discusses the relative efficiencies of nearly all possible estimators of a normal population standard deviation $\sigma = \sigma_X$ for M = 1 and M > 1 subgroups. As has been fairly known in statistical literature, we found that at M = 1 subgroup the maximum likelihood estimator (MLE) of $\sigma = \sigma_X$ is the most efficient estimator. However, for M > 1 subgroup the maximum likelihood estimator of σ is the least efficient estimator of σ_X. Further, for M > 1 subgroups of differing sample sizes (the case of unbalanced design) it is shown that the unbiased estimator of $\sigma = \sigma_X$ is given by $\hat{\sigma}_p = S_p/c_5$, where c_5 is a Quality-Control constant discussed in section 5, and S_p is the sample pooled estimator of σ for M > 1 subgroups of differing sizes. As a result, some slight modifications to control limits of S-& $\bar{x}$-charts and an S^2-chart are recommended.

Key Words: Population and Sample Standard Deviations, Mean Square Error, Relative-Efficiencies, M > 1 Subgroups, Pooled Unbiased Estimators.

Estimators of σ from a Normal Population

1. Historical Background

Throughout this article we are assuming that the underlying distribution (or population) is normal with the pdf (probability density function) given by the probability law

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\frac{x-\mu}{\sigma})^2/2} \quad , \quad -\infty < x < \infty$$

The above density was first discovered by Abraham de Moivre (a French mathematician) in 1738 as the limiting distribution of the binomial pmf (probability mass function) and as such he did not study all its properties. During 1809, Marquis P. S. de Laplace published the Central Limit Theorem (CLT) that stated the limiting sampling distribution of the sample mean $\bar{x}$ is of the form $f(\bar{x}; \mu, \sigma/\sqrt{n})$, n being the sample size. Then, in 1810 Carl F. Gauss published his *Theoria Motus* in which he discussed the statistical properties of the above f(x; μ, σ). For more complete historical details, the reader should refer to Hogg and Tanis (2010), eighth edition, pp. 268-270, and Johnson, Kotz, and Balakrishnan (1994), Vol. 1, pp.85-88, where we have extracted nearly all of the above historical background. As stated by Maurice G. Kendall and Alan Stuart (1963), second edition, Vol. 1, p. 135 footnote, and we exactly quote:

"The description of the distribution as the "normal," due to K. Pearson, is now almost universal by English writers. Continental writers refer to it variously as the second law of Laplace, the Laplace distribution, the Gauss distribution, the Laplace-Gauss distribution and the Gauss-Laplace distribution. As an approximation to the binomial it was reached by Demoivre in 1738 but he did not discuss its properties."

We are not certain as to exactly what year Karl Pearson referred to $f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\frac{x-\mu}{\sigma})^2/2}$ as the normal distribution, but we surmise it must have been around the turn of the 20th century, perhaps in 1894. In statistical literature, the designation $X \sim N(\mu, \sigma^2)$ implies that the variate X is normally distributed with population mean μ and population variance $\sigma^2 = \sigma_X^2$.

Because the normal (or bell-shaped) density is the most important of all statistical distributions, below we summarize its practical properties, all of which are well known, mostly for the practitioner. Nearly every dimension that man manufactures can be approximately modeled by a normal underlying population.

(1) The normal density is exactly symmetrical about its population mean μ and because of symmetry the population median $x_{0.50} = \mu$, and the population modal (or maximum) point of the density also occurs at $MO = \mu$, i.e., $MO = x_{0.50} = \mu$ at which the 1st derivative of the density is zero. The height of the modal density is equal to $f(\mu; \mu, \sigma) = 1/(\sigma\sqrt{2\pi}) \cong 0.398942280401433/\sigma$. It can easily be verified that the expected-value of X, E(X), is equal to μ, and hence μ is the location parameter.

(2) It can easily be verified that the population variance of X, V(X), is equal to $\sigma_X^2 = \sigma^2$, and hence $\sigma = \sigma_X$ is the population scale-parameter. Amongst the most commonly encountered statistical distributions in-practice, the normal density is the one which has its standardized-value, $Z = (x-\mu)/\sigma$, appear in its exponent as $f(x; \mu, \sigma) =$

$\text{Exp}(-Z^2/2)/(\sigma\sqrt{2\pi}) = e^{-Z^2/2}/(\sigma\sqrt{2\pi})$; further, because all normal pdfs have identical bell-shape, then they have no shape parameter.

(3) The distance between the line x = μ and the points of inflection is exactly equal to σ, i.e., the second derivative of f(x; μ, σ) vanishes at x = $\mu \pm \sigma$.

(4) The normal density is concave for all $\mu - \sigma < x < \mu + \sigma$ and is convex for all $|x-\mu| > \sigma$. That is, the second derivative is strictly negative in the range $\mu - \sigma < x < \mu + \sigma$ and positive definite for $|x-\mu| > \sigma$.

(5) Due to symmetry, the skewness of a normal density is identically equal to zero, i.e., $\alpha_3 = E\{[(X-\mu)/\sigma]^3\} \equiv E(Z^3) = \mu_3/\sigma^3 \equiv 0$, where $\mu_3 = E[(X-\mu)^3]$ is the 3rd central moment.

(6) The forth standardized central moment of X is given by $\alpha_4 = \mu_4/\sigma^4 \equiv E(Z^4) = 3$, where $\mu_4 = E[(X-\mu)^4]$, and hence the kurtosis of all normal distributions is equal to $\beta_4 = \alpha_4 - 3 = 0$. Because the normal kurtosis $\beta_4 = \alpha_4 - 3 = 0$, then the kurtosis of all other statistical distributions is compared against 0 in order to assess their tail-thickness . Kendall & Stuart (1963), pp. 85-86, denote kurtosis by γ_2 and name curves with zero γ_2 as mesokurtic; curves with $\gamma_2 < 0$ as platykurtic, and those with $\gamma_2 > 0$ as leptokurtic. However, they do emphasize that leptokurtic curves are not necessarily more sharply peaked in the middle than the normal curve, and vice a versa for platykurtic curves.

(7) The moment generating function of X $\sim$ N(μ, σ^2) is given by mgf(t) = $\int_{-\infty}^{\infty} e^{tx} f(x;\mu,\sigma)\,dx$ = E(e^{tx}) and is equal to mgf(t) = $e^{\mu t + (\sigma^2 t^2/2)}$. It should be clear to the reader as to why $\int_{-\infty}^{\infty} e^{tx} f(x)\,dx$, where f(x) is the density of any continuous variate X, is called the mgf. Without carrying out the integration, only if the integral $\int_{-\infty}^{\infty} e^{tx} f(x)\,dx$ exists, one can partially differentiate inside the integral with respect to t

and then set t = 0 in order to obtain the population 1st-moment about the origin, μ_1'.

The second partial derivative with respect to t inside the integral and setting t = 0 results in μ_2' = E(X^2), the 2nd population origin-moment, etc.

(8) The standard normal density is almost universally denoted by $\phi(z) = e^{-Z^2/2} / \sqrt{2\pi}$, and the corresponding cumulative distribution function (cdf) is universally denoted by $\Phi(z) = \int_{-\infty}^{z} e^{-u^2/2} du / (\sqrt{2\pi})$. Some authors, such as Johnson *et al.* (1994) use U, instead of Z, to denote the unit normal variate N(0, 1). Because the cdf is an integral and the integrand has no closed-form antiderivative (excluding an infinite series), it is clear that unlike many other statistical distributions (such as the Uniform, Triangular, Exponential, Laplace, Cauchy, Extreme-value, Logistic, Weibull, etc.) the normal cdf is not directly invertible. That is, its quantile (or inverse) function $\Phi^{-1}(p) = Z_{1-p}$, $0 < p < 1$, must be obtained numerically by a computer-program. Nearly all statistical software (and Matlab) invert the standard normal cdf to at least 15 decimal accuracy. For example, Matlab gives $\Phi^{-1}(0.95) = Z_{0.05}$ = 1.644853626951473, and $\Phi^{-1}(0.975) = Z_{0.025}$ = 1.9599639845401.

(9) The location parameter pth quantile, or 100×p percentile, is given by $x_p = \mu + Z_{1-p}\times\sigma$, where Z_{1-p} is the pth quantile of a N(0, 1)-distribution. For example, the first decile is given by $x_{0.10} = \mu + Z_{0.90}\times\sigma$, where $Z_{0.90}$ = – 1.28155157.

(10) As stated before item (1) above the fact that how important the normal distribution is in developing statistical theory. In fact if X_i's, i = 1, 2, …, n are iid (independently and identically distributed) variates with means μ and variances σ^2, then one form of the CLT states that the limiting, in terms of n, sampling distribution (SMD) of the linear combination, $Y_n = \sum_{i=1}^{n} c_i X_i$ where c_i's are known constants, approaches normality with mean $E(Y_n) = \mu \sum_{i=1}^{n} c_i$, and variance $V(Y_n) = \sigma^2 \sum_{i=1}^{n} c_i^2$. The standard form of the CLT has

the form for which all c_i's = 1, i.e., $S_n = \sum_{i=1}^{n} X_i \to N(n\mu, n\sigma^2)$ as $n \to \infty$. The rate of approach to normality depends strictly on the skewness and kurtosis of the individual X_i's (Hool and Maghsoodloo (1980)), each of which is identically distributed like X. It can be proven that the skewness of $S_n = \sum_{i=1}^{n} X_i$ is given by $\alpha_3(S_n) = \mu_3(X_i)/(\sigma^3 \times \sqrt{n}) = \alpha_3(X)/\sqrt{n}$ and the kurtosis of S_n is given by $\beta_4(S_n) = [\alpha_4(X_i) - 3]/n = \beta_4(X)/n$ (the proof is available on request), where $\mu_4(X_i) = E[(X_i - \mu)^4] = \mu_4(X)$ and $\sigma^2 = V(X)$. For example, if X_i's are uniformly and independently distributed over the same interval, then $\alpha_3(S_n) \equiv 0$ and the kurtosis $\beta_4(S_n) = [\mu_4(X_i)/\sigma^4 - 3]/n = [(1/80)/(1/12)^2 - 3]/n = (144/80 - 3)/n = (1.80 - 3)/n = -1.20/n$. Thus, in this case we have a perfect fit of the first 3 moments of S_n with that of the normal, and an n = 10 forces the kurtosis of $S_{10} = \sum_{i=1}^{10} X_i$ to the value of –0.120, which is sufficiently close to zero for adequate normal approximation of $\sum_{i=1}^{10} X_i$ SMD (Sampling Distribution). On the other hand, if X_i's are independently and exponentially distributed (at the same identical parameter- rate λ), then it is well known that the n-fold convolution of $\sum_{i=1}^{n} X_i$ (or the Gamma density) has a skewness of $2/\sqrt{n}$ and a kurtosis of 6/n, where $\alpha_3(X_i) = 2$ and $\beta_4(X_i) = 6$ for all i. Therefore, in this case an n > 276 is needed for approximate normality of S_n because the skewness of $\sum_{i=1}^{277} X_i$ equals to 0.120 (to 3 decimals) and the skewness plays a much more important role in the Exponential case in approach to normality than does its kurtosis. Note that, to be on the conservative side, we have forced the skewness of this last S_{277} to become almost equal to the kurtosis of Uniform-convolution because the uniform-convolution is symmetrical and its skewness $\equiv$ 0 exactly matches that of the normal. Our experience shows that for nearly all statistical distributions, $\beta_4 > \alpha_3^2 - 2$. We do not yet know how much larger β_4 should be compared to α_3^2, before kurtosis plays a more important role in the approach to normality.

For the most complete analytical and mathematical discussions on the normal distribution, the reader is referred to Johnson, Kotz, and Balakrishnan, Vol. 1, 2nd edition, the entire Chapter 13, pp. 80-206. These three authors also provide a complete and comprehensive bibliography on the normal distribution for the time-period 1738-1994 on pp. 174-206.

2. Introduction

Consider a simple random sample of size n form a Laplace-Gaussian (or normal) population, denoted $N(\mu, \sigma^2)$, where both process mean μ and process standard deviation $\sigma = \sigma_X$ are unknown. There are numerous estimators of μ such as the sample arithmetic-mean, sample median, the sample mode, mid-range, trimmed mean, etc. It is well known from statistical theory that for sample sizes beyond 2 the most efficient of all estimators of a normal population mean μ is the sample mean $\overline{x} = \sum_{i=1}^{n} x_i / n$, the sample first origin-moment. For example, only at the trivial sample size n = 2, the relative efficiency (REL-EFF) of the sample median, $\hat{x}_{0.50}$, to $\overline{x}$ is 100%. Kendall and Stuart (1963), second edition, Vol. 1, pp. 236-243, show that the approximate variance of $\hat{x}_{0.50}$ in large samples is given by $V(\hat{x}_{0.50}) \cong 1/(4n\, f_{med}^2)$, where $f_{med} = f(x_{0.50})$ is the median ordinate of the underlying density. Because in the normal case $f_{med} = f(x_{0.50}) = f(\mu) = 1/(\sigma\sqrt{2\pi})$, then $V(\hat{x}_{0.50}) \cong \frac{\pi}{2}(\sigma^2 / n) = \frac{\pi}{2}V(\overline{x})$. Consequently, the asymptotic relative efficiency (ARE) of the sample median to sample mean is $ARE(\hat{x}_{0.50}, \overline{x}) = 2/\pi \cong 63.662\%$ for n > 30. This last ARE is provided by others such as Johnson *et al.* 's (1994), p. 124, and improves to 67.85%, 70.74%, 72.19%, 74.32%, 77.63%, 69.45%, 83.86%, 74.07% for sample sizes n = 20, 12, 10, 8, 6, 5, 4, & 3, respectively (see Kendall and Stuart (1963), 2nd Ed., p. 327, and Vol. 2 (1967), p. 7). Throughout this article, discussions pertain only to normal underlying populations with unknown process mean μ and unknown population standard deviation σ, and the symbol V stands for the nonlinear population-variance operator.

Another estimator of the parameter μ, proposed by Hogg & Tanis (2010), p. 127, is the Trimean = (the first sample quartile +2×the sample median + the 3rd sample quartile)/4. Although the authors do not provide an expression for the variance of this estimator, our intuition tells us that it cannot be as efficient as $\overline{x}$ because, unlike the sample mean, the Trimean uses the sample information of only a maximum of 6 order-statistics. So, we suspect that its relative efficiency to $\overline{x}$ will be less than 100%. We illustrate this below by applying the population-variance (V) and covariance (COV) operators.

V(Trimean) = $\frac{1}{16}[V(Q_1)+4V(Q_2)+V(Q_3)+4COV(Q_1,Q_2)+2COV(Q_1,Q_3)+4COV(Q_2,Q_3)]$

where Q_1 is the sample 1st-quartile, Q_2 is the sample median, and Q_3 is the sample 3rd-quartile. Kendall & Stuart (1963), Vol. 1, p. 237-239, provide approximate formulas for both the variance of sample quantiles and covariance of any two sample percentiles (more details forthcoming). Using their results, we obtain:

$$V(\text{Trimean}) \cong \frac{1}{16}\left[\frac{0.25\times 0.75}{nf_1^2}+\frac{4(0.5)^2}{nf_2^2}+\frac{0.75\times 0.25}{nf_3^2}+4\frac{0.25\times 0.50}{nf_1 f_2}+2\frac{0.25\times 0.25}{nf_1 f_3}+4\frac{0.5\times 0.25}{nf_2 f_3}\right]$$

$$V(\text{Trimean}) \cong \left[0.50/(nf_1^2)+1/(nf_2^2)+1/(nf_1 f_2)\right]/16,$$

where $f_1 = f_3$ is the density-ordinate at either $x_{0.25}$ or $x_{0.75}$, and $f_2 = 1/(\sigma\sqrt{2\pi}) = 0.3989423/\sigma$. Substituting $f_1 = f_3 = 0.317776573/\sigma$ and f_2, we obtain V(Trimean) $\cong$ (1.19516164)σ^2/n = (1.19516164) $V(\overline{x})$. Further, we surmise that the Trimean is a biased estimator of μ, and hence its REL-EFF must be less than 83.671% for all n > 3. Johnson and *et al.* (1994), Vol. 1, p. 13, state that only the limiting bias in sample quantiles is roughly zero. The value of n must exceed 3 because 3 point estimates must be made in order to compute the value of Trimean. (Perhaps to be on the safe side, one must not compute Trimean unless n > 4.)

Just like the midrange $[x_{(n)} + x_{(1)}]/2$ and Trimean, one could define numerous other estimators of μ (using the symmetrical property of the normal distribution) as $\hat{\mu}_Q$ = $(\hat{x}_p + \hat{x}_{1-p})/2$, p > 0.50, and Q for quantile. However, except at n = 2 for the midrange, none of these estimators of μ are nearly as efficient as the sample mean $\overline{x}$. Kendall & Stuart (1963, p.

327) clearly state that the midrange becomes less and less efficient with increasing sample size, and the limiting value of its relative-efficiency, as $n \to \infty$, is zero. We used their result on p. 239 to compute the approximate variance of $\hat{\mu}_Q$ in the normal case, which is given by $V(\hat{\mu}_Q) = \pi q e^{Z_p^2}(\sigma^2 / n)$, where $q = 1-p < p$, and REL-EFF$(\hat{\mu}_Q, \bar{x}) = e^{-Z_p^2} / (\pi q)$. For example, at $p = 0.75$, the REL-EFF $(\hat{\mu}_{0.75}, \bar{x})$ = 80.786%, while the REL-EFF $(\hat{\mu}_{0.90}, \bar{x})$ = 61.599%; as $p \to 1$, the REL-EFF $(\hat{\mu}_Q, \bar{x})$ goes to zero. Because the arithmetic mean $\bar{x}$ uses all the information in the sample of size n, as has been reported in statistical literature, there does not exist another estimator, except at n = 2, of the normal μ which is more efficient than $\bar{x}$ with the exception of samples that contain outliers. In such samples, the trimmed mean may be more efficient.

Similarly, it is well known that the statistic $S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 / (n-1)$ is an unbiased estimator of σ^2 iff the underlying population is infinite. However, $S = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 / (n-1)}$ is always a biased estimator of σ from a normal universe. In fact, it is well documented both in Statistical and Quality Control (QC) literatures that for a normal universe the expected-value of S, $E(S) = c_4(n)\sigma$, where the biasing-factor $c_4(n) = \sqrt{\frac{2}{n-1}} \times \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]} = \sqrt{\frac{n-1}{2}} \times \frac{\Gamma(n/2)}{\Gamma[(n+1)/2]}$ is a function of n, $0.797884560802 < c_4 < 1$ for all $2 \le n < \infty$, and the expected-value operator E is linear, and the symbol $\Gamma(\cdot)$ stands for the Gamma function. We have verified that a good approximation for $c_4(n) \cong (4n^2 - 8n + 3.8775)/(4n^2 - 7n + 3)$. This last approximation is accurate to 3 decimals for $n \ge 7$, to 4 decimals for $n \ge 13$, and to 5 decimals for $n \ge 19$, the approximation improving with increasing n. It is widely known that the amount of bias in S as an estimator of σ is given by $B(S) = E(S) - \sigma = (c_4 - 1)\sigma$. It is also well known (we are not providing reference because there are too many) that for a $N(\mu, \sigma^2)$-universe, $V(S) = (1 - c_4^2)\sigma_x^2$ so that the Mean Square Error of S is equal to $MSE(S) = E[(S - \sigma)^2] = V(S) + [B(S)]^2 = 2(1-c_4)\sigma^2$, where MSE measures the accuracy of an estimator. Note that the MSE(S) almost equals to

V(S) for n > 100 to at least 5 decimals. The reader should bear in mind that the operator E computes the weighted-average over all elements in the range space , and hence the end-result will always invariably be a population parameter. The relationship E(S) = $c_4(n)\sigma$ clearly shows that an unbiased estimator of σ is given by $\hat{\sigma}_{UB} = S/c_4$ so that the MSE of $\hat{\sigma}_{UB}$ is given by MSE($\hat{\sigma}_{UB}$) = V($\hat{\sigma}_{UB}$) = $(c_4^{-2} - 1)\sigma^2$. Thus, REL-EFF($\hat{\sigma}_{UB}$, S) = $2(1-c_4)/(c_4^{-2}-1)$ = $2c_4^2/(1+c_4)$, which is equal to 70.819% at n = 2, and is monotonically increasing, in terms of n, so that at n = 5 its value is REL-EFF($\hat{\sigma}_{UB}$, S, n = 5) = 91.091%, REL-EFF($\hat{\sigma}_{UB}$, S, n = 25) = 98.449%, and increases to REL-EFF($\hat{\sigma}_{UB}$, S, n = 50) = 99.237%.

It is also well known that for an underlying normal universe, the maximum likelihood estimator (MLE) of σ is given by the square-root of the sample variance (or the sample 2nd central moment) $m_2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$, i.e., $\hat{\sigma}_{mle} = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/n}$, the so called *root mean square*. Note that nearly all authors refer to S^2 as the sample variance because it is an unbiased estimator of σ^2 only for all infinite populations, but we agree with Kendall & Stuart's (1967), Vol. 2, designation on their pages 4-5, because S^2 is not the sample second central moment, consequently it is not exactly the sample variance, and the population second central moment is $\mu_2 = E[(X-\mu)^2]$; hence, the moment-estimator of μ_2 is the sample variance m_2, where m_2 is the weighted-average of deviations-from-sample-mean squared. It is widely known that $\hat{\sigma}_{mle}$ is also the moment-estimator of σ, and that the amount of bias B($\hat{\sigma}_{mle}$) = $(c_4\sqrt{(n-1)/n} - 1)\sigma$, its variance is given by V($\hat{\sigma}_{mle}$) = $(n-1)(1-c_4^2)\sigma^2/n$, and the MSE($\hat{\sigma}_{mle}$) = $[2(1 - c_4\sqrt{(n-1)/n}) - 1/n]\sigma^2$. This shows that $\hat{\sigma}_{mle}$ is a more accurate point estimator of σ than S because the REL-EFF($\hat{\sigma}_{mle}$, S) = $(1-c_4)/[1 - c_4\sqrt{(n-1)/n} - (2n)^{-1}]$ > 1 for all finite n. In fact, REL-EFF($\hat{\sigma}_{mle}$, S, n = 2) = 108.775%, REL-EFF($\hat{\sigma}_{mle}$, S, n = 5) = 101.288%, REL-EFF($\hat{\sigma}_{mle}$, S, n = 25) = 100.050% and exceeds 100% for all finite n. Only for $n \geq 80$, the REL- EFF(S, $\hat{\sigma}_{mle}$, n) $\cong$ 100.00% to 2 decimal places.

Another estimator of σ from the field of QC is $\hat{\sigma}_X = R/d_2$, where R is the sample range and $d_2 = E(W) = E(R/\sigma)$, ($W = R/\sigma$ = Sample Relative Range). The values of d_2 are reproduced in Table 1 and are also tabulated in every text on QC. Clearly, $\hat{\sigma}_X = R/d_2$ is an unbiased estimator of σ and hence its MSE is equal to its variance. The SMD of the sample range R from a normal universe is extremely complicated and is provided by Kendall & Stuart (1963), pp. 337-338. Numerical integration must be used to compute both the population mean and variance of W from a normal universe at different values of n. Nevertheless, $V(\hat{\sigma}_X) = V(R/d_2) = V(R)/d_2^2 = \sigma^2 V(R/\sigma)/d_2^2 = \sigma_X^2 \times d_3^2/d_2^2$, where $d_3^2 = V(W = R/\sigma)$ from a normal universe. The values of d_2 & d_3 were obtained by L. H. C. Tippett (1925), E. S. Pearson (1926), E. S. Pearson (1932) and H. L. Harter (1960). Their results are also reproduced herein Tables 1 and 2. We used their values in Tables 1 & 2 in order to to obtain the REL-EFF($\hat{\sigma}_X = R/d_2$, S) = MSE(S)/V(R/d_2) at n = 2(1)20(5)35 and are given in Table 3 in percent. Not surprisingly, Table 3 shows that the estimator $\hat{\sigma}_X = R/d_2$ attains its maximum REL-EFF at the nominal sample size n = 5 in the field of QC for R-& $\bar{x}$ -charts. Due to the fact that $R = x_{(n)} - x_{(1)}$ uses only the n^{th} and 1^{st}-order statistics of the sample, while S uses all the information in the sample, R/d_2 is never as efficient as S in estimating $\sigma = \sigma_X$. Further, as Kendall & Stuart (1963), p. 339, mention the SMD of sample range R diverges from normality as n increases, and is well-documented in statistical literature that the SMD of R is also unstable for moderate to large sample sizes (n > 20). Hence, one should use R/d_2 as an estimator of σ_X for only small sample sizes $2 \le n \le 20$. The REL-EFF(R/d_2, S) is monotonically decreasing after n = 5 and is a mere 34.661% at n = 100.

Table 1. The Expected-Value of Relative Range ($W = R/\sigma$) of a $N(\mu, \sigma^2)$ from Tippett (Dec. 1925)

n	2	3	4	5	6	7	8	9	10	11
d_2	1.128379	1.692569	2.058751	2.325929	2.534413	2.704357	2.847201	2.970026	3.077505	3.172873
n	12	13	14	15	16	17	18	19	20	21
d_2	3.258455	3.335980	3.406763	3.471827	3.531983	3.587884	3.640064	3.688963	3.734950	3.778336
n	22	23	24	25	26	27	28	29	30	35
d_2	3.819385	3.858323	3.895348	3.930629	3.964316	3.996539	4.027414	4.057044	4.085522	4.213219

Another point estimator of σ, proposed by Kendall & Stuart (1963), p. 239, is $\hat{\sigma}_{MD}$ = $MD \times \sqrt{\pi/2}$, where the sample mean-deviation (or average-deviation) is given by *MD* = $\sum_{i=1}^{n}|x_i - \bar{x}|/n = \sum_{i=1}^{n}\text{abs}(x_i - \bar{x})/n$. Below, we will try to argue for their recommendation.

Table 2. The Standard Deviation of W = R/σ of a Normal Universe from Table 2 of H. Leon Harter*

n	2	3	4	5	6	7	8	9	10	11
d_3	0.852502	0.888368	0.879808	0.864082	0.848040	0.833205	0.819831	0.807834	0.797051	0.787315
n	12	13	14	15	16	17	18	19	20	21
d_3	0.778478	0.770416	0.763023	0.756211	0.749908	0.744052	0.738591	0.733481	0.728686	0.724173
n	22	23	24	25	26	27	28	29	30	35
d_3	0.719915	0.715887	0.712068	0.708441	0.704988	0.701697	0.698553	0.695546	0.692665	0.679871

**Harter's (1960) Table 2 provides 10-decimal accuracy for both $d_2 = E(W=R/\sigma)$ and $d_3^2 = V(R/\sigma)$; further, we used Table 2 of H. L. Harter to obtain the 6th decimal in the above Table 1 which was extracted from Tippett (1925) at 5 decimal accuracy.*

Table 3. The REL-EFF(R/d_2, S) in percent to two decimals.

n	2	3	4	5	6	7	8	9	10	11	12
REL-EFF	70.82	82.60	86.17	86.97	86.58	85.61	84.35	82.96	81.52	80.07	78.63
n	13	14	15	16	17	18	19	20	25	30	35
REL-EFF	77.23	75.87	74.55	73.28	72.05	70.88	69.74	68.65	63.78	59.71	56.26

Applying the expected-value operator, we obtain E(*MD*) = $\frac{1}{n}E\sum_{i=1}^{n}|x_i - \bar{x}|$ = $\frac{1}{n}\sum_{i=1}^{n}E|x - \bar{x}|$ = $\frac{1}{n}nE|x - \bar{x}|$ = $E|x - \bar{x}|$ = $\int_{-\infty}^{\infty}|x - \bar{x}|\frac{\text{Exp}[-(x-\mu)^2/(2\sigma^2)]}{\sigma\sqrt{2\pi}}dx$. However, carrying out the exact integration of this last integral is beyond our grasp, and as has been done by numerous others, we will make an approximation as follows:

$$E(MD) = \int_{-\infty}^{\infty} |(x-\mu)-(\overline{x}-\mu)| \frac{Exp[-(x-\mu)^2/(2\sigma^2)]}{\sigma\sqrt{2\pi}} dx$$

The expected value of $|(x-\mu)-(\overline{x}-\mu)|$ is extremely difficult to handle, and therefore, we argue that in simple random samples of moderate to large sizes n (> 20), the value of (x–μ) should be much larger than $\overline{x}-\mu$, and hence $|(x-\mu)-(\overline{x}-\mu)| \cong |x-\mu|$ to the order of $n^{-1/2}$. A simulation-study will be carried out to obtain a better approximation of E(*MD*) at different values of n. Applying this approximation, we obtain E(*MD*) ≈ $\int_{-\infty}^{\infty} |x-\mu| \frac{Exp[-(x-\mu)^2/(2\sigma^2)]}{\sigma\sqrt{2\pi}} dx$, where ≈ denotes rough approximation. Using the definition of absolute values and making the transformation of Z = (x–μ)/σ results in E(*MD*) ≈

$$\int_{-\infty}^{0} -(x-\mu)\frac{Exp[-Z^2/2]}{\sqrt{2\pi}} dZ + \int_{0}^{\infty} (x-\mu)\frac{Exp[-Z^2/2]}{\sqrt{2\pi}} dZ = 2\sigma\int_{0}^{\infty} Z\frac{e^{-Z^2/2}}{\sqrt{2\pi}} dZ = \frac{2\sigma}{\sqrt{2\pi}}\left[-e^{-Z^2/2}\right]_0^{\infty} =$$

$\frac{2\sigma}{\sqrt{2\pi}} = \frac{\sigma}{\sqrt{\pi/2}}$. This last result is also provided by many authors; for example, see Johnson and *et al.* (1994), Vol. 1, 2nd edition, p. 91. Setting this last approximate-expectation to the corresponding sample statistic $\frac{1}{n}\sum_{i=1}^{n}|x_i-\overline{x}|$ results in $\hat{\sigma}_{MD} = \sqrt{\pi/2} \times MD = \sqrt{\pi/2} \times \sum_{i=1}^{n}|x_i-\overline{x}|/n$. Kendall & Stuart (1963), p. 243, give an approximation for the variance of *MD* (they also provide this information for almost any other statistic) as $V(\hat{\sigma}_{MD}) = V\left(\frac{\sqrt{\pi/2}}{n} \times \sum_{i=1}^{n}|x_i-\overline{x}|\right) = \frac{\pi}{2n}(1-2/\pi)\sigma^2 = (0.50\pi-1)\sigma^2/n$. Clearly, the estimator $\hat{\sigma}_{MD}$ must be biased, and as shown above, the amount of bias is $B(\hat{\sigma}_{MD}) \approx 0$ to order of $n^{-1/2}$. In practice the value of S will exceed $\hat{\sigma}_{MD}$ iff the data range is sufficiently large (we do not yet know how large). A simulation study can quantify the amount and sign of the bias. Below we will attempt to obtain the approximate variance of $\hat{\sigma}_{MD}$ and compare it against that of Kendall & Stuart's.

$$V(\hat{\sigma}_{MD}) = V(\sqrt{\pi/2}\ \sum_{i=1}^{n}|x_i - \bar{x}|/n) = \frac{\pi}{2n^2} V\sum_{i=1}^{n}|x_i - \bar{x}|$$

$$= \frac{\pi}{2n^2}\left[E(\sum_{i=1}^{n}|x_i - \bar{x}|)^2 - (E\sum_{i=1}^{n}|x_i - \bar{x}|)^2\right] \qquad (1)$$

We first compute the 1st expectation inside brackets in equation (1) on the right-hand side (RHS): $E[(\sum_{i=1}^{n}|x_i - \bar{x}|)^2] = E[\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i\neq j}^{n}|x_i - \bar{x}||x_j - \bar{x}|] = E[(n-1)S^2] +$

$\sum_{i\neq j}^{n}[E|x_i - \bar{x}| \times E|x_j - \bar{x}|] \approx (n-1)\sigma^2 + \sum_{i\neq j}^{n}[\sigma/\sqrt{\pi/2} \times \sigma/\sqrt{\pi/2}] = (n-1)\sigma^2 + n(n-1)2\sigma^2/\pi.$

The 2nd expectation on the far-right of Eq. (1) is equal to $E\sum_{i=1}^{n}|x_i - \bar{x}| = nE|x_i - \bar{x}| \approx \frac{n\sigma}{\sqrt{\pi/2}}$;

thus, insertion into (1) yields

$$V(\hat{\sigma}_{MD}) \cong \frac{\pi}{2n^2}\left[(n-1)\sigma^2 + 2n(n-1)\sigma^2/\pi - (\frac{n\sigma}{\sqrt{\pi/2}})^2\right] = \frac{\pi}{2n^2}\left[(n-1)\sigma^2 - 2n\sigma^2/\pi\right]$$

$$V(\hat{\sigma}_{MD}) \cong \frac{\pi}{2}\left[(n-1)/n - 2/\pi\right]\frac{\sigma^2}{n}, \qquad (2)$$

as compared to $\frac{\pi}{2}(1-2/\pi)\sigma^2/n$ given by Kendall & Stuart (1963). If we take the conservative route and replace (n–1)/n by 1 in Eq. (2), n > 20, then our variance result in Eq. (2) also reduces to $V(\hat{\sigma}_{MD}) \approx (\pi/2-1)\sigma^2/n$, which is now consistent with hat of Kendall & Stuart's. Therefore, we compute an approximate REL-EFF by comparing $V(\hat{\sigma}_{MD})$ to V(S). That is, REL-EFF($\hat{\sigma}_{MD}$, S) $\cong$ $\frac{1-c_4^2}{(0.50\pi-1)/n} = \frac{n(1-c_4^2)}{0.50\pi-1}$. This REL-EFF is maximum at n = 2, equaling 127.324%, diminishes to 101.987% at n = 5, and is less than 100% for all n > 5. At n = 30, its value is 89.823%; at n = 50, it monotonically decreases to 88.924%, and is equal to 88.257% at n = 100.

3. Estimates of a Laplace-Gaussian-σ Based on Quantiles

Is seems that there are uncountably infinite estimators of a Laplace-Gaussian-σ because of the fact that the p^{th} quantile of a $N(\mu, \sigma^2)$ is given by $x_p = \mu + Z_{1-p}\times\sigma$, where Z_{1-p} is the $(1-p)\times100$ percentage point, or the p^{th} quantile, of the N(0, 1)-density. For example, the 0.90 quantile of $X \sim N(\mu, \sigma^2)$ is given by $x_{0.90} = \mu + Z_{0.10}\times\sigma$, where $Z_{0.10} = 1.28155157$ is the 0.90-quantile of a N(0, 1), and thus the 1^{st} decile is given by $x_{0.10} = \mu - 1.28155157\times\sigma$, by symmetry. Nearly all statistical software, such as Minitab, also Matlab & even Microsoft Excel, provide the inverse functions of nearly all statistical distributions that are not directly invertible to 15 decimal accuracies. Although, Minitab17 has a slight discrepancy starting in the 11^{th} decimal place. The two equations for the 9^{th} and 1^{st} deciles clearly show that the value of population interdecile range is given by $IDR = x_{0.90} - x_{0.10} = 2Z_{0.10}\times\sigma$, which leads to estimator $\hat{\sigma}_{idr} = (\hat{x}_{0.90} - \hat{x}_{0.10})/(2Z_{0.10})$, where $\hat{x}_p$ is the corresponding $(p\times100)^{th}$ sample percentile. Therefore, σ-estimates for a normal universe based on quantiles are given by $\hat{\sigma}_Q = (\hat{x}_p - \hat{x}_{1-p})/(2Z_{1-p})$, $p > 0.50$, and $Z_{1-p} > 0$.

It is well known that the sample quantiles are obtained from sample order statistics $x_{(i)}$, $i = 1, 2, 3, \ldots, n$, as follows: (1) Multiply (n+1) by p; if $(n+1)\times p$ is an exact integer, say equal to I, then $\hat{x}_p = x_{(I)}$. (2) If (n+1)p is not an exact integer such that $I < (n+1)p < I+1$, then the sample p^{th}-quantile is given by the convex combination $\hat{x}_p = a x_{(I)} + (1-a)x_{(I+1)}$, where $0 < a = (I+1) - (n+1)p < 1$. As shown here, sample quantiles are obtained from sample order-statistics, whose sampling distributions depend on the parental density f(x). Further, such point estimates are biased, and as stated by Kendall & Stuart, (1963, 2e), p. 236, the amount of bias $B(\hat{x}_p) = E(\hat{x}_p) - x_p$ diminishes as n increases to order of 1/n. Kendall and Stuart (1963, pp. 236-7) proceed to obtain the variance of a sample quantile for any continuous underlying distribution, for large n, as $V(\hat{x}_p) \cong pq/[nf^2(x_p)]$, where $q = 1-p$ and $f(x_p)$ is the ordinate (or height) of the density at x_p. This last approximation is given in their Eq. (10.29) on page 237 of their second edition, but they do not state how large n should be for adequate approximation of their Eq. (10.29). If the underlying distribution is unknown, then Kendall & Stuart (1963), p. 237, recommend that a frequency distribution (n > 30) be developed and $f(x_p)$ be estimated by the frequency per unit interval at $\hat{x}_p$. Of course, in the normal case, a logical estimator of p^{th} quantile is $\hat{x}_{QUB}$ =

$\bar{x} + Z_{1-p}S / c_4$. This estimator is obviously unbiased and its exact variance, because $\bar{x}$ and S are stochastically independent, is given by

$$V(\hat{x}_{QUB}) = MSE(\hat{x}_{QUB}) = \sigma^2 / n + Z_{1-p}^2 (c_4^{-2} - 1)\sigma^2 \quad (3)$$

Presently, we do not know the exact MSE($\hat{x}_p$), and therefore, we compare its variance $pq / [nf^2(x_p)]$ against Eq. (3). The REL-EFF($\hat{x}_p, \hat{x}_{QUB}$) < 100% for all n and all 0 < p < 1.

The biased estimator of x_p given by $\hat{x}_Q = \bar{x} + Z_{1-p}S$, has a MSE($\hat{x}_Q$) = $[1/n + 2(1-c_4) Z_{1-p}^2]\sigma^2$ and is a bit more efficient than $\hat{x}_{QUB}$ for all $p \neq 0$, 0.50, and 1.

Although both estimators $\hat{x}_p$ and $\hat{x}_{1-p}$ of x_p and x_{1-p}, respectively, are biased, we suspect that the sign of bias in them is the same , and thus the sample inter-percentile range IPR = $\hat{x}_p - \hat{x}_{1-p}$, p > 0.50, suffers from less bias as an estimator of the population $x_p - x_{1-p}$ than individual sample percentiles. Hence, the MSE($\hat{x}_p - \hat{x}_{1-p}$) $\cong$ V($\hat{x}_p - \hat{x}_{1-p}$) in large samples, p > 0.50. Kendall and Stuart (1963, pp. 238-9) also proceed to obtain the variance of difference of two distinct sample quantiles, x_1 and x_2, and their equation is exactly repeated below in their notation:

$$\text{“var } \delta = \frac{1}{n}\left\{\frac{p_1 q_1}{f_1^2} + \frac{p_2 q_2}{f_2^2} - \frac{2 p_2 q_1}{f_1 f_2}\right\}\text{”}$$

where $\delta = x_1 - x_2$, f_1 is the height (or ordinate) of density at the quantile x_1, $p_2 = 1 - q_2$, and they do not number their above equation on p. 239. We used their above result to obtain

$$V(IPR) = V(\hat{x}_p - \hat{x}_{1-p}) \cong 2(3p - 2p^2 - 1) / [nf^2(x_p)], \; p > 0.50, \quad (4a)$$

where for $X \sim N(\mu, \sigma^2)$ the value of $f(x_p) = f(x_{1-p}) = \text{Exp}(-Z_p^2 / 2) / (\sigma\sqrt{2\pi})$. Hence,

$$V(\hat{x}_p - \hat{x}_{1-p}) \cong 2(3p - 2p^2 - 1) / [nf^2(x_p)] = 4\pi\sigma^2 (3p - 2p^2 - 1) e^{Z_p^2} / n, \; p > 0.50. \quad (4b)$$

As a result, $V(\hat{\sigma}_Q) = V[(\hat{x}_p - \hat{x}_{1-p}) / (2Z_{1-p})] = [\pi(3p - 2p^2 - 1) e^{Z_p^2} / Z_p^2]\sigma^2 / n$, (5)

where $Z_p = -Z_{1-p}$ and p > 0.50. For example, for p = 0.95, $Z_{0.95} = -1.644853627$.

We first consider the σ-estimator based on sample IQR, namely $\hat{\sigma}_{iqr} =$

$(\hat{x}_{0.75} - \hat{x}_{0.25})/(2Z_{0.25})$. From Eq. (5), V($\hat{\sigma}_{iqr}$) = V[$(\hat{x}_{0.75} - \hat{x}_{0.25})/(2Z_{0.25})$] ≅ 1.3604593043σ²/n = 1.3604593043 $\sigma^2_{\bar{x}}$. As stated above $\hat{\sigma}_Q$ is biased and we should compare its variance against V(S), but we take the conservative route and compare it against the MSE(S), i.e., REL-EFF($\hat{\sigma}_{iqr}$, S) ≅ 2n(1−c_4)/1.3604593. This last REL-EFF is larger than V(S)/V($\hat{\sigma}_{iqr}$) = n(1−c_4^2)/1.3604593043. Thus $\hat{\sigma}_{iqr}$ is a very poor estimator of σ; in fact, the REL-EFF($\hat{\sigma}_{iqr}$, S) monotonically diminishes with increasing n from 50.177% at n = 3 to 37.849% at n = 30. This should not be surprising because as n increases the estimator $\hat{\sigma}_{iqr}$, unlike S, uses less and less proportion of the sample information. In fact, $\hat{\sigma}_{iqr}$ uses a maximum of 4 order-statistics of the sample to obtain $\hat{x}_{0.75}$ and $\hat{x}_{0.25}$. Similarly, another point estimator of σ_X is given from sample interdecile range (IDR) as $\hat{\sigma}_{idr} = (\hat{x}_{0.90} - \hat{x}_{0.10})/(2Z_{0.10})$, where $Z_{0.10}$ = 1.28155157. Eq. (5) gives V($\hat{\sigma}_{idr}$) = 0.790754795603×σ²/n, and REL-EFF($\hat{\sigma}_{idr}$, S) ≅ 2.5292290491×n(1−c_4). This REL-EFF is far superior to that of $\hat{\sigma}_{iqr}$ and is equal to 86.327% at n = 3, and monotonically diminishes to 65.117% at n = 30.

4. Estimates Based on Mean of Successive Squared Differences

Minitab 17, the easiest-to-use Statistical Package in our opinion, provides another estimator of σ² in their "Display Descriptive Statistics" menu called the Mean of Successive Squared Differences, which they denote by MSSD, and is defined below

$$\text{Minitab's MSSD} = \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2$$

where x_i is the i^{th} random observation (not the i^{th} order-statistic $x_{(i)}$) in the sample of size n > 1. We have verified, a proof is available on request, that E(MSSD) = σ², and hence MSSD is an unbiased point estimator of σ². Therefore, a point-estimator of σ_X is given by

$$\hat{\sigma}_{SSD} = \sqrt{\frac{1}{2(n-1)}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2}\ ,$$

where SSD stands for Successive-Squared-Differences. Because x_{i+1} and x_i are iid, then $E(x_{i+1} - x_i) \equiv 0$ and $V(x_{i+1} - x_i) \equiv \sigma^2 + \sigma^2 = 2\sigma^2$. Further, we are random-sampling only a normal underlying population, then $(x_{i+1} - x_i)^2/(2\sigma^2)$ has a central chi-square distribution with 1 degree of freedom. However, the successive squared-differences $(x_2 - x_1)^2/(2\sigma^2)$, $(x_3 - x_2)^2/(2\sigma^2)$, $(x_4 - x_3)^2/(2\sigma^2)$, ..., $(x_n - x_{n-1})^2/(2\sigma^2)$, except in the trivial case n = 2, are obviously heavily and hopelessly autocorrelated, and hence $\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2/(2\sigma^2)$ does not have a central-chi-square distribution, although it may have n–1 degrees of freedom (*df)*, and as a result we have not yet obtained the approximate expected-value of $\hat{\sigma}_{SSD}$. Moreover, because the $V(\hat{\sigma}_{SSD}) = \sigma^2 - [E(\hat{\sigma}_{SSD})]^2 > 0$, then the sign of the bias in $\hat{\sigma}_{SSD}$ must be negative. We may define another estimator of σ_X^2 for even sample sizes n = 2r (r = 1,2, …, n/2) as MSSDD = $\sum_{i=1}^{n/2}(x_{2i} - x_{2i-1})^2/n$, where MSSDD stands for Mean of Successive Squared Distinct Differences such that the successive squared variates $(x_2 - x_1)^2/(2\sigma^2)$, $(x_4 - x_3)^2/(2\sigma^2)$, $(x_6 - x_5)^2/(2\sigma^2)$, ..., $(x_n - x_{n-1})^2/(2\sigma^2)$ are now independent. It can be shown (the proof is available on request) that this estimator is unbiased and that $\sum_{i=1}^{n/2}(x_{2i} - x_{2i-1})^2/(2\sigma^2)$ has a $\chi^2_{n/2}$ SMD, and the estimator $\hat{\sigma}_{SSDD} = \sqrt{\sum_{i=1}^{n/2}(x_{2i} - x_{2i-1})^2/n}$ has a $MSE(\hat{\sigma}_{SSDD}) = 2\{1 - 2\times\Gamma[(n+2)/4]/[\Gamma(n/4)\sqrt{n}]\}$. The REL-EFF($\hat{\sigma}_{SSDD}$, S) = 100% only at n = 2, and rapidly approaches 50% with increasing n = 4, 6, 8, …, which should be expected. A similar estimator can also be defined, albeit more painstakingly, for odd n = 3, 5, 7, … .

The exact SMD of S is derived by Kendall & Stuart [(1963), p. 256 where their "*s*" is the moment-estimator from the N(0,1)], and is also given by Ostle & Malone (1988), pp. 80-81 for the general $N(\mu, \sigma^2)$, and repeated from them by Chou (1992, p. 10). However, because the exact SMD of $S^2 \sim \sigma^2\chi^2_{n-1}/(n-1)$, statistical properties of variate S can be studied using χ^2_{n-1}.

Although, S is not a linear sum of independent variates, nevertheless it is widely known that its SMD very slowly approaches the $N(\sigma, \sigma^2/2n)$ pdf as $n \to \infty$. It has been verified by the author that a better approximation for the exact V(S) = $(1-c_4^2)\sigma^2$ from a normal universe is V(S) $\cong \sigma^2/[2(n-0.745)]$, n > 50. For practical applications, when X $\sim N(\mu, \sigma^2)$, the normal approximation to the SMD of S is fairly accurate for n > 75. From the above discussion on the approximate SMD of S (for n > 75), we suspect that the limiting SMD of $\hat{\sigma}_{SSD}$ will also be normal with mean roughly close to σ and unknown variance. Further, we do not know how large n must be for adequate normal approximation of $\hat{\sigma}_{SSD}$.

Finally, because the ordinate of normal mode, f(MO), is equal to $1/(\sigma\sqrt{2\pi})$, and if the large sample unimodal frequency per unit of X, $\hat{f}(MO)$, can be obtained, then a rough point estimate of σ_X is given by $\hat{\sigma}_X \approx [\hat{f}(MO)\times\sqrt{2\pi}]^{-1}$. However, we have presently no information about the statistical properties of this last estimator, and we suspect that it is not at all efficient because it uses a small proportion of sample information.

5. Estimates of Normal-σ Based on M > 1 Subgroups

Consider M independent random samples of different sizes, $n_i > 1$, i = 1, 2, ..., M from the same normal universe (with unknown process mean μ and unknown population standard deviation σ), where N = $\sum_{i=1}^{M} n_i$. Such unbalanced sampling occurs frequently in the field of QC, where commonly $20 \le M \le 30$ independent subgroups are obtained to setup trial-control limits. Let $S_p = \sqrt{\sum_{i=1}^{M}(n_i-1)S_i^2 / \sum_{i=1}^{M}(n_i-1)}$ be the pooled-estimator of σ = σ_X. Before computing the Mean-Square Error of S_p, denoted MSE(S_p), we must first compute the amount of bias in S_p as a point estimator of process standard deviation, σ_X. As a result, applying the linear Expected-Value Operator, E, we obtain

$$E(S_p) = E\sqrt{\sum_{i=1}^{M}(n_i-1)S_i^2 / \sum_{i=1}^{M}(n_i-1)} = E\sqrt{\sum_{i=1}^{M}CSS_i / (N-M)}\,, \tag{6}$$

where the Corrected Sum of Squares $CSS_i = (n_i-1)S_i^2 = \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$ and $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}$.

Using the linear properties of the operator E and additive property of χ^2, Eq. (6) yields

$$E(S_p) = \frac{1}{\sqrt{N-M}} E\sqrt{\sum_{i=1}^{M} CSS_i} = \frac{\sigma}{\sqrt{N-M}} E\sqrt{\sum_{i=1}^{M} CSS_i / \sigma^2} = \frac{\sigma}{\sqrt{N-M}} E\sqrt{\sum_{i=1}^{M} \chi^2_{n_i - 1}}$$

$$= \frac{\sigma}{\sqrt{N-M}} E\sqrt{\chi^2_{\sum_{i=1}^{M}(n_i-1)}} = \frac{\sigma}{\sqrt{N-M}} E\sqrt{\chi^2_{(N-M)}}, \tag{7}$$

where χ^2_ν denotes a Chi-square variate with ν *df*. It is widely known the $E\sqrt{\chi^2_\nu}$ =

$\frac{\sqrt{2}\times\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)}$ (a proof is available from the author), and thus Eq. (7) reduces to

$$E(S_p) = \frac{\sigma}{\sqrt{N-M}} E\sqrt{\chi^2_{N-M}} = \frac{\sigma}{\sqrt{N-M}} \times \frac{\sqrt{2}\times\Gamma[(N-M+1)/2]}{\Gamma[(N-M)/2]}, \text{ i.e.,}$$

$$E(S_p) = c_5 \times \sigma, \tag{8}$$

where we have denoted the already-known QC constant

$$c_5 = \sqrt{\frac{2}{N-M}} \times \frac{\Gamma[(N-M+1)/2]}{\Gamma[(N-M)/2]} = \sqrt{\frac{N-M}{2}} \times \frac{\Gamma[(N-M+1)/2]}{\Gamma[(N-M+2)/2]} \tag{9}$$

Therefore, from Eq. (8) the pooled unbiased estimator of $\sigma = \sigma_x$ is given by

$$\hat{\sigma}_p = S_p / c_5 > S_p \tag{10}$$

where $c_5 = \sqrt{\frac{2}{N-M}} \times \frac{\Gamma[(N-M+1)/2]}{\Gamma[(N-M)/2]} < 1$ for all $N = \sum_{i=1}^{M} n_i > M$ and is given in Eq. (9).

Letting $\nu = N - M$, it can be verified that for $N - M > 20$ a close approximation (to at least 4 decimal accuracy) for c_5 is given by $c_5 = \sqrt{\frac{2}{\nu}} \times \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)} \cong \frac{4\nu^2 - 8\nu + 4.8675}{(4\nu-3)(\nu-1)}$, $\nu > 20$,

and accuracy improves to at least 5 decimals once $\nu > 32$. Note that as $N - M \to \infty$, then $c_5 \to 1$ from below. The amount of bias in S_p, $B(S_p)$, as an estimator of $\sigma = \sigma_X$ is given by

$$B(S_p) = E(S_p - \sigma) = c_5 \times \sigma - \sigma = (c_5 - 1)\sigma < 0. \tag{11}$$

Eq. (11) clearly shows that S_p on-the-average underestimates σ in the long-run, while $\hat{\sigma}_p$ given in Eq. (10) is an unbiased estimator of σ_X. In order to compare the accuracy of the two estimators S_p and $\hat{\sigma}_p$, we proceed to compute their Mean Square Errors.

$$\text{MSE}(\hat{\sigma}_p) = V(\hat{\sigma}_p) = E(S_p^2/c_5^2) - \sigma^2 = E(S_p^2)/c_5^2 - \sigma^2 = \sigma^2/c_5^2 - \sigma^2$$

$$\text{MSE}(\hat{\sigma}_p) = (c_5^{-2} - 1)\sigma^2 \quad (12)$$

We next compute the $\text{MSE}(S_p) = V(S_p) + [B(S_p)]^2$. From Eq. (11) and the fact that S_p^2 is an unbiased estimator of σ^2, we obtain

$$V(S_p) = \sigma^2 - [E(S_p)]^2 = \sigma^2 - (c_5 \times \sigma)^2 = (1 - c_5^2)\,\sigma^2 \quad (13)$$

Combining equations (11) & (13) yields

$$\text{MSE}(S_p) = (1 - c_5^2)\sigma^2 + [(c_5 - 1)\sigma]^2 = 2(1 - c_5)\sigma^2. \quad (14)$$

Thus, the relative-efficiency of $\hat{\sigma}_p$ to S_p is given by

$$\text{REL-EFF}(\hat{\sigma}_p, S_p) = 2(1 - c_5)/(c_5^{-2} - 1) = 2c_5^2/(1 + c_5) \quad (15)$$

Eq. (15) shows that at N–M = 20, the REL-EFF($\hat{\sigma}_p$, S_p) = 98.141%, while at N–M = 50, it improves to 99.253%. The reader should not be surprised that the unbiased estimator $\hat{\sigma}_p$ is a bit less efficient than the biased estimator S_p because $1/c_5 > 1$ and the amount of bias in S_p (for N–M > 20) is not sufficiently large, especially because N–M generally exceed 50, to increase its MSE to the same level of $\hat{\sigma}_p$. Further, the nominal number of subgroups in the field of QC to set up a trial S-chart to monitor process variability is within M = 20 to 30, and for varying sample sizes n_i = 3 to 7, the value of N–M ranges from 40 to 180 (or larger) for which the REL-EFF($\hat{\sigma}_p$, S_p) ≥ 99.066% and improves to 99.792% at N–M = 180.

Of course, another point estimator of σ from M > 1 independent subgroups is the MLE given by $\hat{\sigma}_{MLE} = \sqrt{\sum_{i=1}^{M}(n_i - 1)S_i^2 / N}$. It can easily be shown the MLE: $\hat{\sigma}_{mle} = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 / n}$ at M = 1 can be generalized to M > 1 subgroups, as given herein. Using the same procedure as in equations (6), (7) and (8) it can be shown that the amount of bias is $B(\hat{\sigma}_{MLE}) = (c_5\sqrt{1-(M/N)} - 1)\sigma$. Using the fact that $E(\chi_\nu^2) = \nu$, it can be proven that the $V(\hat{\sigma}_{MLE}) = (1 - c_5^2)(1 - M/N)\sigma^2$, and hence its MSE($\hat{\sigma}_{MLE}$) = $2[1 - M/(2N) - c_5\sqrt{1-(M/N)}]\sigma^2$. Thus, the REL-EFF($\hat{\sigma}_{MLE}$, S_p) = $(1-c_5)/[1 - M/(2N) - c_5\sqrt{1-(M/N)}\,]$. Unlike the case of M = 1 subgroup, the REL-EFF ($\hat{\sigma}_{MLE}$, S_p) is poor. Only at M = 2 and N = 10, REL-EFF($\hat{\sigma}_{MLE}$, S_p) = 92.935%, and it monotonically decreases with increasing M and N. Its value at M = 25 and N = 79 is a mere 24.537%. Thus, $\hat{\sigma}_{MLE}$ should be avoided as an estimator of σ for M > 1 independent subgroups.

6. Applications to Quality-Control Charting for Unbalanced Designs

Traditionally to monitor process variation with varied sample sizes, the central line (CTLN) of an S-chart is given by $CTLN_S = S_p = \sqrt{\sum_{i=1}^{M}(n_i - 1)S_i^2 / (N - M)}$. However, the control limits will vary according to the size of the sample $n_i > 1$ (i = 1, 2, 3, …, M > 1) as described below. Note that some authors (such as D. C. Montgomery (2009), pp. 258-259) obtain an unbiased estimate of σ by using the largest proportion of the M subgroups with the same n_i as $\overline{S}/c_4$, and he refers to such a sample size as the model n. This procedure, however, does not use the information in all M subgroups.

As stated previously, the standard error of S is SE(S) = σ_S = $(1-c_4^2)^{1/2}\sigma_x$, and hence for a random sample of size n from a normal universe, the sample standard-error of S, *se*(S) = $\hat{\sigma}_S$ = $(1-c_4^2)^{1/2}\hat{\sigma}_x$. Traditionally, the QC literature uses S_p as a point estimator of σ_x , i.e., $\hat{\sigma}_S$ = $(1-c_4^2)^{1/2}S_p$, which is a biased estimator of σ_S = SE(S). As a result, the corresponding historical S-control limits for the case of differing n_i, i = 1, 2, ..., M, are given by

$$LCL(S_i) = S_p - 3[(1-c_4^2(n_i)]^{1/2}S_p \quad (16a)$$

$$UCL(S_i) = S_p + 3[(1-c_4^2(n_i)]^{1/2}S_p \quad 16b)$$

The alternative is to estimate σ_x, congruent with the case of balanced sampling scheme, by its unbiased estimator S_p/c_5 resulting in $\hat{\sigma}_S$ = *se*(S) = $(1-c_4^2)^{1/2}S_p/c_5$, but the $CNTL_S$ cannot be replaced with S_p/c_5 because the S-chart plots the biased estimates S_i on such a chart. If one decides to use the $CNTL_S = S_p/c_5$, then the unbiased estimates $S_i/c_4(n_i)$, where $c_4(n_i)$ is the value of c_4 at n_i, must be graphed to develop the S-chart. However, this is very awkward if one is developing an S-chart by hand-calculators, while statistical packages can easily perform this modification.

On the other hand, we do not encounter the above bias-problem in developing the $\bar{x}$-chart because its CNTL is given by $\text{CNTL}_{\bar{x}} = \bar{\bar{x}} = \sum_{i=1}^{M}\sum_{j=1}^{n_i} x_{ij} / N = \sum_{i=1}^{M} n_i \bar{x}_i / N$ which is an unbiased estimator of μ, and hence it will be best to alter the control limits to

$$\text{LCL}(\bar{x}_i) = \bar{\bar{x}} - 3\hat{\sigma}_p / \sqrt{n_i} \qquad (17a)$$

$$\text{UCL}(\bar{x}_i) = \bar{\bar{x}} + 3\hat{\sigma}_p / \sqrt{n_i} \qquad (17b)$$

where $\hat{\sigma}_p = S_p/c_5$ is given in Eq. (10). In equations (17) both $\bar{x}_i$ and $\bar{\bar{x}}$ are unbiased estimators of μ, while $\hat{\sigma}_p / \sqrt{n_i}$ is the unbiased estimator of $\text{SE}(\bar{x}_i) = \sigma_X / \sqrt{n_i}$.

7. The case of Balanced Sampling Scheme, i.e., n_i = n for all i = 1, 2, ..., M > 1

First, We wish to determine which estimator $\bar{S} = \sum_{i=1}^{M} S_i / M$, $\bar{S}/c_4$, S_p, or $\hat{\sigma}_p = S_p/c_5$ has the smallest MSE; it must be clear that the estimators $\bar{S}$ and $\bar{S}/c_4 = \bar{S}/c_4(n)$ can be used only when all the M > 1 sample sizes are identical, namely $n \geq 2$. Since the estimators $\bar{S}/c_4$ and S_p/c_5 are unbiased, their MSE's are equal to their variances; further, the value of $N = M \times n$. We have already shown that $\text{MSE}(\hat{\sigma}_p) = V(\hat{\sigma}_p) = (c_5^{-2} - 1)\sigma^2$ and $\text{MSE}(S_p) = 2(1 - c_5)\sigma^2$. We next compute the $\text{MSE}(\bar{S}/c_4) = V(\bar{S}/c_4) = E[(\bar{S}/c_4)^2] - \sigma^2 = E(\bar{S}^2)/c_4^2 - \sigma^2$. Therefore, we must compute $E[(\bar{S})^2]$ which implies that we must first square $\bar{S}$, as shown below before applying the linear operator E.

$$(\bar{S})^2 = \left(\frac{1}{M}\sum_{i=1}^{M} S_i\right)^2 = \frac{1}{M^2}\left[\sum_{i=1}^{M} S_i^2 + 2\sum_{i=1}^{M-1}\sum_{j>i}^{M} S_i S_j\right] \qquad (18)$$

Applying the expected-value operator E to the RHS of (18) results in

$$E[(\overline{S})^2] = \frac{1}{M^2}\Big[\sum_{i=1}^{M}\sigma^2 + 2\sum_{i=1}^{M-1}\sum_{j>i}^{M}E(S_i)\times E(S_j)\Big] = \frac{1}{M^2}[M\sigma^2 + 2\sum_{i=1}^{M-1}\sum_{j>i}^{M}(c_4\sigma)(c_4\sigma)]$$

$$= \frac{1}{M^2}[M\sigma^2 + M(M-1)c_4^2\sigma^2] = [1+(M-1)c_4^2]\sigma^2/M \qquad (19)$$

Combining Eq. (19) and $V(\overline{S}/c_4) = E(\overline{S}^2)/c_4^2 - \sigma^2$ results in

$$MSE(\overline{S}/c_4) = V(\overline{S}/c_4) = [c_4^{-2} + (M-1)]\sigma^2/M - \sigma^2 = (c_4^{-2} - 1)\sigma^2/M. \qquad (20a)$$

Similarly, $V(\overline{S}) = (1 - c_4^2)\sigma^2/M$ and $B(\overline{S}) = (c_4 - 1)\times\sigma$ lead to

$$MSE(\overline{S}) = [(1 - c_4^2)/M + (c_4 - 1)^2]\times\sigma^2 \qquad (20b)$$

Equations (20) clearly show that the $MSE(\overline{S}/c_4)$ and $MSE(\overline{S})$ diminish as n and M become larger. The corresponding REL-EFFs are given by

$$\text{REL-EFF}_1 = \text{REL-EFF}(\overline{S}/c_4, S_p) = 2M(1-c_5)/(c_4^{-2} - 1), \qquad (21a)$$

$$\text{REL-EFF}_2 = \text{REL-EFF}(\hat{\sigma}_p, S_p) = 2c_5^2/(1+c_5), \qquad (21b)$$

$$\text{REL-EFF}_3 = \text{REL-EFF}(\overline{S}, S_p) = \frac{2(1-c_5)}{(1-c_4^2)/M + (c_4-1)^2} \qquad (21c)$$

The 1st two REL-EFFs given in equations (21) are less than 100%, the smallest occurring at n = 2 and M = 2. However, they both rapidly approach 100% as n and M increase. The 3rd REL-EFF is 102.25% only at n = M = 2, and is less than 100% for all other combinations of n and M. Table 4 provides their values in percent for typical values of n at M = 25 subgroups. The REL-EFFs at n = 5 for 20 and 30 subgroups are REL-EFF_1 = 94.71%, REL-EFF_2 = 99.53%, REL-EFF_3 = 66.22%, and REL-EFF_1 = 94.76%, REL-EFF_2 = 99.69%, REL-EFF_3 = 55.63%, respectively.

Historically, two possibilities occur in practice: (1) The process variance is targeted (or

Table 4. The REL-EFFs given in Equations 21 for M = 25 Subgroups

n	3	4	5	6	7	8	9	10	11	12	13	14	15*
REL-EFF_1	91.26	93.42	94.74	95.63	96.26	96.73	97.10	97.40	97.64	97.84	98.01	98.15	
REL-EFF_2	99.25	99.50	99.63	99.70	99.75	99.79	99.81	99.83	99.85	99.86	99.88	99.88	
REL-EFF_3	46.33	54.38	60.46	65.16	68.88	71.89	74.37	76.46	78.23	79.75	81.08	82.24	

* Matlab ran out of computational capacity.

desired) at σ_T^2 in which case historical control limits are given by LCL(S) = $c_4\sigma_T - 3\times$ $(1-c_4^2)^{1/2}\sigma_T$ and UCL(S) = $c_4\sigma_T + 3(1-c_4^2)^{1/2}\sigma_T$. Note that historically the $CNTL_S$ is adjusted by the biasing-factor c_4 because E(S) = $c_4\sigma$, i.e., because S underestimates σ in the long-run, then the corresponding $CNTL_S$ is adjusted downward toward S. (2) The central line of the S-chart is estimated from M subgroups each of size n > 1 and historically is given by $CTLN_S = \overline{S}$, and because *se*(S) = $\hat{\sigma}_S = (1-c_4^2)^{1/2}\hat{\sigma}_x = (1-c_4^2)^{1/2}\overline{S}/c_4$, the historical control limits are given by LCL(S) = $\overline{S} - 3[(1-c_4^2(n)]^{1/2}\,\overline{S}/c_4 = B_3\overline{S}$ and UCL(S) = $\overline{S} + 3[(1-c_4^2(n)]^{1/2}\,\overline{S}/c_4 = B_4\overline{S}$, where $B_3 = 1 - 3[(1-c_4^2(n)]^{1/2}\,\overline{S}/c_4$ and $B4 = 1 + 3[(1-c_4^2(n)]^{1/2}\,\overline{S}/c_4$. These last 2 control limits are given in nearly every text book on QC. Correcting the $CTLN_S$ to the unbiased estimator $\overline{S}/c_4$ is not warranted because it is the biased S_i's that are plotted on an S-chart, and $\overline{S}$ is also a biased estimator of the same amount. If the $CNTL_S$ is adjusted to $\overline{S}/c_4$, then the values of S_i/c_4, the unbiased estimators, must be plotted on the S-chart.

8. Probability Limits for an S^2-Chart

In the field of QC, sometimes variation is monitored with the use of an S^2-chart. It is widely known that for a random sample of size n from a N(μ, σ^2)-universe the SMD of the variate $(n-1)S^2/\sigma^2$ follows a central chi-square distribution with (n–1) *df*, i.e., $(n-1)S^2/\sigma^2$ is distributed as χ^2_{n-1}. Consequently,

$$\Pr[\chi^2_{1-\alpha/2,n-1} \le \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}] = 1-\alpha \qquad (22)$$

where $\chi^2_{\alpha/2,n-1}$ is the upper $(\alpha/2)\times100$ percentage point (or the $1-\alpha/2$ quantile) of a central chi-square with (n–1) *df*. The most common values of False-Alarm probability, α, in the field of QC are α = 0.0027 and 0.001. Henceforth in this section, for simplicity of notation, we let $\chi^2_{\alpha/2}$ = $\chi^2_{\alpha/2,n-1}$ and Eq. (22) now yields

$$\Pr[\frac{\sigma^2}{n-1}\chi^2_{1-\alpha/2} \le S^2 < \frac{\sigma^2}{n-1}\chi^2_{\alpha/2}] = 1-\alpha \qquad (23)$$

Eq. (23) clearly shows that in order to maintain the confidence probability-level at $1-\alpha$, one must keep the subgroup sizes identically the same, namely n. Thus, we presently restrict discussions to the balanced sampling scheme. Again historically, two possibilities occur in practice: (1) The process variance is targeted (or desired) at σ_T^2 in which case Eq. (23) yields

$$LCL(S^2) = \frac{\sigma_T^2}{n-1}\chi^2_{1-\alpha/2}, \quad \text{and} \quad UCL(S^2) = \frac{\sigma_T^2}{n-1}\chi^2_{\alpha/2} \qquad (24)$$

(2) The process variance must be estimated from M > 1 initial-subgroups of identical sample sizes n > 1. Traditionally, the QC literature estimates σ^2 in Eq. (23) with $(\bar{S})^2$ as shown below.

$$LCL(S^2) = \frac{(\bar{S})^2}{n-1}\chi^2_{1-\alpha/2}, \quad \text{and} \quad UCL(S^2) = \frac{(\bar{S})^2}{n-1}\chi^2_{\alpha/2} \qquad (25)$$

The process variance σ^2 in Eq. (23) should not be estimated by $(\bar{S})^2$ because $(\bar{S})^2$ is a woefully biased estimator of σ^2. In fact, Eq. (19) shows that the amount of bias in $(\bar{S})^2$ as an estimator of σ^2 is given by

$$B[(\bar{S})^2] = E[(\bar{S})^2 - \sigma^2] = [1+(M-1)c_4^2]\sigma^2 / M - \sigma^2 = [1+(M-1)c_4^2](1/M-1)\sigma^2 \qquad (26)$$

Eq. (26) shows that the $B[(\bar{S})^2] < 0$ for all n and M, and unfortunately, it increases rapidly and monotonically in absolute value from $|-0.94179|\sigma^2$ at n = 5 and M = 2, to $|-21.31752|\sigma^2$ at n = 5 and M = 25. We can correct this biasing-problem by using the unbiased estimator $S_p^2 = \sum_{i=1}^{M} S_i^2 / M$ in lieu of $(\bar{S})^2$ in Eq. (25), notwithstanding the fact that it is the unbiased estimators S_i^2, i = 1, 2, …, M that are plotted on the S^2-chart. The alternative of replacing $(\bar{S})^2$ with $(\bar{S}/c_4)^2$ in Eq. (25) dramatically reduces the amount of bias. For example, at n = 5 and M = 2, the amount of bias is $0.065884\sigma^2$ and reduces to $0.005271\sigma^2$ at n = 5 and M = 25. Of course, these two recommended alternatives will lead to less powerful control charts [because both S_p^2 and $(\bar{S}/c_4)^2 > (\bar{S})^2$] but less false alarms. If the sampling scheme is unbalanced, probability-limits for control charts should be avoided.

9. Summary and Conclusions

In section 2, we established that the maximum likelihood estimator, $\hat{\sigma}_{\text{mle}}$, for M = 1 subgroup is the most accurate estimator of a normal population standard deviation $\sigma = \sigma_X$; in fact V($\hat{\sigma}_{\text{mle}}$) < V($\hat{\sigma}_{MD}$) for all n ≥ 2. In sections 5, 6 & 7, for M > 1 subgroup, the $\hat{\sigma}_{\text{MLE}}$ was found to be the least accurate estimator of σ. We found that the unbiased estimator of $\sigma = \sigma_X$ is given by $\hat{\sigma}_p = S_p / c_5$, where c_5 is given by Eq. (9). As a result, it is recommended that for M > 1 subgroups of differing sample sizes, the SE(S) = $(1-c_4^2)^{1/2}\sigma_X$ be estimated by $\hat{\sigma}_S$ = *se*(S) = $(1-c_4^2)^{1/2} S_p / c_5$; of course this leads to wider control limits and less powerful control charts. Nevertheless, $(1-c_4^2)^{1/2} S_p / c_5$ is an unbiased estimator of SE(S). In section 8, we showed that the QC literature should consider replacing $(\bar{S})^2$ in probability limits of the S^2-chart, in the case of balanced sampling scheme, with the unbiased estimator $S_p^2 = \sum_{i=1}^{M} S_i^2 / M$, M > 1.

References

Chou, Chao-Yu (June 1992). Exact Control Limits for Monitoring Process Variation, Auburn University dissertation, directed by S. Maghsoodloo.

Harter, H. L. (Dec. 1960). The Annals of Mathematical Statistics, Vol. 31, No. 4, pp. 1122-1147.

Hool, J.N. and Maghsoodloo, S., (June 1980). Normal Approximation to Linear Combinations of Independently Distributed Random Variables, *AIIE Transactions*, Vol. 12, No. 1, pp. 140-144.

Hogg, V. H. and Tanis E. A. (2010). *Probability and Statistical Inference*, 8th edition, Prentice Hall.

Johnson N. L., Kotz, S. and Balakrishnan, N. (1994). Continuous Univariate Distributions, Vol. 1. Second Edition, John Wiley & Sons, Inc.

Kendall, M.G. and Stuart A., (1963). *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin & Company Limited.

Kendall, M.G. and Stuart A., (1967). *The Advanced Theory of Statistics*, Vol. 2, Charles Griffin & Company Limited.

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*, 6th edition, John Wiley & Sons, Inc.

Ostle, B. and Malone, L. C. (1988). Statistics in Research: Basic Concepts and Techniques for

Research Workers, 4th ed., Iowa State University Press/Ames, pp. 80-81.
Pearson, E. S. (July 1926). A further note on the distribution of range in samples taken from a normal population. Biometrika, Vol. 18, p. 173.
Pearson, E. S. (Nov. 1932). The percentage limits for the distribution of range in samples from a normal population. Biometrika, Vol. 24, p. 404.
Tippett, L. H. C. (Dec. 1925). On the extreme individuals and the range of samples taken from a normal population, Biometrika, Vol. 17, p. 364.