2009

Estimating σ and Improving Control Limits with RWAV for Normal and Nonnormal Processes

Xia Pan

Jeffrey E. Jarrett
University of Rhode Island, jejarrett@uri.edu

Citation/Publisher Attribution
Pan, Xia and Jarrett, Jeffrey E.(2009)'Estimating and Improving Control Limits with RWAV for Normal and Nonnormal Processes',Quality Engineering,21:1,72 — 78. doi: 10.1080/08982110802450769
Available at: http://dx.doi.org/10.1080/08982110802450769

Quality Engineering
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597292

Estimating and Improving Control Limits with RWAV for Normal and Non-normal Processes

Xia Pan [a]; Jeffrey E. Jarrett [b]
[a] Linghan College, Sun Yat-sen University, Guangzhou, China [b] University of Rhode Island, Kingston, Rhode Island

Online Publication Date: 01 January 2009

To cite this Article Pan, Xia and Jarrett, Jeffrey E.(2009)'Estimating and Improving Control Limits with RWAV for Normal and Non-normal Processes',Quality Engineering,21:1,72 — 78
To link to this Article: DOI: 10.1080/08982110802450769
URL: http://dx.doi.org/10.1080/08982110802450769

Quality Engineering, 21:72–78, 2009

ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982110802450769

Estimating σ and Improving Control Limits with RWAV for Normal and Non-normal Processes

Xia Pan[1],
Jeffrey E. Jarrett[2]
[1]Linghan College, Sun Yat-sen University, Guangzhou, China
[2]University of Rhode Island, Kingston, Rhode Island

ABSTRACT In this article, we compare the statistical properties of $\overline{R}$, $\overline{S}$, and RWAV methods for estimation of variance of a process for quality control purposes. We investigate effects of non-normality for different estimators. Our results indicate that RWAV gives the best estimates of the standard deviations for both normal and non-normal processes. We recommend the construction of control charts with RWAV.

KEYWORDS estimation, mean and range charts, process control, quality control

INTRODUCTION

Statistical process control (SPC) is useful for quality assurance in production and other operations. The Shewhart control charts, i.e., $\overline{X}$ charts, R charts, and S charts, are one of the most popular tools of SPC. The control charts have two phases. Phase I is the chart construction phase, and Phase II is the monitoring phase. In Phase I, one needs to make sure that the underlying process is in control before he formally estimates the central line and control limits. We also imply that one assumes (and verifies) that the underlying process is normal because the sample size employed in Shewhart control chart methods is usually small (less than 30 in most situations). In Phase II, the quality analyst monitors the process for assignable causes of variation associated with unexpected changes in the process parameters; i.e., the process mean and process variability (Shewhart, 1931). Any advantage an estimator may have relates to the accuracy of parameter estimates (Del Castillo, 1996).

For m samples of size n, the sample range and sample standard deviation are

$$\overline{R} = \sum_{i=1}^{m} R_i/m \qquad [1]$$

$$\overline{S} = \sum_{i=1}^{m} S_i/m \qquad [2]$$

and are estimates of the process standard deviation. If the underlying process is normal, the unbiased estimates are, respectively,

$$\hat{\sigma}(R) = \overline{R}/d_2(n) \qquad [3]$$

$$\hat{\sigma}(S) = \overline{S}/c_4(n) \qquad [4]$$

Address correspondence to Jeffrey E. Jarrett, University of Rhode Island, College of Business Administration, 7 Lippitt Road, Ballentine, Kingston, RI 02881. E-mail: jejarrett@mail.uri.edu

where $d_2(n)$ and $c_4(n)$ are dependent on sample sizes, and

$$c_4(n) = \sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma((n-1)/2)} \quad [5]$$

In Phase I of SPC, we distinguish within-sample variation from between-sample variation. One advantage of this distinction is that information acquired within the samples reflects the process variation when the process mean is in control and the process variation is constant. The second advantage is that both $\overline{R}/d_2(n)$ and $\overline{S}/c_4(n)$ measure within-sample variation since estimators in Phases I and II may not be, or do not have to be, the same. If necessary, we employ a different estimator for Phase I than for Phase II to obtain a better estimate of the within-sample variability.

Sample sizes may change from sample to sample for a variety of reasons (Montgomery, 2005, see pp. 109–111 and 123). To accommodate variable sample sizes, Montgomery (2005, pp. 194–247) suggests several estimators applying R and S . Their recommendation weights the variance of each sample and corrects the bias with the coefficient of aggregated sample size, $f+1=\sum_{i=1}^{m} n_i - m + 1$. Therefore, the unbiased estimator for within sample variability is

$$\hat{\sigma}(S) = \frac{1}{c_4(f+1)}\sqrt{\sum_{i=1}^{m}(n_i-1)S_i^2\Big/\left(\sum_{i=1}^{m} n_i - m\right)} \quad [6]$$

and

$$c_4(f+1) = \sqrt{\frac{2}{f}}\,\frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma(f/2)}. \quad [7]$$

A proof of [7] is in Appendix 1. When f is very large then [7], is equal to unity and becomes the root weighted average of variance, RWAV:

$$\mathrm{RWAV} = \sqrt{\sum_{i=1}^{m}(n_i-1)S_i^2/left(\sum_{i=1}^{m} n_i - m)}. \quad [8]$$

As $\hat{\sigma}(S)$, or RWAV$/c_4(f+1)$, is unbiased, we consider the slightly biased estimator by RWAV in the following sections. We begin by analyzing possible values of $c_4(f+1)$ other than unity in Table 1. For the sample size n, denote $\bar{n}=\sum_{i=1}^{m} n_i/m$ as the mean sample size. Hence, we have $f=(\bar{n}-1)m$. The bottom row of Table 1 lists the reference values of $c_4(n=\bar{n})$ for comparison. As suggested by Burr (1976) and Nelson (1984), $c_4(f+1)$ is approximately one; therefore, we use RWAV (Eq. [8]) to estimate the process's standard deviation in Phase I. This is convenient since it does not require one to check or calculate the bias correction factor c_4.

Among the many estimators proposed in Burr (1976) and Nelson (1984), all except RWAV/$c_4(f+1)$ and RWAV reduce to $\overline{R}/d_2$ or $\overline{S}/c_4(n)$, when sample sizes are equal. However, RWAV/$c_{4](f+1)}$ is more efficient than $\overline{R}/d_2(n)$ and $\overline{S}/c_4(n)$ since equal sample sizes is a special case. Moreover, Burr (1976) did suggest the use of RWAV$/c_4(f+1)$ and RWAV when sample sizes are not equal. Later, we will recommend using RWAV for equal sample sizes as well when we focus on the quality of estimators.

EFFICIENCY AND ACCURACY OF ESTIMATORS

With the definition of statistical relative efficiency that the more efficient estimator has smaller mean squared error (MSE)—i.e., the expected squared

TABLE 1 Values of $c_4(f+1)$

	$\bar{n}$						
m	2	4	5	6	10	15	20
10	0.9754	0.9917	0.9938	0.995	0.9972	0.9982	0.9987
20	0.9876	0.9958	0.9969	0.9975	0.9986	0.9991	~1
30	0.9917	0.9972	0.9979	0.9983	0.9991	~1	~1
50	0.9950	0.9983	0.9988	0.9990	~1	~1	~1
80	0.9969	0.9990	0.9992	~1	~1	~1	~1
100	0.9975	0.9992	~1	~1	~1	~1	~1
150	0.9983	~1	~1	~1	~1	~1	~1
$c_4(n=\bar{n})$	0.7979	0.9213	0.9400	0.9515	0.9727	0.9823	0.9868

TABLE 2 **Relative Efficiency between RWAV/$c_4(f+1)$ and $\overline{S}/c_4(n)$**

	$\bar{n}$ or n								
m	2	3	4	5	6	7	8	9	10
10	0.8967	0.9261	0.9435	0.9545	0.9620	0.9674	0.9715	0.9747	0.9772
20	0.8866	0.9206	0.9397	0.9516	0.9596	0.9654	0.9698	0.9732	0.9759
30	0.8831	0.9187	0.9384	0.9506	0.9588	0.9648	0.9692	0.9726	0.9754
40	0.8814	0.9178	0.9378	0.9501	0.9584	0.9644	0.9689	0.972	
50	0.8803	0.9172	0.9374	0.9498	0.9582	0.9642			

deviation of the sample estimate from the population parameter—we notice that the relative efficiency for unbiased estimators just means that the estimator with smaller variance is more efficient. It is known that when the underlying processes are normal, $\overline{R}/d_2(n)$ and $\overline{S}/c_4(n)$ are unbiased, and $\overline{R}/d_2(n)$ is less efficient than $\overline{S}/c_4(n)$. We can further show that the relative efficiency between the two unbiased estimators, RWAV/$c_4(f+1)$ and $\overline{S}/c_4(n)$, is

$$\frac{\mathrm{Var}(\frac{\mathrm{RWAV}}{c_4(f+1)})}{\mathrm{Var}(\overline{S}/c_4)} = \frac{\sigma^2(1-c_4^2(f+1))/c_4^2(f+1)}{\sigma^2(1-c_4^2(n))/(m*c_4^2(n))} = \frac{mc_4^2(n)*(1-c_4^2(f+1))}{c_4^2(f+1)*(1-c_4^2(n))} \quad [9]$$

For certain m, the number of samples in Phase I, the variance of RWAV/$c_4(f+1)$ does not depend on how the sample sizes vary. Instead, it depends on the mean sample size $\bar{n}$, since $f=(\bar{n}-1)m$. The ratio [9] compares the efficiency of RWAV/$c_4(f+1)$ at ($\bar{n}$, m) where sample sizes vary and that of $\overline{S}/c_4(n)$ at (n, m) where sample sizes are equal. In Table 2, we find the numerical values of [9] indicating that RWAV/$c_4(f+1)$ is more efficient than $\overline{S}/c_4(n)$ in measuring within-sample variability (hence, it is also more efficient than $\overline{R}/d_2(n)$). The relative efficiency in [9] is more sensitive to sample size, n, than the number of samples, m, while the relative efficiency between $\overline{R}/d_2(n)$ and $\overline{S}/c_4(n)$ is not related to m. In Phase I where n is often 4 or 5 and m varies between 30 and 50, we observe in Table 2 that $\overline{S}/c_4(n)$ is roughly 5% less efficient than RWAV/$c_4(f+1)$.

Previously, others debated for decades on whether the mean square error (MSE) or mean absolute error (MAE) should be used when assessing the quality of estimates and/or forecasts (Brandon and Jarrett, 1974, 1986; Hanke and Wichern, 2005; Jarrett, 1991). The MSE penalizes greatly for large errors rather than penalizing proportionally for all errors. MSE is most appropriate when economic circumstances require severe penalties for large errors. The use of MAE is the common in industrial applications. In fact, the same logic of using MAE is also found in Burr (1976) where he emphasized that the reason for using $\overline{R}$ and $\overline{S}$ in the estimation of control limits is to avoid the greater penalties for larger errors. And this idea has been adopted in SPC since Shewhart (1931).

Our purpose in this study is not to debate whether MSE or MAE is more appropriate. We do note that for a given probability distribution there exists a relationship between the MSE and MAE; i.e., the expected absolute deviation of the sample estimate from the population parameter. For example, the variance of a normal distribution is 1.25 times the MAE. And for other distributions the multiplier is different from 1.25. Therefore, if one estimator has smaller MSE than the other, it must also have smaller MAE. In other words, the two criteria are consistent and will lead to the same decision in the comparison of estimators. In the following we will use MAE as our criterion.

When the underlying processes are not normal, the estimators mentioned above for process standard deviation are all biased. Hence, using variance to assess the relative efficiency as in [9] is not available anymore. To assess the estimation quality of an estimator for the parameter σ, we use the estimated MAE,

$$\mathrm{MAE} = \left(\sum_{j=1}^{k} \left|\hat{\sigma}_j - \sigma\right|\right)/k \quad [10]$$

where k is the number of replications of taking the same sampling of m samples of size n, and $\hat{\sigma}_j$ is the result of estimate of σ for the jth sampling, no matter what estimator is used. The process is

TABLE 3 MAE of Estimators

$k=100$											
$m=10$											
$\bar{n}$ or n	2	3	4	5	6	7	8	10	15	20	25
$\overline{R}/d_2(n)$	19.0%	12.5%	10.6%	9.9%	8.9%	8.3%	7.1%	6.8%	5.3%	4.7%	4.5%
$\overline{S}/c_4(n)$	19.0%	12.6%	10.6%	9.8%	8.3%	7.6%	6.7%	5.8%	4.8%	3.6%	3.9%
RWAV	17.7%	12.7%	10.6%	9.5%	8.2%	7.4%	6.4%	5.7%	4.8%	3.7%	3.8%
$m=20$											
$\bar{n}$ or n	2	3	4	5	6	7	8	10	15	20	25
$\overline{R}/d_2(n)$	11.7%	9.3%	7.6%	6.3%	5.7%	5.1%	5.4%	4.7%	3.9%	3.6%	3.9%
$\overline{S}/c_4(n)$	11.7%	9.3%	7.6%	6.1%	5.3%	4.9%	4.7%	4.5%	3.3%	2.8%	3.1%
RWAV	10.1%	8.7%	7.8%	5.9%	5.1%	4.8%	4.6%	4.3%	3.3%	2.8%	3.0%
$m=50$											
$\bar{n}$ or n	2	3	4	5	6	7	8	10	15	20	25
$\overline{R}/d_2(n)$	9.02%	6.43%	5.14%	4.59%	3.45%	3.25%	3.20%	2.91%	2.33%	2.09%	2.15%
$\overline{S}/c_4(n)$	9.02%	6.36%	5.01%	4.55%	3.33%	3.27%	3.10%	2.66%	2.07%	1.92%	1.64%
RWAV	8.17%	5.98%	4.77%	4.40%	3.33%	3.20%	2.97%	2.69%	2.02%	1.86%	1.61%
$m=100$											
$\bar{n}$ or n	2	3	4	5	6	7	8	10	15	20	25
$\overline{R}/d_2(n)$	5.98%	3.94%	3.47%	3.14%	2.37%	2.02%	2.13%	2.24%	1.86%	1.54%	1.33%
$\overline{S}/c_4(n)$	5.98%	3.93%	3.45%	3.04%	2.34%	1.97%	2.07%	2.01%	1.58%	1.36%	1.02%
RWAV	5.33%	3.70%	3.29%	2.95%	2.25%	1.97%	2.02%	1.98%	1.56%	1.35%	1.00%

simulated for certain probability distributions that are not necessarily normal.

Now that all the estimators mentioned above are biased for non-normal populations, the use of the bias correction faction c_4 in RWAV/$c_4(f+1)$ is not necessary anymore. We argue that for simplification and convenience we can directly use RWAV to estimate the process standard deviation. We will use MAE as the criterion to compare RWAV and the other estimators mentioned above in non-normal environments. We will show that RWAV is the most efficient estimator in general situations regardless of whether the sample sizes are equal or not and regardless of whether the underlying process is distributed normally or not. The benefit of using RWAV instead of RWAV/$c_4(f+1)$ is that the values of $c_4(f+1)$ are not available in common reference tables.

We first compare RWAV with the other estimators in normal environment. Table 3 shows the simulation results of MAE for the unbiased $\overline{R}/d_2(n)$ and $\overline{S}/c_4(n)$ and the biased RWAV. It is clear that RWAV, although biased, is better than the other two in that it has smaller MAE. Note, also, that $\overline{R}/d_2(n)$ has smaller MAE than $\overline{S}/c_4(n)$ as the sample size, n, increases, which is consistent with the result in [9]. In the comparison, $\overline{R}/d_2(n)$ and $\overline{S}/c_4(n)$ are obtained in the environment of equal sample size simulation. The RWAV is obtained in the environment of equal and unequal sample size simulation, while keeping $\bar{n}=n$. Generally, we let 50% of the samples have size n, while the other 25% samples have greater sizes and another 25% have smaller sizes.

NON-NORMALITY EFFECTS

A good control chart should be robust for the underlying processes of non-normal distributions. Burr (1967) and Schilling and Nelson (1976) reported that $\bar{x}$ chart is robust and useful when processes are not normal. We studied the effects of non-normality of the underlying process to ascertain whether RWAV is again the most accurate estimator. By simulation, we study the effects of logarithmic normal and gamma processes for different parameter values and compare the MAEs of RWAV, $\overline{R}/d_2(n)$, and $\overline{S}/c_4(n)$. Table 4 shows the winning estimator (marked as WINNER in the table) among these three estimators that has the smallest MAE for different lognormal processes and sample sizes. The MAEs are calculated for 200 repeated simulations. Again, RWAV is simulated in the way of equal and unequal sample sizes while keeping $\bar{n}=n$.

In reality, the underlying process may vary greatly; hence, all the estimators are not entirely unbiased. At the bottom row of the table, we report

TABLE 4 **Comparison Following Lognormal Distribution**

m	20		80	
$\bar{n}$ or n	minMAE	WINNER	minMAE	WINNER
$\mu = 0$		True std = 1		
3	0.1782	RWAV	0.107	RWAV
4	0.1608	$\bar{S}/c_4$	0.0974	RWAV
5	0.1429	RWAV	0.0816	RWAV
10	0.1071	$\bar{S}/c_4$	0.0543	RWAV
$\mu = 0.6931$		True std = 2		
3	0.1902	$\bar{S}/c_4$	0.1068	RWAV
4	0.1711	$\bar{S}/c_4$	0.085	RWAV
5	0.1422	RWAV	0.0777	RWAV
10	0.106	RWAV	0.0687	RWAV
$\mu = 0$		True std = 0.7434		
3	0.1556	$\bar{S}/c_4$	0.08	RWAV
4	0.1313	$\bar{S}/c_4$	0.0801	RWAV
5	0.1276	RWAV	0.0652	RWAV
10	0.0793	RWAV	0.0421	RWAV
$\mu = 0.6931$		True std = 1.4868		
3	0.1488	$\bar{S}/c_4$	0.0864	RWAV
4	0.128	RWAV	0.0721	RWAV
5	0.114	RWAV	0.0668	RWAV
10	0.0917	$\bar{S}/c_4$	0.0425	RWAV
Percentage	$\bar{R}/d_2$	$\bar{S}/c_4$	RWAV	
Shares	0%	19%	81%	

*WINNER means the estimator with smallest MAE among the three under certain process conditions. For example, if a cell is the WINNER and is column RWAV, it means RWAV possesses a smaller MAE than $\bar{S}/c_4$ and $\bar{R}/d_2$. MAE is the WINNER and is shown at the cell on the left.

each estimator's percentage share out of the 200 repeated simulations. From Table 4, we observe that in 81% of the cases where the underlying processes follow a lognormal distribution, RWAV is the most accurate estimator. In only 19% of the cases is $\overline{S}/c_4(n)$ the best. The results for gamma distributions (the data are not reported here) are similar. Therefore, in general, using RWAV yields more accurate control limits than other estimators.

Often, production processes that are not normal contain skewness and are often approximated by either a logarithmic normal or gamma functions. Hence, our results support the use of RWAV. Last, we should note that Does and Schriever (1992) and De Mast and Roes (2004) also attempted similar pooling of variances to obtain the within-sample standard deviation for control charts. Our results are simpler and more straightforward but agree with their general findings.

CONCLUSION

Estimation of the process standard deviation is important in SPC and is fundamental to the construction of control charts. Our aim was to find simple and straightforward estimates to construct the control chart in the best manner. Based on the evidence presented, we recommend the use of RWAV for both equal and unequal sample sizes. Furthermore, we justify the use of the biased RWAV for non-normal underlying processes. Compared with others, RWAV is an efficient estimator for the within-sample variability of all underlying processes, no matter whether normal or non-normal. Consequently, we recommend the use of RWAV to construct Phase I control limits. We feel that employing RWAV in Phase I leads to better construction of quality control charts. In addition, calculating RWAV is no longer a problem and is universally beneficial to those in the quality control profession.

Note that RWAV is not available for Phase II. Since one checks for stability in Phase II, we monitor samples in Phase II one at a time. Estimators such as $\overline{R}/d_2(n)$ and $\overline{S}/c_4(n)$ are still useful in Phase II. Last, the steps in control chart construction are as follows:

- Phase I, the central line is obtained in the usual manner (Central Line $= \bar{\bar{x}}$) and we construct the control limits for sample size, n, by

$$UCL = \bar{\bar{x}} + 3 * \text{RWAV}/\sqrt{n}$$
$$\text{and } LCL = \bar{\bar{X}} - 3 * \text{RWAV}\sqrt{n} \qquad [11]$$

- Phase II, we construct the control chart for R as follows:

$$UCL = \overline{R} + 3 * d_3\hat{\sigma} = \overline{R} + 3 * d_3 * \text{RWAV}$$
$$\text{Central Line} = \overline{R} \qquad [12]$$
$$LCL = \overline{R} - 3 * d_3\hat{\sigma} = \overline{R} - 3 * d_3 * \text{RWAV}$$

Also, we construct the control chart for S by

$$UCL = \overline{S} + 3 * \sqrt{1 - c_4^2}\hat{\sigma}$$
$$= \overline{S} + 3 * \sqrt{1 - c_4^2} * \text{RWAV}$$
$$\text{Central Line} = \overline{S} \qquad [13]$$

$$LCL = \overline{S} - 3 * \sqrt{1 - c_4^2}\hat{\sigma} = \overline{S} - 3 * \sqrt{1 - c_4^2} * \text{RWAV}$$

ACKNOWLEDGEMENTS

The authors thank the editor and two anonymous referees for their assistance.

ABOUT THE AUTHORS

Jeffrey E. Jarrett, Ph.D., is former Chairperson of the Department of Management Science and Professor of Management Science and Statistics at the University of Rhode Island. He holds degrees from the University of Michigan and New York University where he studied with W. Edwards Deming, among others. He has published extensively in *The Accounting Review, Decision Sciences, Journal of Business Finance and Accounting, Journal of Accounting Research, Journal of Finance, Management Science, OMEGA: The International Journal of Management Science, Journal of Business Forecasting, Statistical Software Newsletter, Journal of Business and Economic Statistics, Journal of the American Statistical Association, Atlantic Economic Journal, Review of Business and Economic Research, Statistics and Computing, Midwestern Journal of Business and Economics, Financial Engineering and Japanese Markets, Economic and Financial Modeling, International Journal of Business and Economics, International Journal of Forecasting Journal of Applied Statistics, International Journal of Production Economics, Computational Statistics and Data Analysis, Technometrics, International Journal of Industrial and Systems Engineering*, and *Applied Economics*, among others.

Xia Pan, Ph.D., received his Ph.D., degree from the University of Rhode Island. He published extensively in the *International Journal of Production Economics, Computational Statistics and Data Analysis, Journal of Industrial and Systems Engineering*, and the *Journal of Applied Statistics*. He is a frequent contributor to professional meetings, including the Production and Operations Management Society (POMS), the Decision Sciences Institute (DSI), and regional organizations in Asia. His current teaching and research interests include statistical process control, operations management, and multivariate methods in industrial and business research.

REFERENCES

Brandon, C., Jarrett, J. (1974). Accuracy of financial forecasts. *Financial Review*, 9:29–45.

Brandon, C., Jarrett, J. (1986). Improving accuracy of forecasts by a cost-effective method. *Futures: The Journal of Forecasting and Planning*, 1:78–83.

Burr, I. J. (1967). The effect of nonnormality on constants for xbar and R charts. *Industrial Quality Control*, 23(11):563–569.

Burr, I. W. (1976). *Statistical Quality Control Methods*. New York: Marcel Dekker.

Del Castillo, E. (1996). Evaluation of the run length distribution for X-bar charts with unknown variance. *Journal of Quality Technology*, 28:189–201.

De Mast, J., Roes, K. C. B. (2004). Robust individuals control chart for exploratory analysis. *Quality Engineering*, 16(3):407–421.

Does, R. J. M. M., Schriever, B. F. (1992). variables control chart limits and tests for special causes. *Statistica Neerlandica*, 46(4): 229–245.

Hanke, J., Wichern, D. W. (2005). *Business Forecasting*, 8th ed. Englewood Cliffs, NJ: Pearson Prentice-Hall.

Jarrett, J. E. (1991). *Business Forecasting Methods*, 2nd ed. Oxford: Basil Blackwell.

Montgomery, D. C. (2005). *Introduction to Statistical Quality Control*, 5th ed. New York: John Wiley & Sons.

Nelson, L. S. (1984). The Shewhart control chart—Tests for special causes. *Journal of Quality Technology*, 16(4):237–239.

Schilling, E. G., Nelson, P. R. (1976). The effect of nonnormality on the control limits of $\bar{x}$ charts. *Journal of Quality Technology*, 8(4):183–188.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. New York: D. Van Nostrand Company.

APPENDIX: THE PROOF ON EQ. [7]

Proof: Denote $\text{RWAV} = w$, then $w = \sqrt{\sum_{i=1}^{m} (n_i - 1)s_i^2 \Big/ \left(\sum_{i=1}^{m} n_i - m\right)}$.

Since for each sample of size n_i, $\frac{(n_i-1)s_i^2}{\sigma^2} = \chi^2_{(n_i-1)}$ and the samples are independent of each other, the sum of these variables of $\frac{(n_i-1)s_i^2}{\sigma^2}$, denoted as x, is also a chi-square distribution.

Also, sum of degrees of freedom for each sample is $f = \sum_{i=1}^{m} n_i - m$ and

$$x = \sum_{i=1}^{m} (n_i - 1)s_i^2/\sigma^2 = \frac{f * w^2}{\sigma^2} = \chi^2_f.$$

From the probability density function of chi-square for this random variable,

$$p(x) = \frac{(x)^{f/2-1} e^{-x/2}}{2^{f/2}\Gamma(f/2)},$$

we have

$$p(x)dx = \frac{(x)^{f/2-1} e^{-x/2}}{2^{f/2}\Gamma(f/2)} dx = \frac{(fw^2/\sigma^2)^{f/2-1} e^{-fw^2/2\sigma^2}}{2^{f/2}\Gamma(f/2)} d(fw^2/\sigma^2).$$

Let $v = x/2 = fw^2/(2\sigma^2)$ so $w = \sigma\sqrt{x/f} = \sigma\sqrt{2v/f}$. Then the expectation of w is

$$E(w) = \sigma \int_0^\infty \sqrt{\frac{2}{f}} \frac{v^{(f+1)/2-1} e^{-v}}{\Gamma(f/2)} dv$$

Using the definition $\Gamma(k) = \int_0^\infty v^{k-1} e^{-v}\, dv$, we obtain

$$E(w) = \sigma\sqrt{\frac{2}{f}} \frac{\Gamma(\frac{f+1}{2})}{\Gamma(f/2)} = g_4\sigma$$

where

$$g_4 = \sqrt{\frac{2}{f}} \frac{\Gamma(\frac{f+1}{2})}{\Gamma(f/2)} \quad \text{and} \quad f = \sum_{i=1}^{m} n_i - m.$$

Similarly, the variance of w is $\mathrm{Var}(w) = E(w^2) - E(w)^2$. From the definition of w, we have

$$E(w^2) = \sum_{i=1}^{m} (n_i - 1)E(s_i^2) \Bigg/ \left(\sum_{i=1}^{m} n_i - m \right) = \sigma^2$$

Therefore,

$$\mathrm{Var}(w) = \sigma^2 - E(w)^2 = \sigma^2 - g_4^2\sigma^2 = (1 - g_4^2)\sigma^2$$