# Mean Square Error

Glenn J. Battaglia
*AMP Incorporated*

## ABSTRACT

**Mean square error (MSE) is an accepted measure of control and of quality but its practical use as a measure of quality has been limited. This note presents the fundamentals of MSE, introduces a normalized, new version of MSE which overcomes the most severe shortcomings of MSE, discusses breaking MSE into diagnostic components which can assist to obtain maximum improvement at minimum cost, shows how to use the components of MSE to obtain a valid assessment of statistical control, illustrates how to evaluate six-sigma quality accurately for processes which are not in statistical control, and suggests how to employ MSE effectively.**

## THE MEAN SQUARE ERROR

Mean square error (MSE) is an old, proven measure of control and quality.[1] In statistical process control (SPC) literature it is often referred to as mean square deviation (MSD). MSE equals the mean of the squares of the deviations from target, i.e.,

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m}(x_i - T)^2 \qquad (1)$$

where

$x_i$ = ith value of a group of m values,

$T$ = target or intended, i.e., desired, value for the product variable of interest.

MSE calculated from a sample of product taken out of a process is just an estimate of the MSE of the entire population manufactured by the process. To know the true MSE for a manufacturing period *all* product made during that period would have to be measured. Similarly, the true capability of a process would be known in its entirety only if 100% of the output product were measured.

In principle it is possible to estimate MSE by applying equation (1) to a sample of process output. In this case $x_i$ would signify the value of the $i^{th}$ member of the sample. However, a different formula that breaks MSE into useful components should be used. It is introduced and discussed below.

## REASONS FOR USING MSE

A valid index of control and quality is essential for effective management of processes. It is necessary for

- assessing process control and product quality
- setting meaningful goals for control and quality
- monitoring and motivating progress toward goals
- evaluating differences in equipment, procedures, etc.
- prioritizing study and corrective effort
- managing efforts for continuous improvement.

The components of MSE explained below also help to identify means for improving processes and quality.

Other common measures of quality and control are valid only under special conditions. For example, $C_{pk}$ is not meaningful for a process which is not in statistical control, and PPM (parts per million nonconforming), as it is often estimated, can be grossly wrong unless the process of interest is in statistical control. Furthermore, both of these common measures are questionable when they are applied to populations that are not normally distributed. For such processes, the capability indices do not describe what fraction of the process output will fall between specification limits and the PPM estimates can be severely in error.

More important, these measures depend to some extent upon an underlying assumption which is usually fallacious namely that a product is perfectly good up to some specification limit and completely bad beyond that limit. It obviously is not true that a nominally 30-µinch gold plated contact is perfect if the gold layer is 30.001 µinch thick but useless if the layer thickness is only 29.999 µinch. Taguchi has shown that, as a reasonable approximation, the quality cost associated with a product is proportional to its MSE and does not change abruptly from zero to a maximum as a specification limit is crossed.

The idea that the utility of a product decreases with the square of the deviation from the target value is also consistent with common sense. Consider the diameter of a bolt. If it is only a little too large or a little too small, the bolt will still function. If the diameter becomes substantially too large, however, the bolt will not fit properly into some of the nuts or tapped holes into which it is to be inserted. In some cases it will be hard or even impossible to insert the bolt. On the other hand the probability of a bolt pulling out of a tapped hole increases ever more strongly as the diameter approaches the inner diameter of the threaded hole. The results are potentially disastrous.

MSE is a good index of control and of quality even when the process is not in statistical control and even when the product is not normally distributed. Components of MSE, which are explained below, can be used to determine whether a process is operating in statistical control. MSER, which is a derivative of MSE that is described below, is of particular interest because it can be used to establish whether a process meets six-sigma criteria.

MSE coexists comfortably with $C_{pk}$, since MSE is not a competing capability index like $C_{pi}$ or $C_{pm}$. Consequently, no confusion is created by using MSE as an internal measure of control and simultaneously reporting $C_{pk}$ to those who prefer it. MSE can be estimated from the same data that are already available and are used to calculate $C_{pk}$.

## MSER, A NORMALIZED FORM OF MSE

### Practical Difficulties with MSE

In application, MSE suffers from three problems, all of which can be avoided by using MSER, a normalized form of the MSE. These disadvantages are:

- Because MSE depends upon unit of measure, it cannot be averaged meaningfully over a mixture of processes, such as all the processes within a given plant, and there usually is no point in comparing the MSE of one process with the MSE of a different process.
- The utility of MSEs is reduced by the fact that they are often awkward numbers like $5.734 \times 10^{-6}$ and have no apparent significance except in comparison with previous values for the same process.
- It is not obvious how to set goals for MSE or how to use MSE to compare the quality of a process with the six-sigma criteria that have gained such prominence over the last few years.

### Definition of MSER

MSER is a normalized form of MSE which circumvents the three problems just listed. MSER is calculated by dividing MSE by the square of the difference between the target and the nearer specification limit. Specifically,

$$\mathrm{MSER} = \frac{\mathrm{MSE}}{(\mathrm{NSL} - \mathrm{T})^2} \tag{2}$$

where

$$\mathrm{T} = \text{target value},$$

$$\mathrm{NSL} = \text{specification limit nearer to T}.$$

The target T is the intended value for the variable of interest and is sometimes called aim, goal or nominal. Dividing the mean square error by $(\mathrm{NSL} - \mathrm{T})^2$ normalizes it relative to the specification range and turns it into a dimensionless number that is independent of the unit of measurement applied. Division by the square of $(\mathrm{NSL} - \mathrm{T})$ is required because the MSE is itself the average of squared quantities.

### Significance of MSER

For a perfect process that produces all product exactly at the target value, both MSE and MSER equal zero. If the MSER value of a process is 1.0, it is equivalent in quality to a process that yields product just at the nearer specification limit NSL. A process that makes all product out of specification will have an MSER greater than 1.0.

There are two criteria for six-sigma quality:

1. The standard deviation of the process must not exceed 1/12 of the difference between specification limits.
2. The mean of the process can be no more than 1.5 standard deviations from the target.These criteria are based on an assumption that the target value is located midway between the specification limits. A process that meets both of these criteria will have an MSER of 0.09028 or less, as explained in the Appendix, so that a worthwhile goal is to keep all MSERs under 0.09028.

The MSER from a given process can be compared with the two checkpoints, which were just given, 0.09028 and 1.0; and it can be compared with the MSER of a different process. The MSERs from different processes can be averaged, as when it is desired to evaluate the overall quality for an entire plant by averaging the MSERs of all processes run in the plant.

## COMPONENTS OF MSE

When MSE calculations are based upon the subgrouped data that are usually produced by SPC, the MSE can be decomposed into three components:

$M_S$, the short-term component of MSE, estimates the variance within subgroups and is a measure of the variation within subgroups or over short periods of time.

$M_L$, the long-term component of MSE, estimates the component of variance between subgroup averages and is a measure of the variation over long periods of time.

$M_B$, the bias component of MSE, estimates the quantity $(\mu - T)^2$, where μ is the true mean of the process.

$M_B$ estimates the square of the bias of the process and is a measure of how badly the process is centered relative to the target.

### Calculating MSE and MSER Components

Calculate the subgroup variance $V_c$, the variance of averages $V_a$, and the grand average $\overline{\overline{X}}$. [1] Then the three components of MSE are

$$M_S = V_c,$$

$$M_L = V_a - \frac{V_c}{n},$$

$$M_B = (\overline{\overline{X}} - T)^2 - \frac{V_a}{k}$$

where

$n$ = number of items per subgroup,

$k$ = number of subgroups,

$T$ = target value for the variable.

The MSE is the sum of these three components so that

$$MSE = M_S + M_L + M_B.$$

If process subgroups each consist of just one item, then the subgroup average is equal to the single observed value and $M_S$, which equals $V_c$, must be considered equal to zero. The calculated long-term component then reflects both the short-term and long-term components.

It is an identity that

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(x_i - T)^2 \equiv M_S + M_L + M_R$$

with

$$m = kn,$$

so the sum of the components always equals the MSE calculated directly from the individual measurements.

Since the calculations yield only estimates of the true components, negative values of $M_L$ or $M_B$ may be encountered, particularly when only a small number of subgroups are included in the calculations. Because a negative component is impossible, 0 is always a better estimate than any negative number. It is preferable to report 0 rather than a negative number in all reports of MSE, not only because zero is a better estimate but also because it avoids confusing the reader. However, if components are to be averaged with those from other MSE calculations, as discussed below under *Period Included in Calculations*, the negative values should be retained and not be adjusted to zero before the average is computed.

According to equation (2) each component of MSER equals the corresponding component of MSE divided by $(NSL - T)^2$. These components are

$MR_S$ = short-term component of MSER = $M_S/(NSL - T)^2$

$MR_L$ = long-term component of MSER = $M_L/(NSL - T)^2$

$MR_B$ = bias-squared component of MSER = $M_B/(NSL - T)^2$

The meaning of these components is easier to understand and remember when they are expressed as a percentage of total MSE or MSER. The percentage components of MSE are

$$M_{S\%} = 100 \cdot M_S/MSE$$

$$M_{L\%} = 100 \cdot M_L/MSE$$

$$M_{B\%} = 100 \cdot M_B/MSE,$$

and those of MSER are

$$MR_{S\%} = 100 \cdot MR_S/MSER$$

$$MR_{L\%} = 100 \cdot MR_L/MSER$$

$$MR_{B\%} = 100 \cdot MR_B/MSER.$$

Since the percentage components of MSER are equal to the percentage components of MSE, it is unnecessary to calculate both percentages.

### Calculation from Subgroup Averages and Ranges

In some cases, the data may be available only as subgroup averages and ranges, i.e., $\overline{X}$ and R data rather than as individual measurements. In such cases, if the underlying populations are normally distributed, the three components of sample mean square error can be approximated as

$$M_s \approx C\overline{R}^2$$

$$M_L \approx V_a - B\overline{R}^2$$

$$M_B \approx (\overline{\overline{X}} - T)^2 - \frac{V_a}{k}.$$

In these approximations $\overline{R}$ denotes the average of the k subgroup ranges, the sample grand average $\overline{\overline{X}}$ is the average of the k subgroup averages, and B and C denote constants whose values are listed in Table 1.

**Table 1.** Constants B and C used in approximate formulas for $M_S$, $M_L$, and $M_B$.

| Subgroup Size n | B | C |
|---|---|---|
| 2 | 0.393 | 0.786 |
| 3 | 0.116 | 0.349 |
| 4 | 0.059 | 0.236 |
| 5 | 0.037 | 0.185 |
| 6 | 0.026 | 0.156 |
| 7 | 0.0195 | 0.137 |
| 8 | 0.0154 | 0.123 |
| 9 | 0.0126 | 0.113 |
| 10 | 0.0106 | 0.106 |
| 12 | 0.00785 | 0.0942 |
| 15 | 0.00553 | 0.0830 |
| 20 | 0.00358 | 0.0717 |
| 25 | 0.00259 | 0.0647 |

### Period Included in Calculations and Reports

Imagine a process that is typically run for one day at a time, producing 18 subgroups, and then not run for some time. Suppose that the usual procedure is to allow the process to run far off target for the entire day, even though it would be easy to correct it. Further, suppose that the high days are essentially cancelled out by the low days over a period of a year, so that the grand average for all the days in a year tends to be rather close to target. An MSE calculated for the entire year would show a large long-term component but very little bias, even though bias is a major and easily correctable problem.

Consequently, it is preferable to calculate MSEs for individual runs, or for moderate periods like one day if the runs are quite long, and then to average the results, weighted by degrees of freedom for $V_a$, provided that not too many degrees of freedom are lost by breaking the data into smaller groupings. For these purposes, a run ends when there is a physical break in operation, as when the equipment is shifted to making another product or when it is shut down for the weekend.

An example of a weighted average calculation is given in Table 2. The averages in the table were calculated by dividing each total by the sum of all the degrees of freedom, which happens to be 30 for this example.

**Table 2.** Example of a weighted average calculation. k = number of subgroups for each of the MSEs to be averaged, $\Phi$ = number of degrees of freedom for $V_a$, $\Phi = k - 1$.

| | k | $\Phi$ | $M_S$ | $M_L$ | $M_B$ | $\Phi \cdot M_S$ | $\Phi \cdot M_L$ | $\Phi \cdot M_B$ |
|---|---|---|---|---|---|---|---|---|
| | 11.00 | 10.00 | 2.00 | 1.00 | 1.00 | 20.00 | 10.00 | 10.00 |
| | 21.00 | 20.00 | 3.00 | 2.00 | 1.00 | 60.00 | 40.00 | 20.00 |
| TOTAL | | 30.00 | | | | 80.00 | 50.00 | 30.00 |
| AVERAGE | | | | | | 2.67 | 1.67 | 1.00 |

Whenever a period is broken into two smaller periods, one degree of freedom is lost for the long-term component, $V_a$. Therefore, breaking long runs into periods of 3 subgroups per period would result in sacrificing approximately 33% of the degrees of freedom for $V_a$, but breaking long runs into periods of 11 subgroups per period costs only approximately 9% of the degrees of freedom for $V_a$.

Although it is acceptable to calculate MSE for runs with very few subgroups when they are to be averaged with results from other runs, MSE should never be reported if fewer than 10 degrees of freedom for $V_a$ are represented. The potential sampling error of variances based on less than 10 degrees of freedom is just too great. However, MSERs representing fewer degrees of freedom may be included in averages with the MSERs of other processes, as for a plant-wide average MSER.

### Data To Be Included in Calculations

Unlike $C_p$ and $C_{pk}$, which are sometimes considered to be indicators of how well a process can be run under ideal conditions, MSE is an indicator of how good the control and quality actually were. Consequently, the data included in calculations must be representative.

Where computerized data acquisition and reporting systems are in use, it is both desirable and practical to include all subgroups. If manual acquisition and computation are required, select a subset of the data on a basis which will not bias the results. However, do *not* select random subgroups. All subgroups included in one MSE calculation must be consecutive. A simple and satisfactory procedure to follow is

1. select the last 18 to 24 subgroups in the reporting period, which would usually be a month or a quarter,
2. break the subgroups into sets of data each containing only subgroups from one consecutive run,
3. calculate the components of MSE for each run,
4. average the corresponding components for the runs, weighted by degrees of freedom for $V_a$.

## ASSESSING STATISTICAL CONTROL

The long-term component of MSE for a process that truly operates in statistical control is zero. Since one must work with estimates of the actual components, however, calculated values of $M_L$ which are greater than zero will be observed for many processes which actually are in statistical control. To determine which processes are not in statistical control,

1. scan MSE reports for values of $M_L$ which are greater than zero,
2. for such processes, compute the ratio $V_a/(V_c/n)$,
3. compare $V_a/(V_c/n)$ with a table of F ratios to assess the null hypothesis that $V_a$ and $V_c/n$ are both estimates of the same population variance, in which case the process is in statistical control,
4. if $V_a/(V_c/n)$ is too large, the null hypothesis must be rejected and the process cannot be considered to be in statistical control.

Figure 1 gives a chart suitable for quickly testing statistical control. If the process really is in statistical control, there is a probability of only 0.05 of observing a value of $V_a/(V_c/n)$ above the line corresponding to the number of degrees of freedom for $V_a$, which is the numerator in the F ratio. Any process that produces an F ratio above the line cannot safely be considered to be in statistical control, although there is some possibility that it may be. Refer to a table of F ratios to learn just how improbable the observed value of $V_a/(V_c/n)$ is, assuming that the process is in statistical control.

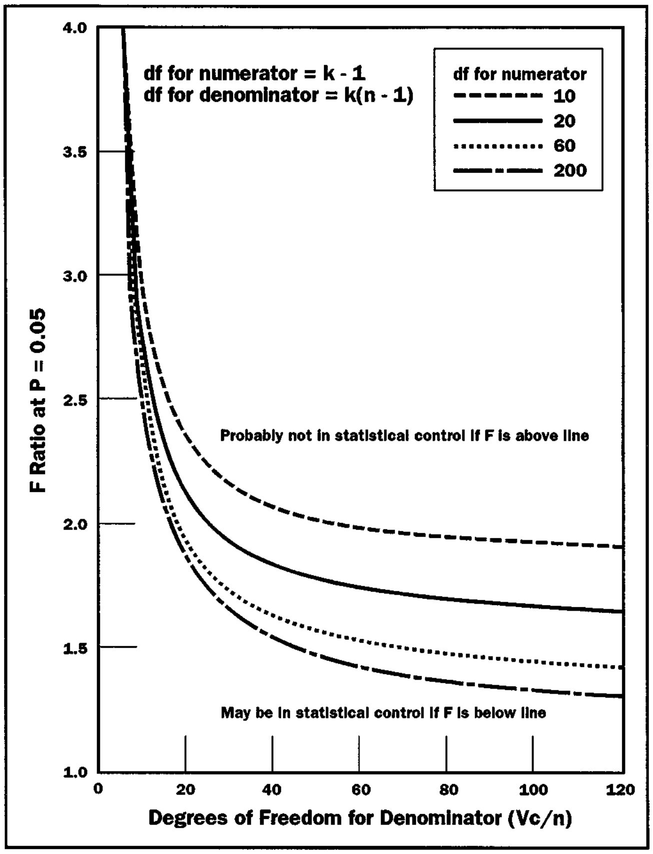


**Figure 1.** F-ratios for testing statistical control. Form test F ratio as $V_a/(V_c/n)$.

If the F ratio falls below the line, the data provide no firm basis for concluding that the process is not in statistical control. Such an F ratio does not prove that the process is in statistical control, but it is reasonable to assume that it is close to statistical control.

## REPORTING AND USING MSE/MSER

MSE and MSER are tools for managing processes, not for controlling them on an hour-by-hour basis. With the highly automated processes common today, managerial action will often be required to achieve significant process improvement whether the action is effected by an empowered operator, a team, or a plant manager. Managerial action includes such steps as changing maintenance schedules, installing new controls or equipment, choosing a new supplier, altering formulations, scheduling to minimize upsets, setting MSE goals, improving training, changing factory staffing, designing products for manufacturability, and revising procedures for control and inspection.

For MSE and MSER to have beneficial impact, they must be reported in a timely and effective fashion and should be used actively by those responsible for process improvement. Several suggestions can be made for effectively employing MSE and MSER.

- Use MSER as the internal measure of control and quality instead of $C_{pk}$, which should still be calculated because of its appeal to many customers. As long as $C_{pk}$ is the official measure of quality within an organization, no one will care about or use any other criterion.
- Within each organization, designate someone to drive MSE reduction.
- Set goals for MSER, ultimately working toward having all processes below the six-sigma MSER of 0.09028.
- Present, review and discuss MSER on a top-down basis, looking first at the plant-wide average, then at detail to explain problems. Reviews can be monthly or quarterly, or possibly even less frequent, depending upon circumstances. The important objective is to take constructive action, not to generate the numbers frequently. Presentations should be in graphical format whenever practical.
- When setting priorities for process and quality improvement, a heavy weighting should be given to MSER.
- Use the components of MSE to help achieve the maximum improvement in MSE with a minimum of effort and cost. Many processes will have a large bias component, in which case it may be possible to achieve a very substantial improvement in MSE by paying more attention to staying on target. If the long-term component is high it is possible that either overcontrol or undercontrol is occurring; a control chart that jumps around may indicate overcontrol, a control chart that drifts may indicate under-control.[a] If the short-term component is large, useful improvement may be attainable by looking into measurement equipment and procedures.
- Use MSE to evaluate product design, equipment, materials and procedures. If MSE decreases after a change in such a factor, the change probably has improved both control and quality.MSE is recommended as an internal tool. Although MSE, or MSD as it is called in much SPC literature, is not a new concept, many customers today appear to be more concerned with $C_{pk}$. By using MSE to monitor and motivate internal process improvement while still reporting $C_{pk}$ to those who request it, one can achieve both continuous improvement and customer satisfaction.


## ACKNOWLEDGEMENT

Many of the concepts in this paper were developed in conjunction with Dr. James M. Maynard, AMP Incorporated, Global Engineering Practices.


[a] If the capability to calculate autocorrelation functions is available, positive autocorrelation is a good indicator of drift and undercontrol.

## REFERENCE


1. G.J. Battaglia and J.M. Maynard, ‘‘Mean Square Error: A Useful Tool for Statistical Process Management,’’ AMP J. Technol. **2**, 47-55 (1992).


## APPENDIX
## MSER FOR SIX-SIGMA QUALITY

In processes with exactly six-sigma quality, as defined by Motorola and adopted by many other companies, the upper and lower specification limits are separated by 12 standard deviations, $\sigma$, and the average value of all production is 1.5 $\sigma$ from target. Target is taken to be half-way between upper and lower specification limits.

The MSE of a given quantity of production can be calculated according to equation (1). It is true for the entire population of a variable x that

$$\mathrm{MSE} = \sigma^2 + (\mu - \mathrm{T})^2$$

where

$$\sigma^2 = \text{standard deviation of x},$$

$$\mu = \text{mean of the entire population.}$$

When the mean deviates from target by exactly the amount allowable for six sigma quality, $(\mu - \mathrm{T})$ equals $1.5 \cdot \sigma$ so

$$\mathrm{MSE}_{6\sigma} = \sigma^2 + (1.5\ \sigma)^2 = 3.25\ \sigma^2.$$

When the standard deviation, $\sigma$, has the value which is just consistent with six-sigma quality, it is equal to 1/12 of the difference between the upper and lower specification limits, or to 1/6 of the difference between either specification limit and the target. Consequently,

$$\sigma = \frac{\mathrm{U} - \mathrm{T}}{6}$$

where

$$\mathrm{U} = \text{upper specification limit.}$$

Since the definition of six-sigma quality assumes that the specification limits are equidistant from the target, the lower specification limit could also be used, in which case $\sigma$ would equal $(\mathrm{T} - \mathrm{L})/6$, where L is the lower specification limit.

When a process is just at the conditions imposed for six-sigma quality, its MSE is

$$\mathrm{MSE}_{6\sigma} = 3.25\,\frac{(\mathrm{U} - \mathrm{T})^2}{36} = 0.0902777(\mathrm{U} - \mathrm{T})^2.$$

When the MSE equals $\mathrm{MSE}_{6\sigma}$, then the MSER becomes

$$\mathrm{MSER}_{6\sigma} = \frac{\mathrm{MSE}_{6\sigma}}{(\mathrm{U} - \mathrm{T})^2} = 0.0902777 \approx 0.09028.$$


**Glenn J. Battaglia** is a Project Manager in the Process Control Technology Department at AMP Incorporated in Harrisburg, Pennsylvania.

Mr. Battaglia holds a Bachelor of Science degree in chemical engineering from the Massachusetts Institute of Technology and a Master of Business Administration degree from Harvard University. His industrial career, most of which has been spent at the Polaroid Corporation and at AMP Incorporated, includes work in research, manufacturing and computer systems. Since joining AMP, his primary activity has been the application of statistics to control of processes.