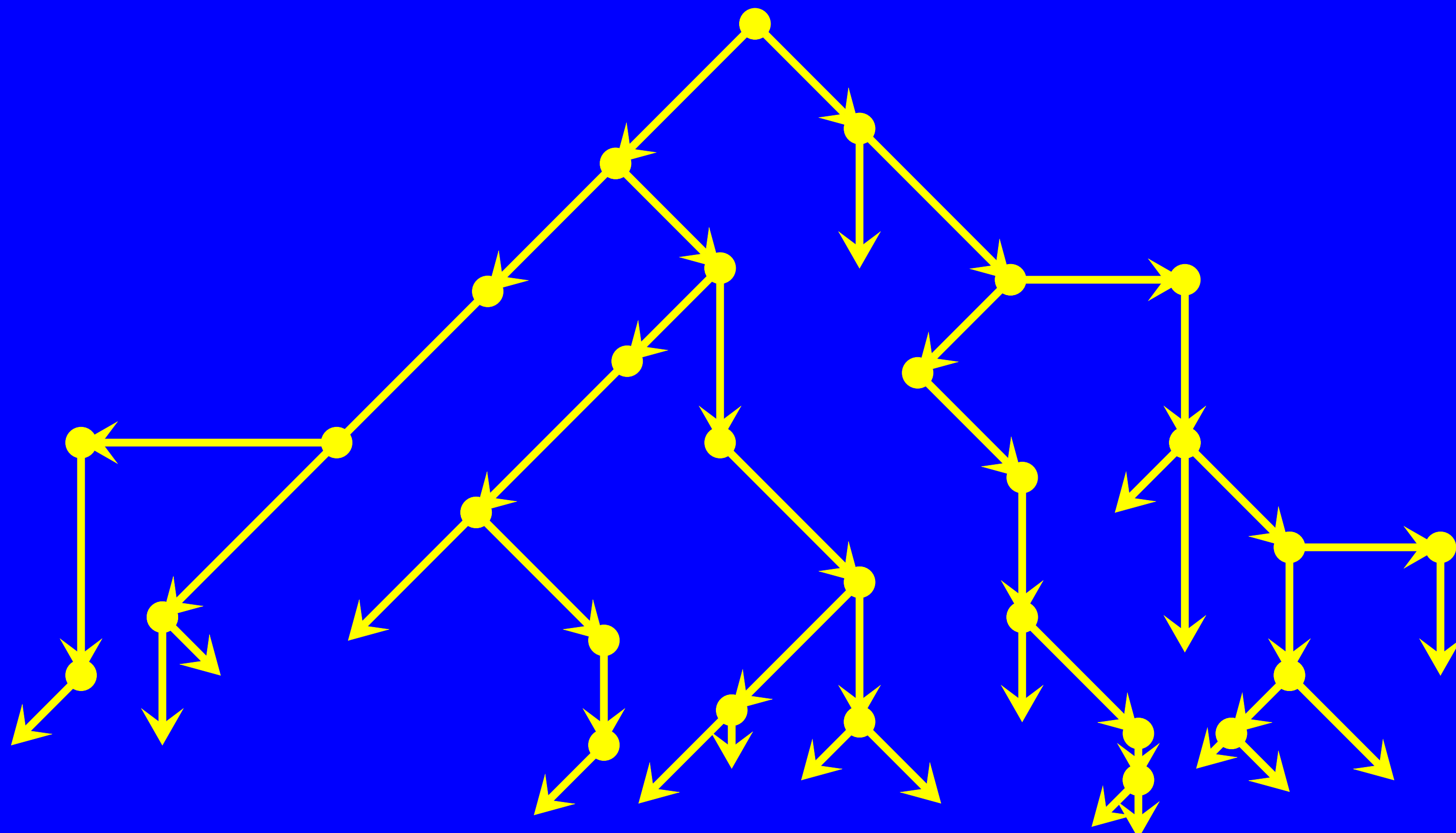
DONALD J. WHEELER

Gauge R&R Studies & Four Classes of Process Monitors

based on Chapters Fifteen and Sixteen of the book
EMP III USING IMPERFECT DATA

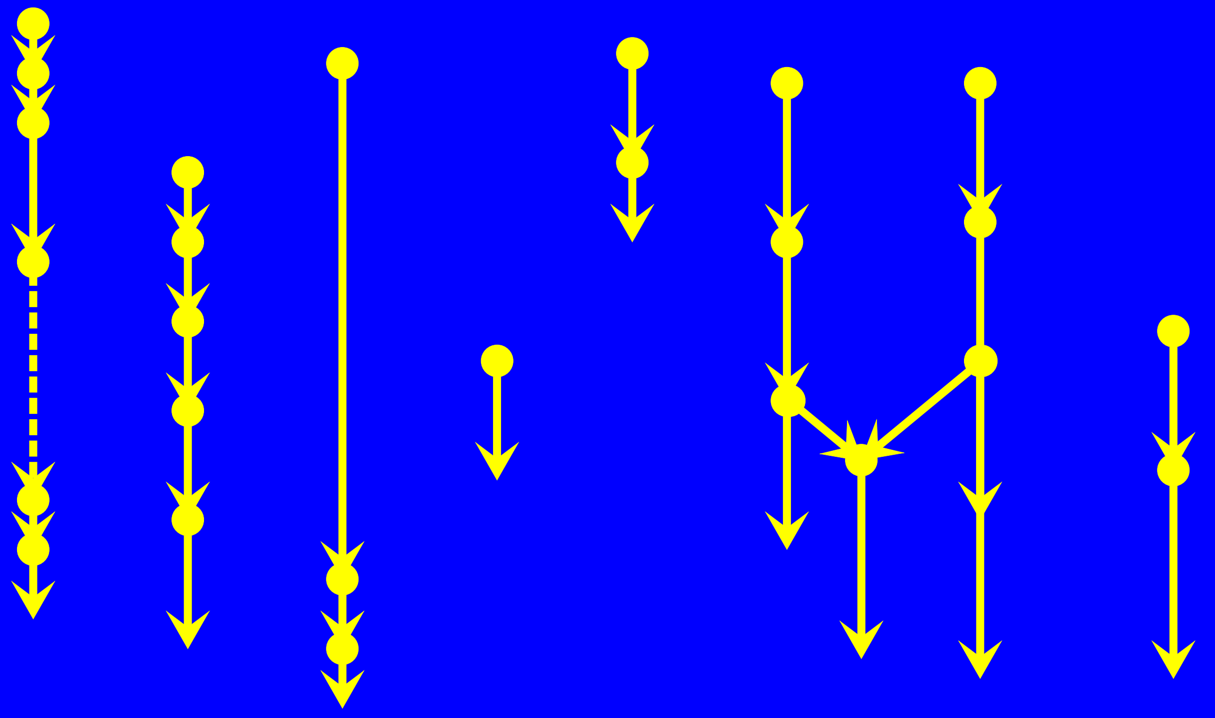
R&R Studies & Four Classes of Monitors

Many ideas develop coherently
from a single, seminal concept...



R&R Studies & Four Classes of Monitors

Other ideas develop incoherently from diverse origins with little cross-fertilization between streams...



Measurement System Analysis belongs to this category, with solutions ranging from naive to theoretical, from simple to complex, and from right to wrong.

The Gauge R&R Study

Typically a Gauge R&R Study will have two or more operators, one gauge, and up to ten parts.

Each operator will measure each part two or three times...

The Gasket Thickness Data

Oper.	A	A	A	A	A	B	B	B	B	B	C	C	C	C	C
Part	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1st	167	210	187	189	156	155	206	182	184	143	152	206	180	180	146
2nd	162	213	183	196	147	157	199	179	178	142	155	203	181	182	154
Aver.	164.5	211.5	185.0	192.5	151.5	156.0	202.5	180.5	181.0	142.5	153.5	204.5	180.5	181.0	150.0
Range	5	3	4	7	9	2	7	3	6	1	3	3	1	2	8

The Average Range is 4.2667.
The Operator Averages are 181.0, 172.5, and 173.9.
The Part Averages are
158.0, 206.167, 182.0, 184.833, and 148.0.

1. Check that all the range values fall below the Upper Range Limit.

2. Divide the Average Range by the appropriate bias correction factor of d_2 = 1.128 to obtain an estimate of the Repeatability or Equipment Variation:

$$EV = \frac{\text{Average Range}}{d_2} = \frac{4.2667}{1.128} = 3.783 \text{ mils}$$

The Gauge R&R Study — Step 3

3. Next the Range of the Operator Averages, R_o, is used to compute an estimate of the Reproducibility or Appraiser Variation:

$$AV = \left\{ \left[\frac{R_o}{d_2^*} \right]^2 - \frac{o}{n\,o\,p}\, EV^2 \right\}^{0.5}$$

$$= \left\{ \left[\frac{8.5}{1.906} \right]^2 - \frac{3}{30}\, 3.783^2 \right\}^{0.5} = 4.296$$

The Operator Averages of 181.0, 172.5, and 173.9 have a range of R_o = 8.5 mils, and d_2^* for one group of size three is 1.906.

4. Next the Combined Repeatability and Reproducibility is estimated by squaring the Repeatability, adding the square of the Reproducibility, and finding the square root:

$$GRR = \{ EV^2 + AV^2 \}^{0.5}$$

$$= \{ 3.783^2 + 4.296^2 \}^{0.5}$$

$$= 5.724 \text{ mils}$$

5. Next the Product Variation is estimated using the Range of the Part Averages, R_p

$$PV = \frac{R_p}{d_2^*} = \frac{58.167}{2.477} = 23.483 \text{ mils}$$

The Part Averages of
158.0, 206.167, 182.0, 184.833, and 148.0
have a Range of R_p = 58.167,
and d_2^* for one group of size five is 2.477.

6. Finally the Total Variation is estimated by combining the square of the Equipment Variation, the square of the Appraiser Variation, and the square of the Product Variation, and taking the square root:

$$TV = \{ EV^2 + AV^2 + PV^2 \}^{0.5}$$

$$= \{ 3.783^2 + 4.296^2 + 23.483^2 \}^{0.5}$$

$$= 24.171 \text{ mils}$$

Up to this point things are okay. While these estimates are not the only estimates we could have found, and while they may not be the "best" estimates possible, they are all reasonable estimates of these various quantities.

EV = 3.783 mils
AV = 4.296 mils
GRR = 5.724 mils
PV = 23.483 mils
TV = 24.171 mils

The train wreck begins when the Gauge R&R Study tries to use these estimates to characterize relative utility.

In the current version of the Gauge R&R Study the first four quantities in the list above are expressed as a percentage of the last value.

7. The Repeatability is divided by the Total Variation:

$$\%EV = 100 \frac{EV}{TV} = 100 \frac{3.783}{24.171} = 15.65\%$$

This number is interpreted to mean that the Repeatability consumes 15.7% of the Total Variation.

EV = 3.783 mils
AV = 4.296 mils
GRR = 5.724 mils
PV = 23.483 mils
TV = 24.171 mils

8. The Reproducibility is divided by the Total Variation:

$$\%AV = 100\frac{AV}{TV} = 100\frac{4.296}{24.171} = 17.77\%$$

This number is interpreted to mean that the Reproducibility consumes 17.8% of the Total Variation.

EV = 3.783 mils
AV = 4.296 mils
GRR = 5.724 mils
PV = 23.483 mils
TV = 24.171 mils

9. The Combined R&R is divided by the Total Variation:

$$\%GRR = 100\,\frac{GRR}{TV} = 100\,\frac{5.724}{24.171} = 23.68\%$$

This number is interpreted to mean that the Combined R&R consumes 23.7% of the Total Variation.

EV = 3.783 mils
AV = 4.296 mils
GRR = 5.724 mils
PV = 23.483 mils
TV = 24.171 mils

10. The Product Variation is divided by the Total Variation:

$$\%PV = 100 \frac{PV}{TV} = 100 \frac{23.483}{24.171} = 97.15\%$$

This number is interpreted to mean
that the Product Variation consumes
97.2% of the Total Variation.

EV = 3.783 mils
AV = 4.296 mils
GRR = 5.724 mils
PV = 23.483 mils
TV = 24.171 mils

But since when does 15.7 plus 17.8 equal 23.7 ?

Likewise, when does 23.7 plus 97.2 equal 100 percent ?

Realizing that they had a problem,
and not knowing what else to do about it,
the authors of the Gauge R&R Study decided
to insert a statement at this point...

"The sum of the percent consumed by each factor
will not equal 100 percent."

"The sum of the percent consumed by each factor will not equal 100 percent."

No explanation is given for this statement.

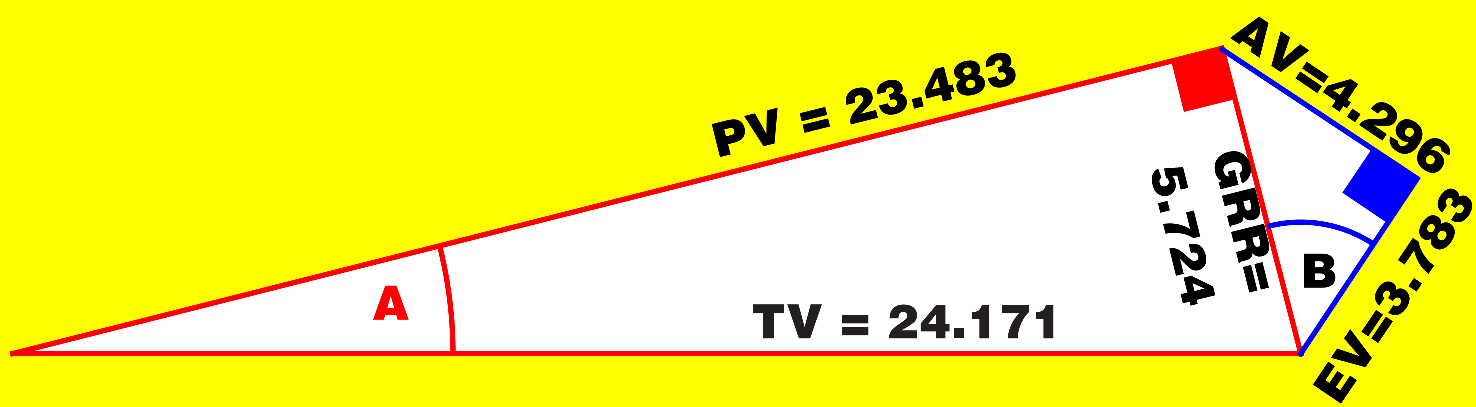
No guide is offered for how to proceed
now that common sense
and every rule in arithmetic have been violated.

Just a simple statement that these numbers
do not mean what they were just interpreted to mean,
and the users are left to their own devices.

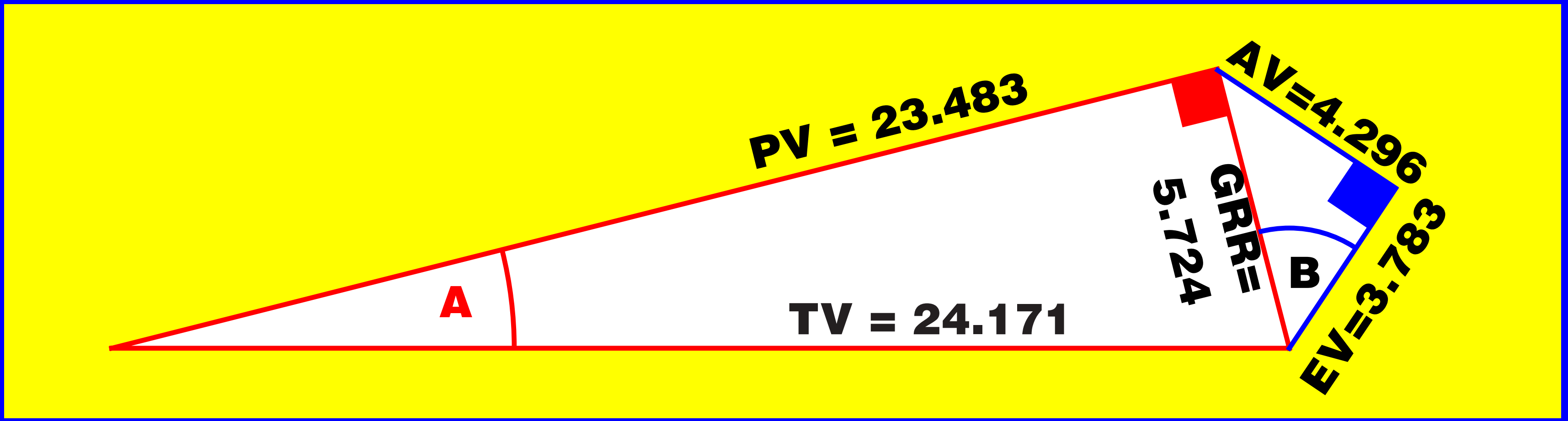
Why the "Percentages" Do Not Add Up

Obviously, the ratios computed in Steps 7 thru 10
are NOT percentages, so what are they?

Considering how the basic quantities
were computed in Steps 2 thru 6
we can show the relationships
between these basic quantities
using a couple of right triangles.



Why the "Percentages" Do Not Add Up

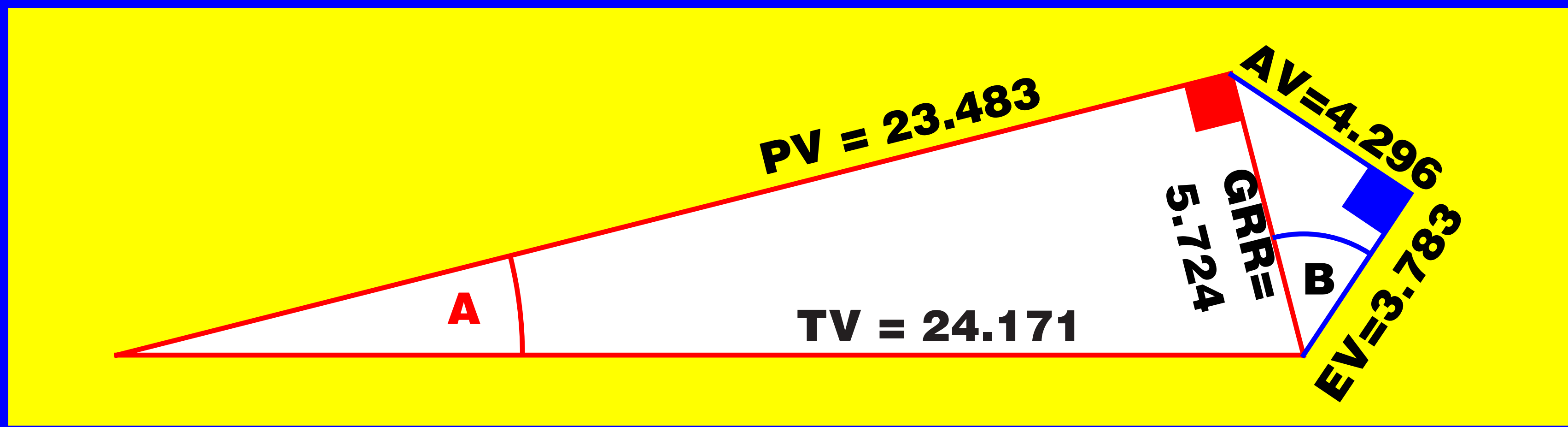


$$\frac{\%EV}{100} = \frac{5.724}{24.171}\,\frac{3.783}{5.724} = (\text{Sine } A)(\text{Cosine } B) = 0.1565$$

$$\frac{\%AV}{100} = \frac{5.724}{24.171}\,\frac{4.296}{5.724} = (\text{Sine } A)(\text{Sine } B) = 0.1777$$

$$\frac{\%GRR}{100} = \frac{5.724}{24.171} = (\text{Sine } A) = 0.2368$$

Why the "Percentages" Do Not Add Up



$$\frac{\%PV}{100} = \frac{23.483}{24.171} = (\text{Cosine } A) = 0.9715$$

While these ratios were interpreted as proportions,
they are clearly trigonometric functions,
and that is why the ratios do not add up.

Why the "Percentages" Do Not Add Up

A set of ratios will be proportions if and only if the common denominator is the sum of the numerators.

$$\frac{a}{a+b+c} + \frac{b}{a+b+c} + \frac{c}{a+b+c} = 1$$

It is the additivity of the numerators that is the essence of proportions.

And in an R&R study
it is *not* the standard deviations,
but rather the *variances* that are additive.

$$TV^2 = EV^2 + AV^2 + PV^2$$

An Honest Gauge R&R Study

Using the relationship between the variances seen in Step 6:

$$TV^2 = EV^2 + AV^2 + PV^2$$

and dividing both sides by the total variance:

$$\frac{TV^2}{TV^2} = \frac{EV^2}{TV^2} + \frac{AV^2}{TV^2} + \frac{PV^2}{TV^2} = 1$$

we discover the true proportions inherent in this problem.

Therefore, the Repeatability actually consumes

$$100\,\frac{EV^2}{TV^2} = 100\,\frac{3.783^2}{24.171^2} = 2.4\%$$

2.4% of the Total Variation,

rather than the 15.7% erroneously found earlier.

EV = 3.783 mils
AV = 4.296 mils
GRR = 5.724 mils
PV = 23.483 mils
TV = 24.171 mils

The Reproducibility actually consumes

$$100\,\frac{AV^2}{TV^2} = 100\,\frac{4.296^2}{24.171^2} = 3.2\%$$

3.2% of the Total Variation,

rather than the 17.8% erroneously found earlier.

$$\begin{array}{r} 2.4 \\ +\underline{3.2} \\ 5.6 \end{array}$$

EV = 3.783 mils
AV = 4.296 mils
GRR = 5.724 mils
PV = 23.483 mils
TV = 24.171 mils

So that the Combined R&R actually consumes

$$100 \frac{GRR^2}{TV^2} = 100 \frac{5.724^2}{24.171^2} = 5.6\%$$

5.6% of the Total Variation,
rather than the 23.7% erroneously found earlier,
and these proportions actually add up,
as proportions should.

EV = 3.783 mils
AV = 4.296 mils
GRR = 5.724 mils
PV = 23.483 mils
TV = 24.171 mils

wrong:

$$\begin{array}{r} 15.7 \\ +\ 17.8 \\ \hline 23.7 \end{array}$$

right:

$$\begin{array}{r} 2.4 \\ +\ 3.2 \\ \hline 5.6 \end{array}$$

$$\begin{array}{r} 100.0 \\ -\ 5.6 \\ \hline 94.4 \end{array}$$

And the Product Variation actually consumes

$$100 \frac{PV^2}{TV^2} = 100 \frac{23.483^2}{24.171^2} = 94.4\%$$

94.4% of the Total Variation,
rather than the 97.2% erroneously found earlier.

Now we have properly accounted for the components of the total variation.

EV = 3.783 mils
AV = 4.296 mils
GRR = 5.724 mils
PV = 23.483 mils
TV = 24.171 mils

2.4%	Repeatability
+3.2%	Reproducibility
+94.4%	Product Variation
100.0%	Total Variation

What Can You Learn From a Gauge R&R Study?

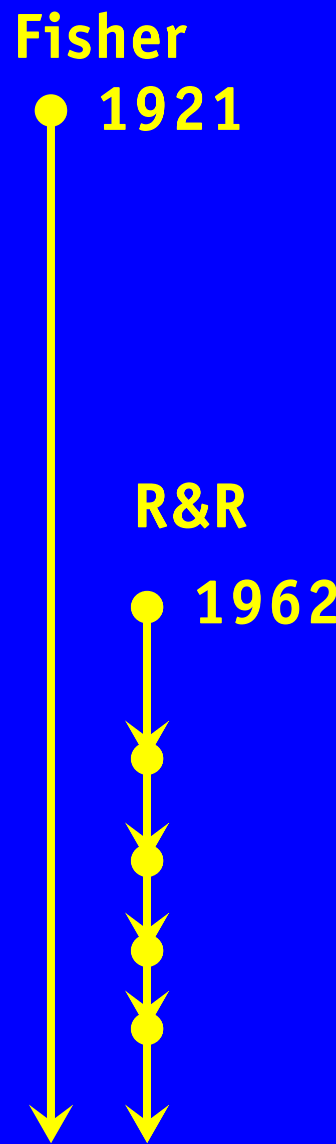
	Traditonal Gauge R&R Values	Honest Gauge R&R Values
Repeatablity	15.7%	2.4%
Reproducibility	17.8%	3.2%
Combined R&R	23.7%	5.6%
Product Variation	97.2%	94.4%

What Can You Learn From a Gauge R&R Study?

	Gauge R&R		The Truth
Repeatablity	15.7%	WRONG	2.4%
Reproducibility	17.8%	WRONG	3.2%
Combined R&R	23.7%	WRONG	5.6%
Product Variation	97.2%	WRONG	94.4%

By ignoring the Pythagorean Theorem
the Gauge R&R Study converts good data
into values that are hopelessly flawed,
resulting in an analysis where
virtually nothing is true, correct, or useful.

The Intraclass Correlation



In 1921 Sir Ronald Fisher introduced
a theoretically sound and easy to understand
way of characterizing
the relative utility
of a measurement system
for a particular application.

This is the Intraclass Correlation, ρ
which may be estimated
using the value from Step H10:

$$\frac{PV^2}{TV^2} = \text{Est. Intraclass Correlation}$$

$$= 0.944$$

The Intraclass Correlation

$$\frac{PV^2}{TV^2} = \text{Est. Intraclass Correlation} = 0.944$$

Clearly the Intraclass Correlation represents that proportion of the total variation that is attributable to variation in the product stream.

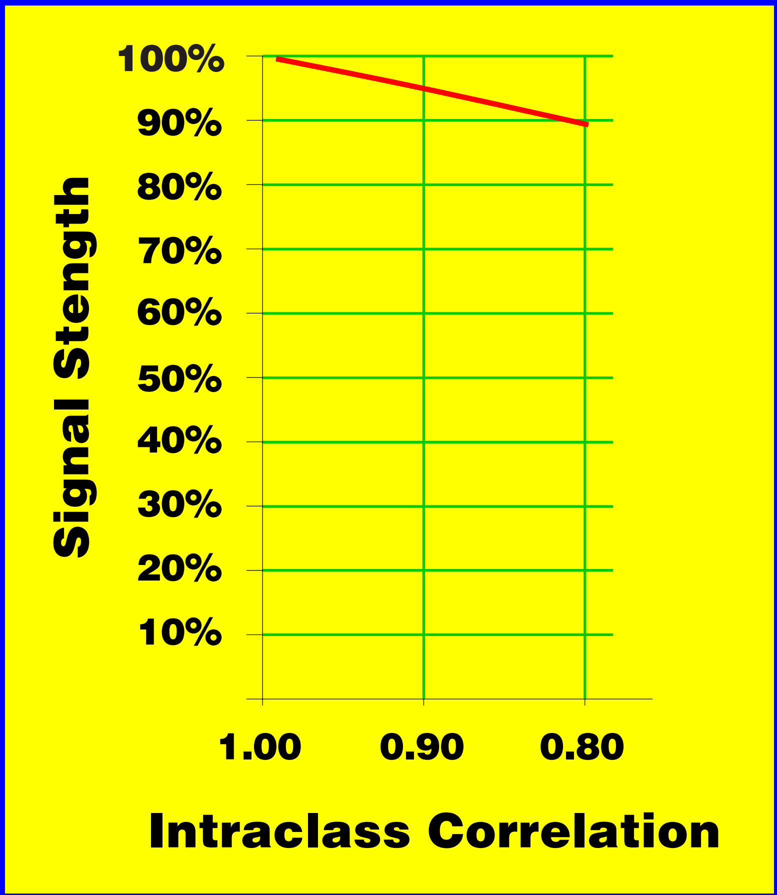
It also represents the correlation between two measurements of the same thing, hence the name of intraclass correlation.

This is the appropriate metric for characterizing the relative utility of a measurement system.

First Class Monitors

$IC > 80\%$

When the Intraclass Correlation exceeds 80%
the measurement system will provide a First Class Monitor.

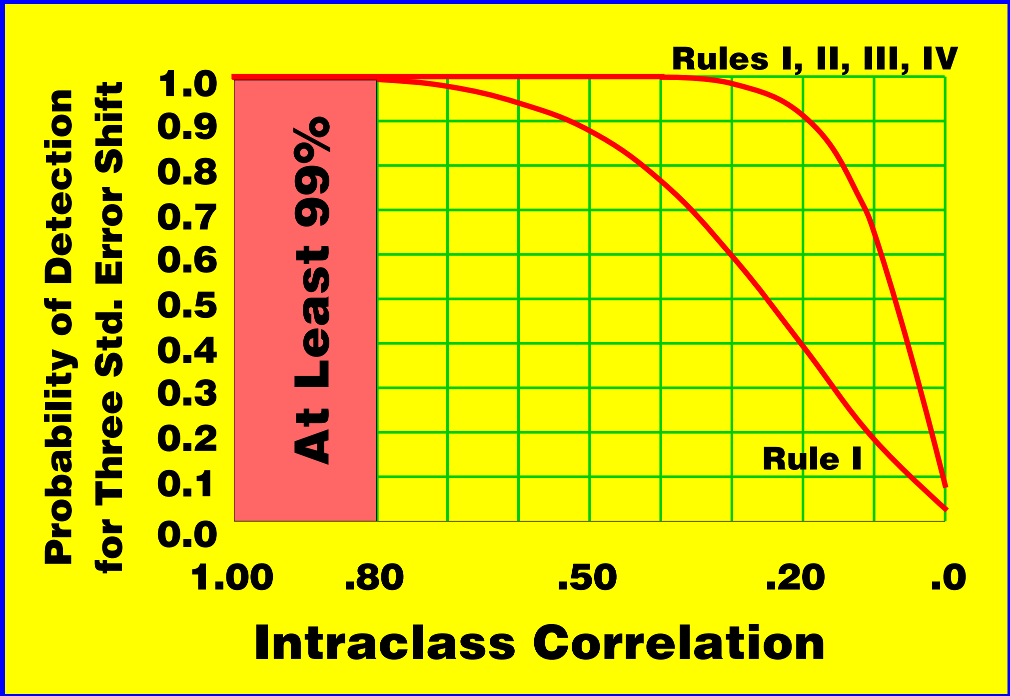


With First Class Monitors
signals coming from
the production process
will be attenuated
by less than 10 percent
due to the effects
of measurement error.

$$\frac{PV^2}{TV^2} = 0.944$$

First Class Monitors *IC > 80%*

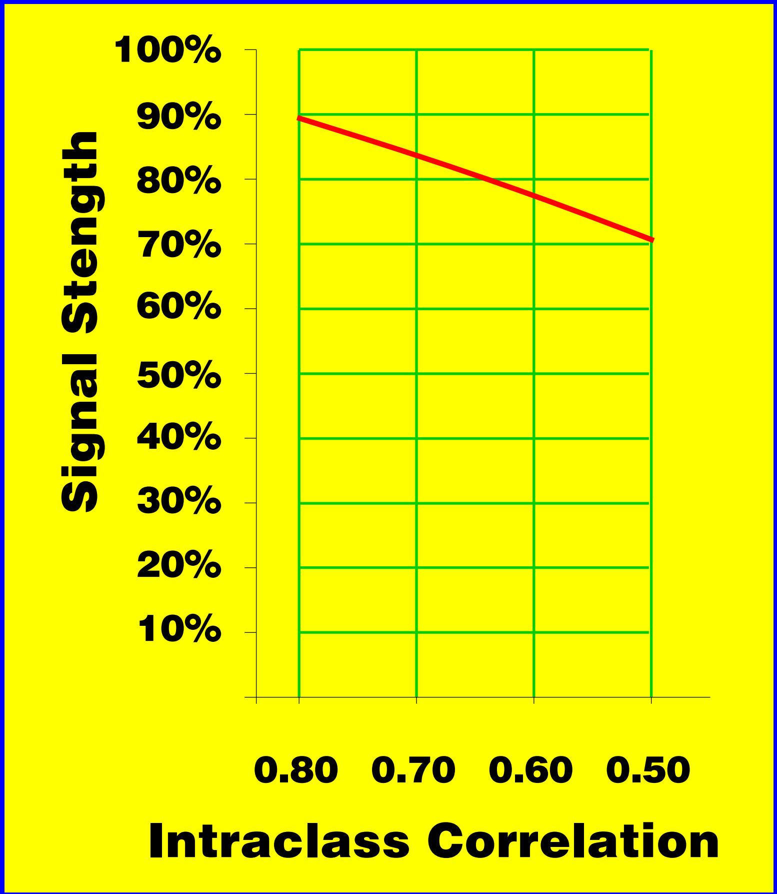
When placed on a Process Behavior Chart, First Class Monitors will have better than a 99% chance of detecting a three standard error shift in the production process using Detection Rule One.



$$\frac{PV^2}{TV^2} = \text{Est. Intraclass Correlation} = 0.944$$

Second Class Monitors 80% > IC > 50%

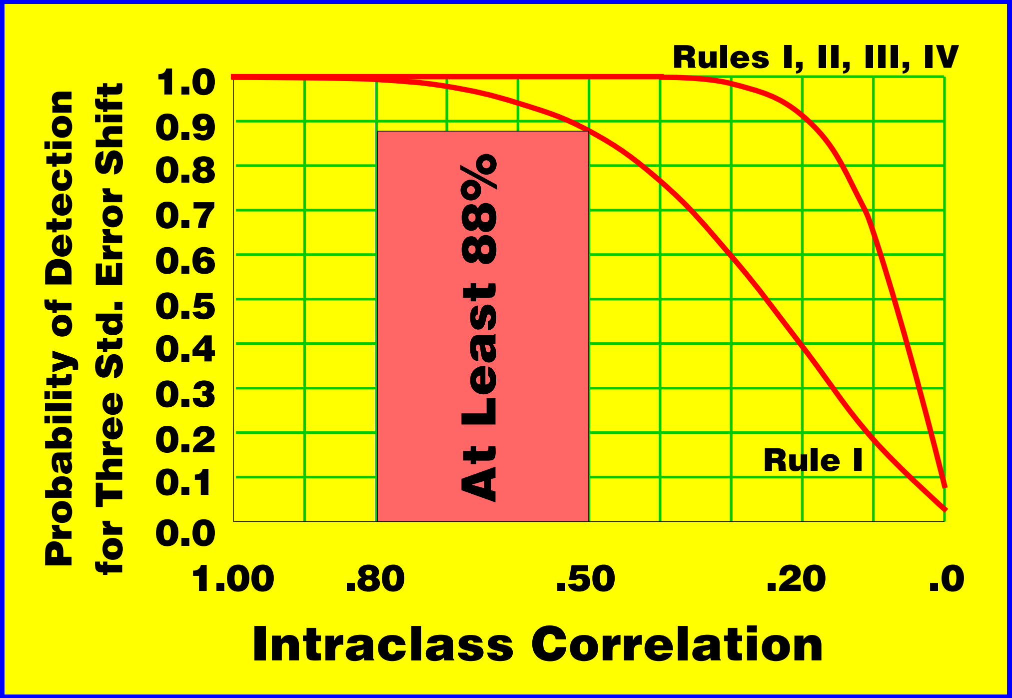
When the Intraclass Correlation is between 80% and 50% the measurement system will provide a Second Class Monitor.



With a Second Class Monitor any signals coming from the production process will be attenuated by 10 to 30 percent due to the effects of measurement error.

Second Class Monitors *80% > IC > 50%*

When used with a Process Behavior Chart, Second Class Monitors will still have better than an 88% chance of detecting a three standard error shift in the process using Detection Rule One alone.

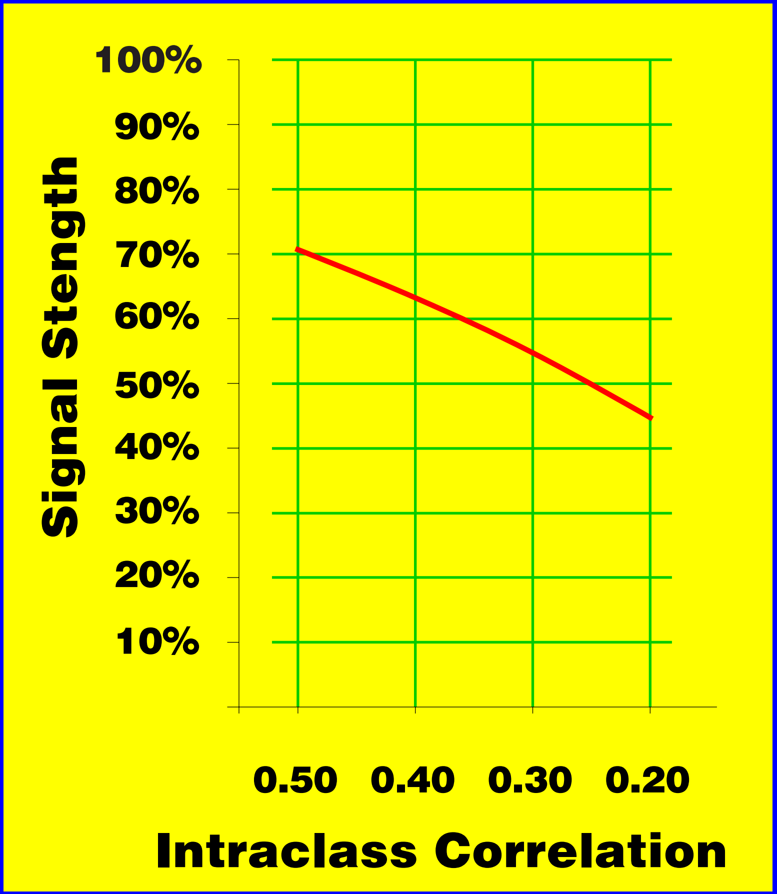


Moreover, they are virtually certain to detect a three standard error shift in the process using Detection Rules 1, 2, 3, and 4.

Third Class Monitors

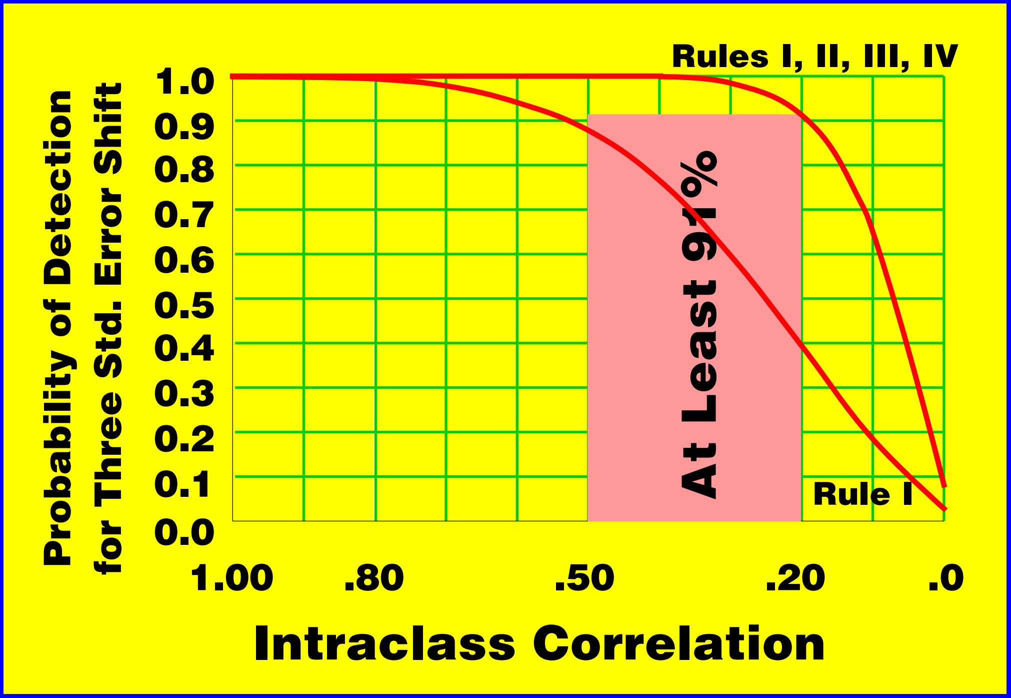
50% > IC > 20%

When the Intraclass Correlation is between 50% and 20% the measurement system will provide a Third Class Monitor.



With a Third Class Monitor any signals coming from the production process will be attenuated by 30 to 55 percent due to the effects of measurement error.

Third Class Monitors *50% > IC > 20%*

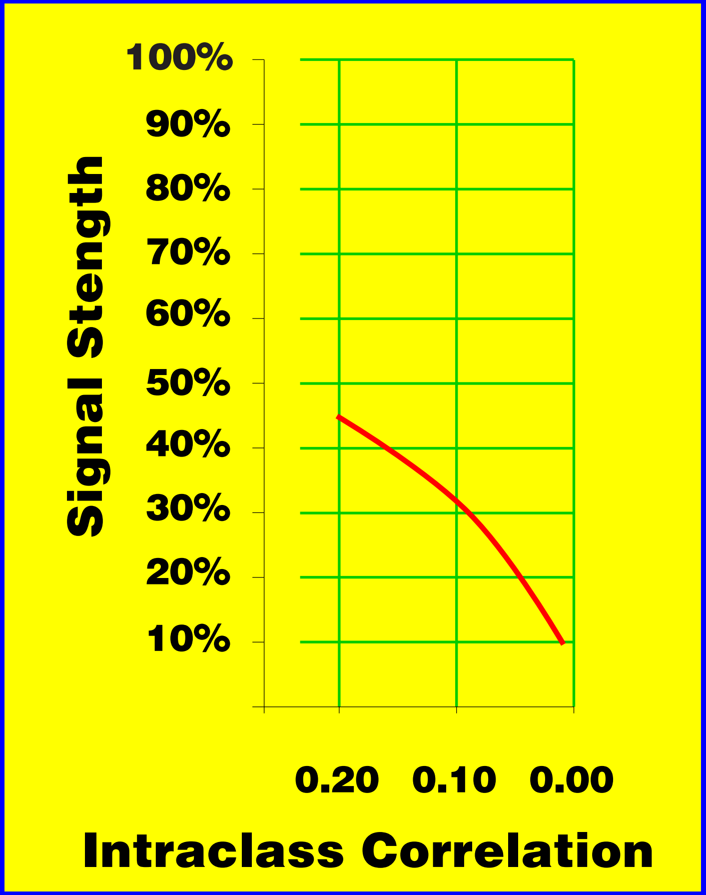


Even though signals from the production process will be attenuated by 30 to 55 percent, when a Third Class Monitor is placed on a Process Behavior Chart it will still have better than a 91% chance of detecting a three standard error shift using Detection Rules 1, 2, 3, and 4.

Fourth Class Monitors

20% > IC

When the Intraclass Correlation is below 20%
the measurement system will provide a Fourth Class Monitor.

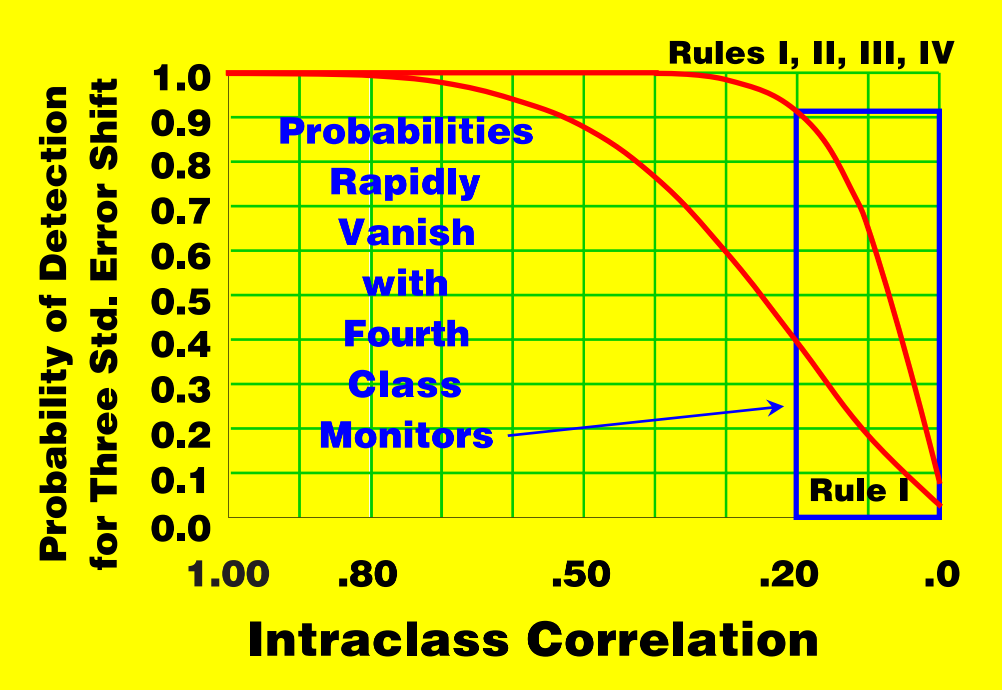


With a Fourth Class Monitor
any signals coming from
the production process
will be attenuated
by more than 55%
due to the dominating
effects of measurement
error.

Fourth Class Monitors

20% > IC

With a Fourth Class Monitor the chances of detecting a three standard error shift using a Process Behavior Chart will rapidly vanish as the measurements come to have less and less information about the production process.



Any use of a Fourth Class Monitor is an act of desperation.

The Four Classes of Process Monitors

Thus the Intraclass Correlation characterizes the relative utility of a measurement system for a given application.

It is theoretically sound,
it is easy to interpret, and
it results in a practical classification scheme.

$$\frac{PV^2}{TV^2} = \textit{Estimated Intraclass Correlation}$$

The Gauge R&R Guidelines

How are these Four Classes of monitors related to the Guidelines given in a Traditional Gauge R&R Study?

These Guidelines are generally applied in the form:

%GRR values that are less than 10% are good,
%GRR values between 10% and 30% are marginal, and
%GRR values that exceed 30% are unacceptable.

To compare these guidelines with the Four Classes we need to know how ***%GRR*** is related to ***IC***.

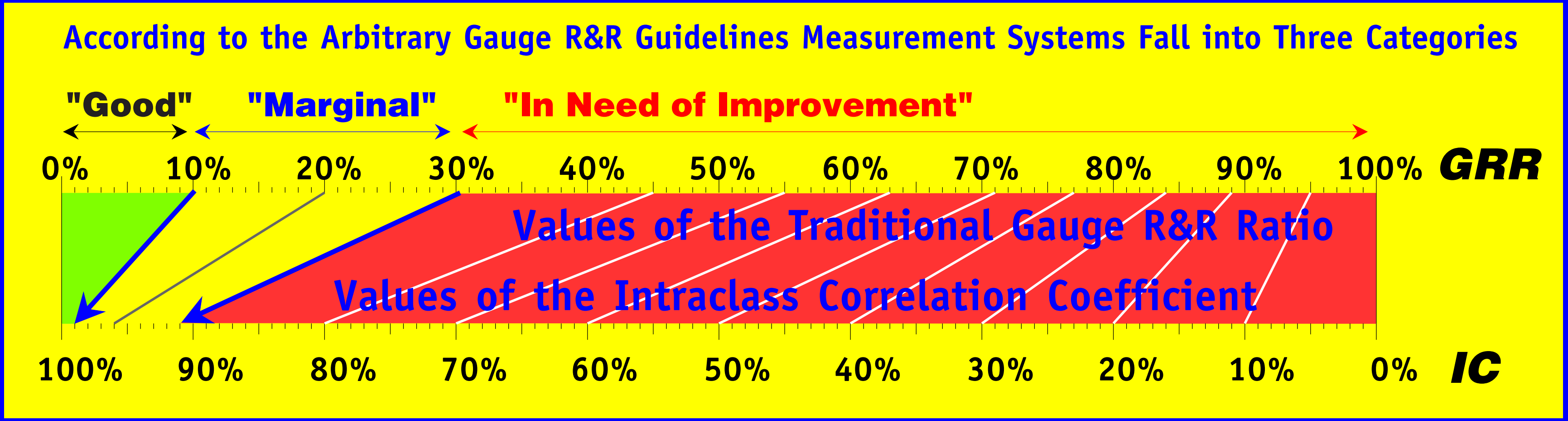
$$\frac{\%GRR}{100} = \frac{GRR}{TV} = \frac{[\,EV^2 + AV^2\,]^{0.5}}{TV} = [\,1 - IC\,]^{0.5}$$

The Gauge R&R Guidelines

$$IC = 1 - \left[\frac{\%GRR}{100}\right]^2$$

So that a *%GRR* value of 10%
corresponds to an Intraclass Correlation of 0.99.

And a *%GRR* value of 30%
corresponds to an Intraclass Correlation of 0.91.



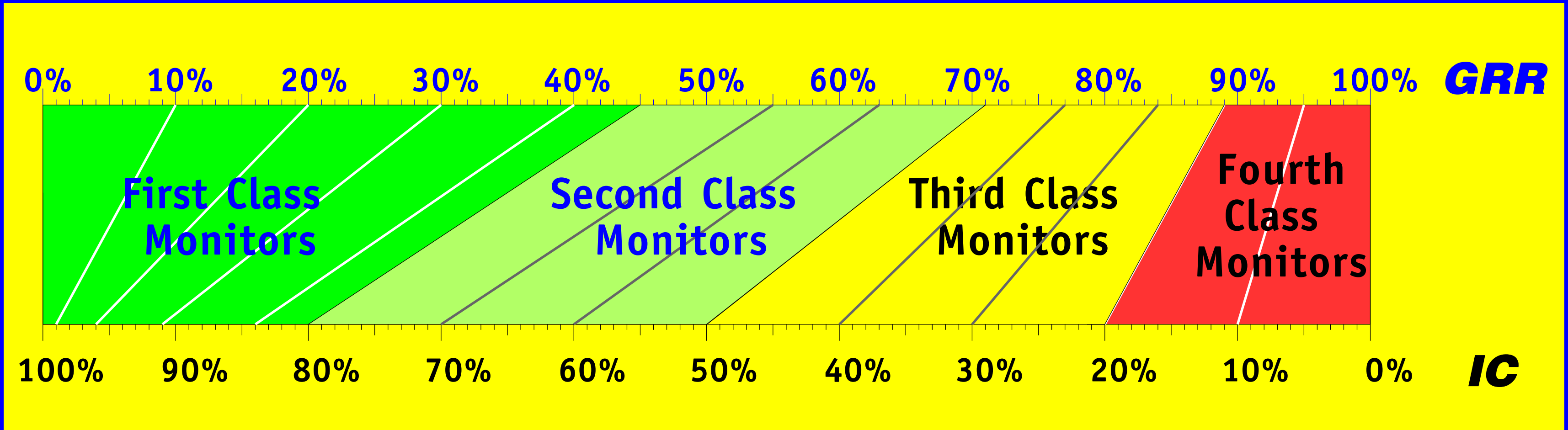
The Four Classes of Monitors

On the other hand:

First Class Monitors will have *%GRR* values below 0.447.

Second Class Monitors will have *%GRR* values below 0.707.

Third Class Monitors will have *%GRR* values below 0.894.



But the best way to see the difference between these two characterizations of relative utility is to consider their impact upon Process Capability.

Quantifying Process Improvements

For the Gasket Thickness the specs are 145 to 225 mils.

With a ***TV*** value of 24.171 mils
we would estimate the current Capability Ratio to be:

$$C_p = \frac{\text{Upper Spec} - \text{Lower Spec}}{6 \ \text{Total Variation}}$$

$$C_p = \frac{225 - 145}{6\,(24.171)} = 0.55$$

With an ***IC*** of 0.944 we have a First Class Monitor,
yet according to the Gauge R&R Study,
with a ***%GRR*** of 23.7% we have
a marginal measurement system.

How much improvement can be tracked according
to these two different approaches to R&R?

Quantifying Process Improvements

Any reduction in the Product Variation
will result in an increase in the Capability Ratio;
an increase in the %GRR value;
and a decrease in the Intraclass Correlation.

For the Gasket Thicknesses:
If the Product Variation dropped from 23.48 to 18.20,
the Capability Ratio would change from 0.55 to 0.70,
the %GRR value would become 30%,
and the Intraclass Correlation would drop to 0.91.

At this point the Gauge R&R Study
would condemn the measurement system.

Therefore, Gauge R&R Studies offer little opportunity
to quantify process improvements.

Quantifying Process Improvements

However, if *PV* dropped from 23.48 to 11.45,
the Capability Ratio would climb from 0.55 to 1.04,
the *%GRR* value would increase to 0.447,
and the Intraclass Correlation would drop to 0.80.

At this point you would be at the crossover from
a First Class Monitor to a Second Class Monitor.

Your measurements would still have a good ability
to detect process changes in a timely manner,
and any such signals from the production process
would only be attenuated by 10 percent
due to the effects of measurement error.

Therefore, First Class Monitors have the ability
to quantify substantial process improvements.

Quantifying Process Improvements

Furthermore, if *PV* dropped from 11.45 to 5.72,
the Capability Ratio would climb from 1.04 to 1.65,
the *%GRR* value would increase to 0.707,
and the Intraclass Correlation would drop to 0.50.

At this point you would be at the crossover from
a Second Class Monitor to a Third Class Monitor.

Your measurements would still have a reasonable ability
to detect process changes in a timely manner,
and any such signals from the production process
would only be attenuated by 30 percent
due to the effects of measurement error.

Therefore, Second Class Monitors still have the ability
to quantify substantial process improvements.

Quantifying Process Improvements

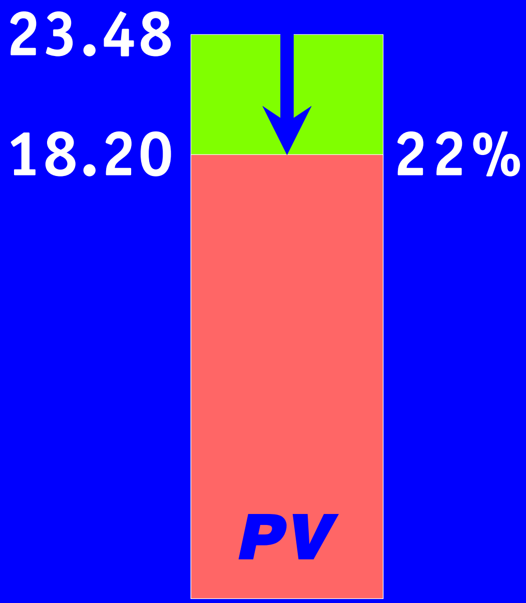
Finally, if *PV* dropped from 5.72 to 2.86,
the Capability Ratio would climb from 1.65 to 2.08,
the *%GRR* value would increase to 0.894,
and the Intraclass Correlation would drop to 0.20.

At this point you would be at the crossover from
a Third Class Monitor to a Fourth Class Monitor.

Signals from the production process
would be attenuated by 55 percent
and the measurement system
would have little remaining utility.

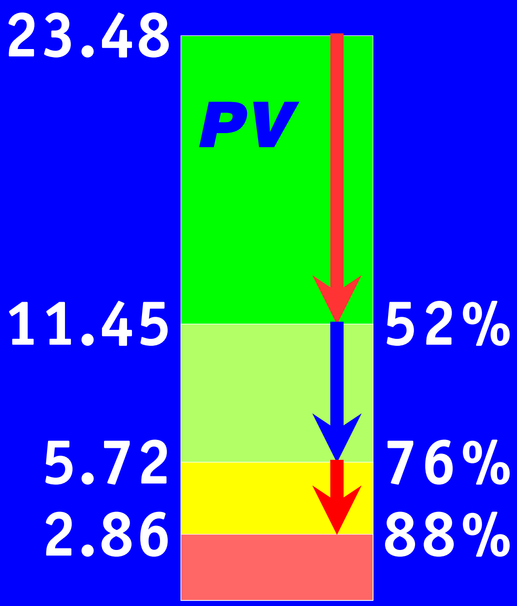
Therefore, Third Class Monitors still have the ability
to quantify process improvements
and detect process changes.

Quantifying Process Improvements



With the Gauge R&R Guidelines
you would condemn
the measurement system used
for the Gasket Thicknesses
after a 22% drop
in the Product Variation.

When in fact a 52% drop in
the Product Variation would only result
in a Second Class Monitor.
A 76% drop would give
a Third Class Monitor, and an
88% drop would be required to make
this measurement system useless.



Quantifying Process Improvements

Thus, by computing the Crossover Capabilities you can determine the ability of a particular measurement system to detect improvements in a particular process.

First Class Monitor

$$C_{p80} = \frac{USL - LSL}{6\sqrt{5}\ GRR}$$

Second Class Monitor

$$C_{p50} = \frac{USL - LSL}{6\sqrt{2}\ GRR}$$

Third Class Monitor

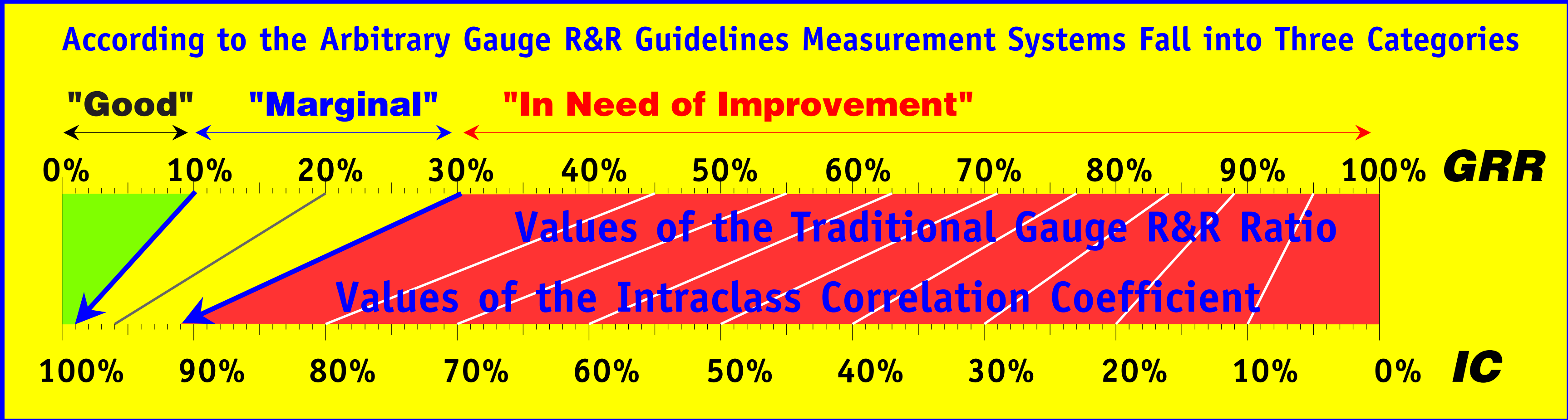
$$C_{p20} = \frac{USL - LSL}{6\sqrt{1.25}\ GRR}$$

Fourth Class Monitor

Increasing Capability

Lessons Learned

Not only is the *%GRR* ratio inflated
by being computed incorrectly,
but the guidelines used to interpret this ratio
are excessively conservative,
and do not even begin to define
the relative utility of the measurement system.



Hence, we must conclude
that the sole purpose of a Gauge R&R Study
is to condemn the measurement process.

Lessons Learned

It has been demonstrated that:

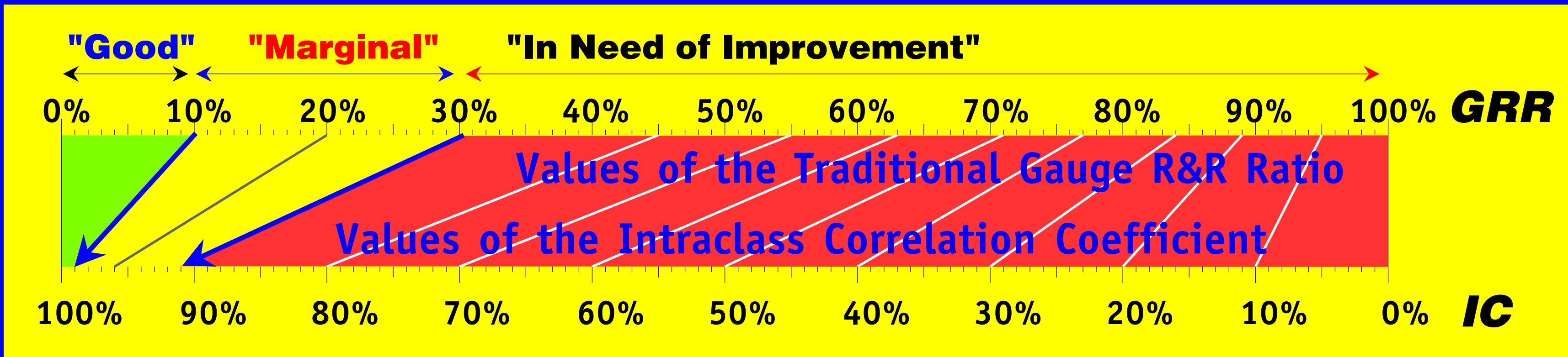
The ratios computed in Steps 7, 8, 9, & 10
of a Traditonal Gauge R&R Study do not represent
what they are said to represent.
(This is true for steps 11 through 14 as well.)

The Guidelines used by Traditional Gauge R&R Studies
are so conservative that they are nonsense.

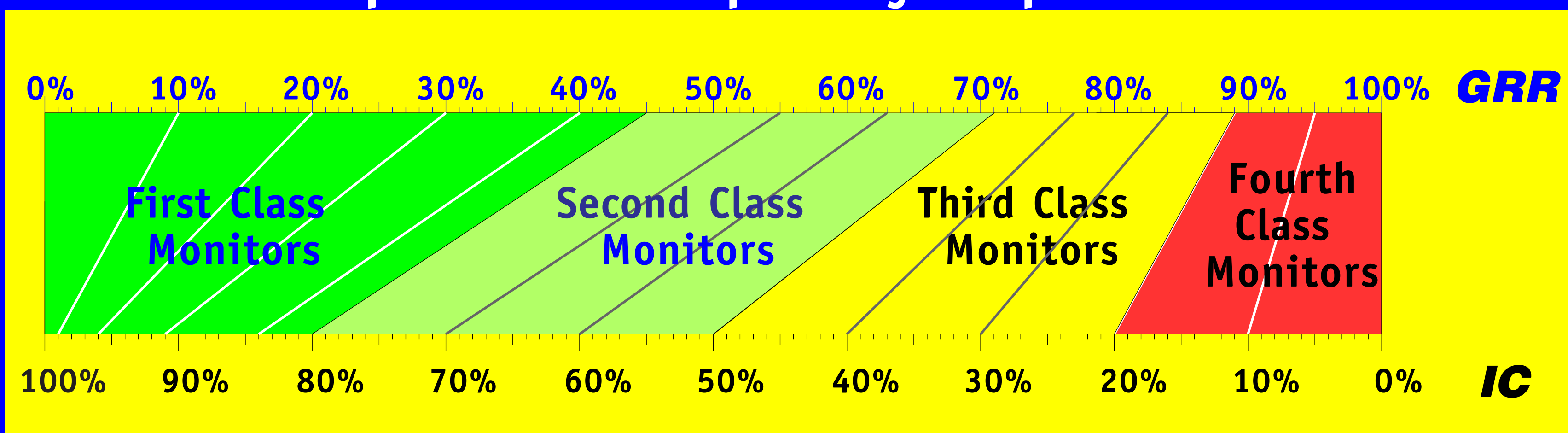
The proper measure of relative utility
is the Intraclass Correlation, which can be used
to define four clear and meaningful classes
of process monitors.

The Choice is Clear

Do you want to condemn your measurement systems?



Or would you prefer to use your less-than-perfect data to operate and improve your processes?



An Honest Gauge R&R Study

Step H.7	Repeatability	$\frac{EV^2}{TV^2}$
Step H.8	Reproducibility	$\frac{AV^2}{TV^2}$
Step H.9	Combined R&R	$\frac{GRR^2}{TV^2}$
Step H.10	Product Variation or Intraclass Correlation	$\frac{PV^2}{TV^2}$

1.00 ← IC → 0.80 ← IC → 0.50 ← IC → 0.20 ← IC → 0.0

First Class	Second Class	Third Class	Fourth Class

An Honest Gauge R&R Study

IC		Attenuation of Process Signals	Chance of Detecting a 3 Std. Error Shift	Track Process Improv.
1.00				
	First Class Monitors	Less than 10 Percent	More than 99% with Rule One	To C_{p80}
0.80				
	Second Class Monitors	10 Percent to 30 Percent	More than 88% with Rule One	To C_{p50}
0.50				
	Third Class Monitors	30 Percent to 55 Percent	More than 91% w/Rules 1,2,3,4	To C_{p20}
0.20				
	Fourth Class Monitors	More than 55 Percent	Rapidly Vanishes	*Cannot Track*
0.00				

djwheeler@spcpress.com