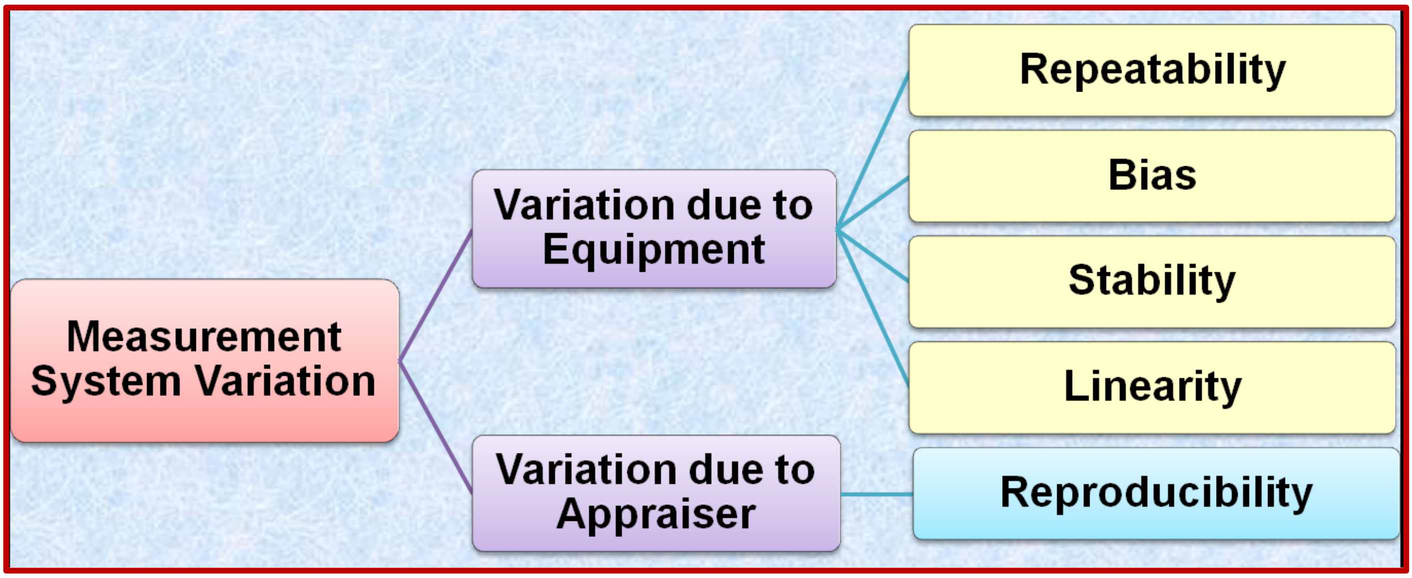
Measurement Assurance

The ability to measure is the foundation of a good quality program. This is especially true as more and more measurement processes are being automated. Before relying on a measurement to make decisions, it is crucial to be confident in the measurement process. Is the variability in the process due to the process itself, or due to measurement error? How precise is the measurement process? Does the ability to measure depend on who is doing the measuring? The total observed variability in an item includes the variability of the measurement process.

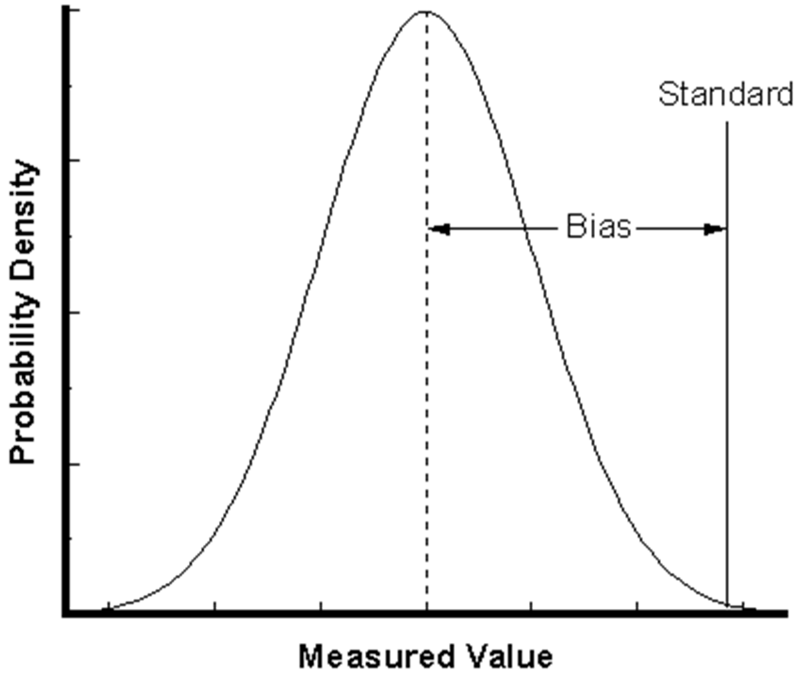
$$\sigma^2_{\text{Total}} = \sigma^2_{\text{Production Process}} + \sigma^2_{\text{Measurement Process}}$$

Measuring is a process, and the variability of the measurement process is a combination of the variability of the instrument being used, the variability of the operators using the gage, and the ability to maintain the accuracy and precision of the instrument being used. Like any other process, the measurement process must be controlled, otherwise the variability of the measurement process will increase.

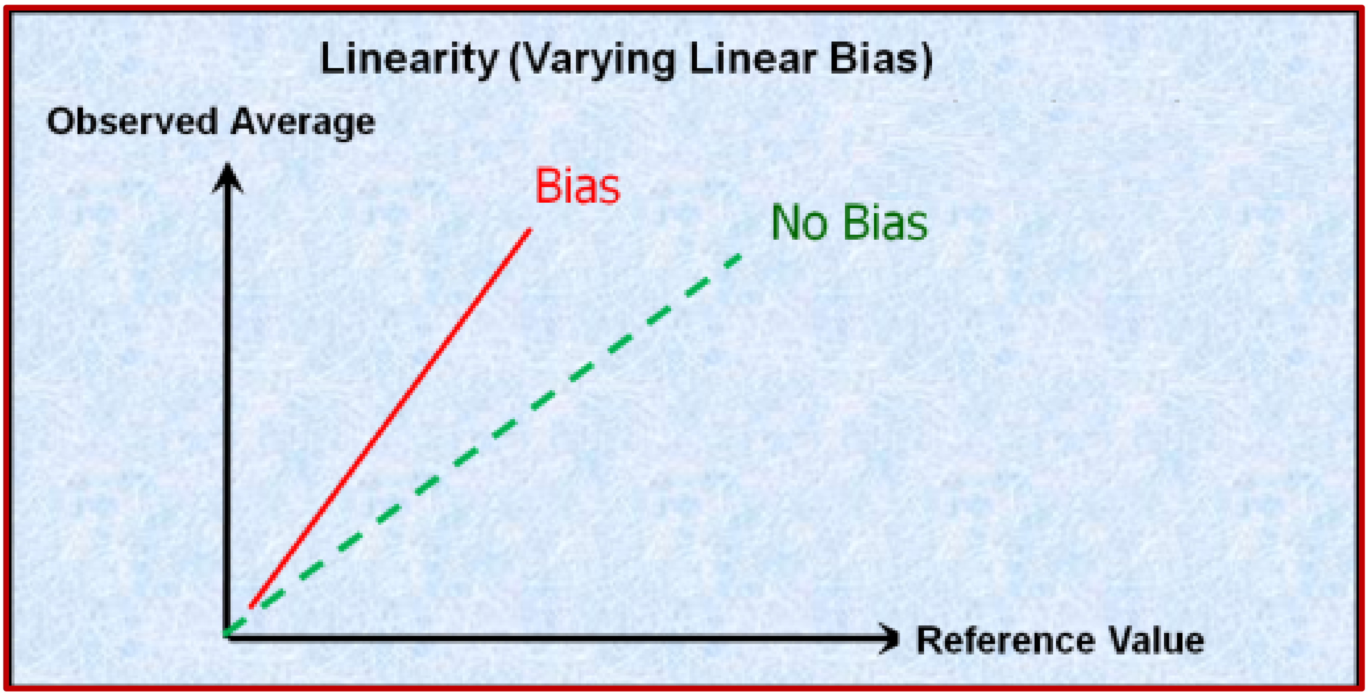
A measurement system is controlled by keeping key aspects of the measurement process in control. These are *bias*, *repeatability*, *reproducibility*, *stability*, and *linearity*.



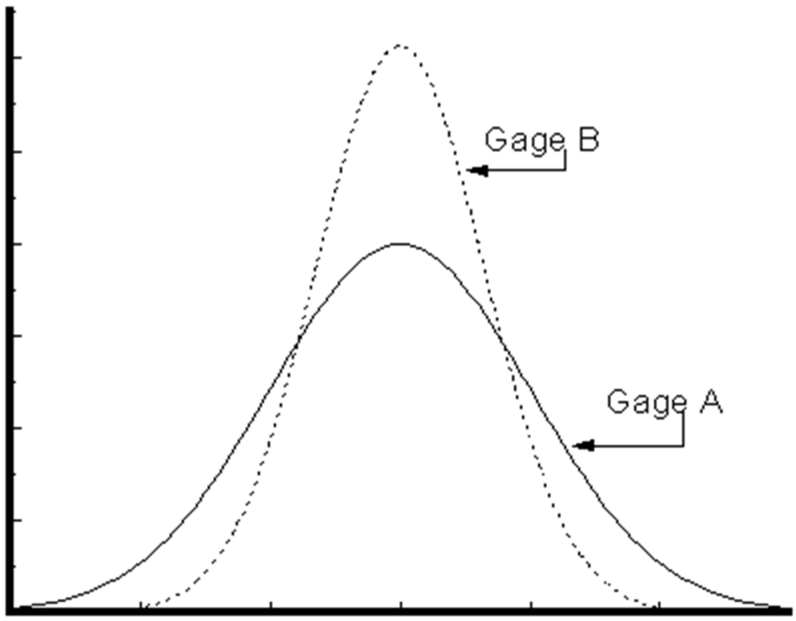
Bias is the difference between the output of the measurement system and the true value and is often referred to as *accuracy*. The concept of bias is shown statistically in the figure below.



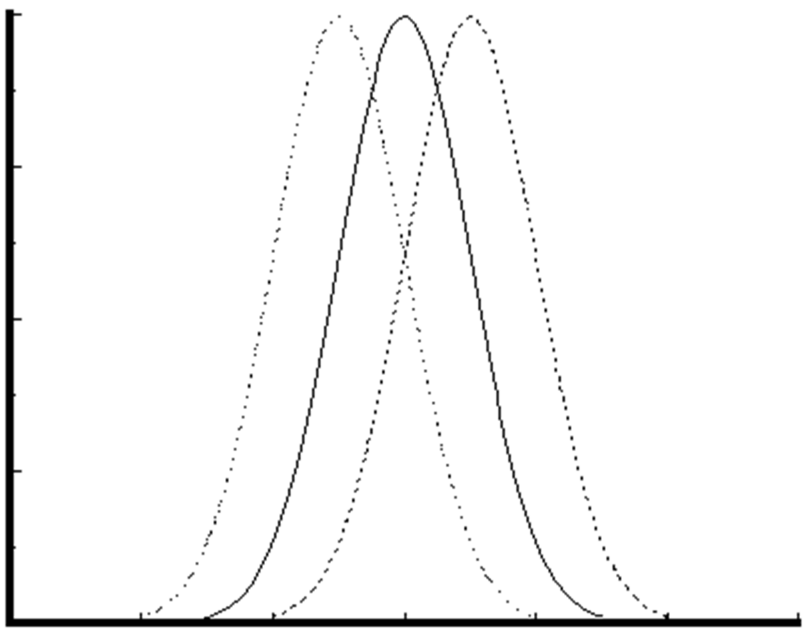
This figure demonstrates negative bias, the gage underestimates. If the gage overestimates, the bias is positive.



Repeatability is the variability of the measurements obtained by one person while measuring the same item repeatedly. This is also known as the inherent precision of the measurement equipment. Consider the probability density functions shown in the figure below. The density functions were constructed from measurements of the thickness of a piece of metal with Gage A and Gage B. The density functions demonstrate that Gage B is more repeatable than Gage A.

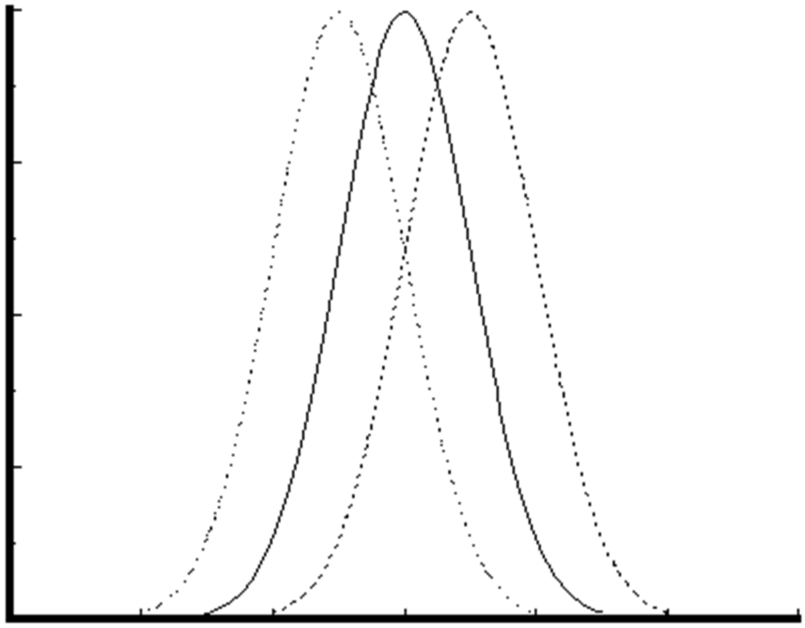


Reproducibility is the variability of the measurement system caused by differences in operator behaviour. Mathematically, it is the variability of the average values obtained by several operators while measuring the same item. The figure below displays the probability density functions of the measurements for three operators. The variability of the individual operators is the same, but because each operator has a different bias, the total variability of the measurement system is higher when three operators are used than when one operator is used.

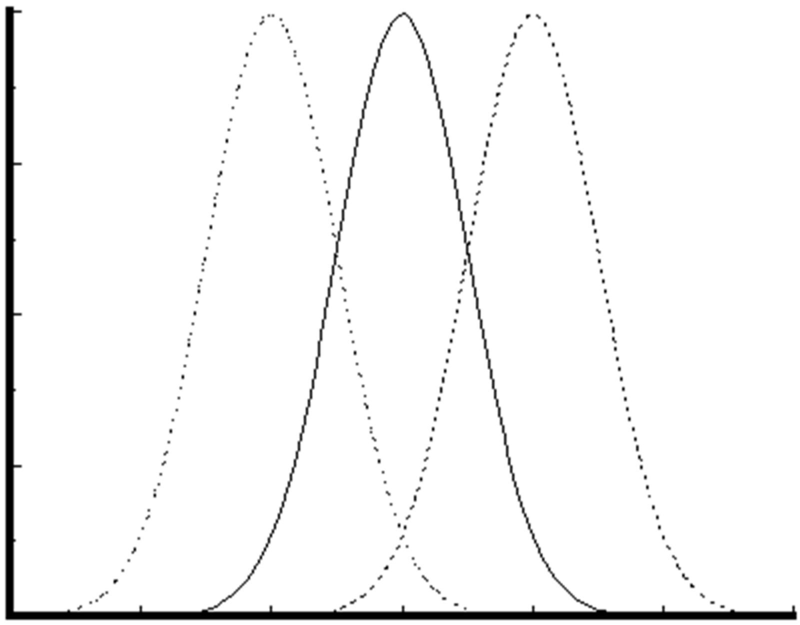


The figure below also displays the probability density functions of the measurements for three operators using the same scale as the previous figure. Notice that there is more difference in the means of the measurements shown in the second figure than the first. The reproducibility of the system shown in the second figure is higher than the reproducibility of the system shown in the first figure.

Stability is a measure of the variability introduced into a measurement system over time; a stable system does not drift with time. Stability is like reproducibility, but the bias shift is caused by the elapse of time rather than the difference in operators. The figure below shows the density functions of a measurement system for three distinct shifts of operation. The system is not stable, because the mean is not the same for all shifts.



The figure below shows the density functions for three distinct shifts of operation with the same scales as the previous figure. The system shown in the figure below is less stable than the system shown in the figure above.

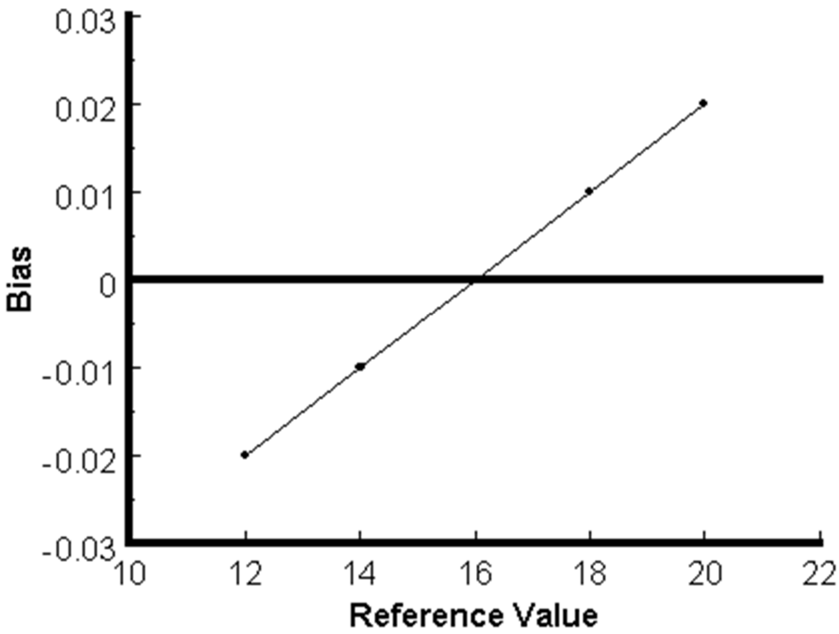


Linearity is a measure of the bias change throughout the operating range of the gage. For example, consider the data in the table below.

Reference Value	Average Reading	Average Bias
12	11.98	-0.02
14	13.99	-0.01
16	16.00	0.00
18	18.01	0.01

20	20.02	0.02

The magnitude of the bias increases with the magnitude of the reference value. This bias is plotted as a function of the reference value in the figure below. The slope of the plotted line, 0.005, is the linearity of the measurement system.



Detailed calculations are available in the Microsoft Excel file "**Chapt-4 Measurement Assurance.xlsx**".

Bias

Bias is a measure of the accuracy of a gage. Bias is computed from the expression:

$$\text{Bias} = \frac{\sum_{i=1}^{n} x_i}{n} - \tau$$

where n is the number of times the standard is measured, x_i is the i-th measurement, and τ is the value of the standard. Bias is usually reported as a percent of process variation or as a percent of tolerance. Given a process standard deviation of σ, bias as a percent of process variation (6σ) is:

$$\text{Bias(Process Variation)} = 100\left[\frac{\frac{\sum_{i=1}^{n} x_i}{n} - \tau}{6\sigma}\right]$$

Given a tolerance (Upper specification - Lower specification) of T, bias as a percent of process tolerance is:

$$\text{Bias(Tolerance)} = 100\left[\frac{\frac{\sum_{i=1}^{n} x_i}{n} - \tau}{T}\right]$$

To perform a bias study, a reference standard with a known value is necessary. Measure the part on the gage being tested a minimum of 10 times, and preferably more than 30 times. The greater the number of measurements, the greater the accuracy. In virtually every case, there will be bias, however, the bias may not be statistically significant. The significance is tested using the t-distribution, and increasing the number of measurements increases the discriminatory power of the t-test.

A gage is biased if the average measure of an item with a known value is significantly different than the known value. The significance of the bias is determined using a *t*-test. The test hypotheses are:

Ho: $\mu = \tau$

H1: μ does not equal τ

The test statistic is:

$$t = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

where X-bar is the average of the measurements, μ is the known measure of the standard, σ is the sample standard deviation of the measurements, and *n* is the number of measurements taken. For a 2-tailed test, the critical values of the t-distribution are:

$-t_{\alpha/2,\, n-1}$ and $t_{\alpha/2,\, n-1}$

where α is the level of significance chosen for the test.

Using a t-test assumes the measured values are normally distributed. This is not a critical assumption unless *n* is small because the t-test is testing the population average, and the distribution of the average tends to normal as n increases regardless of the distribution of the individuals. In most cases *n* = 10 is sufficient, even in cases where the distribution of the individuals greatly differs from normal, a sample size of 30 is usually sufficient.

Example

A block is known to weigh 100.3 pounds. This block is measured 30 times on a scale, and the resulting measures in pounds are given in the table below. The process standard deviation is 0.32, and the process tolerance is 7.4 (Upper Specification - Lower Specification). What is the bias, and is the bias statistically different from zero at a significance of 10%?

100.1	100.6	101.5	101.6	101.8
101.9	100.2	100.8	101.3	100.0
101.0	101.1	102.0	100.5	100.7
100.2	101.6	100.9	100.6	100.4
100.6	100.8	100.5	101.0	100.0
101.5	101.4	101.3	100.3	101.3

Solution

The average weight is 100.92. The bias of the scale is:

Bias = 100.92 - 100.3 = 0.62

Is this true bias, or is it due to the random error? Five of the 30 measurements underestimated the known weight of 100.3. To test for significance, compute samples standard deviation of the 30 values, and compute the t-statistic. The sample standard deviation is, *s* = 0.587, and the t-statistic is

$$t = \frac{\overline{X} - \mu}{s / \sqrt{n}} = \frac{100.92 - 100.3}{0.587 / \sqrt{30}} = 5.78$$

Using a significance level of $\alpha = 0.1$, with 29 degrees of freedom (30 -1), the critical values of the t-distribution are 1.699 and -1.699. These values can be found with the Excel function:

=TINV(0.1,29)

Since the computed t-statistic does not fall between the critical values, the bias is significant, and the difference from the true value is not due to random error. If possible, the scale should be adjusted to yield a bias of zero.

Detailed calculations are available in the Microsoft Excel file "**Chapt-4 Measurement Assurance.xlsx**".

Repeatability & Reproducibility

The analysis of variance method (ANOVA) is the most accurate method for quantifying repeatability and reproducibility. In addition, the ANOVA method allows the variability of the interaction between the appraisers and the parts to be determined. The ANOVA method for measurement assurance is the same statistical technique used to analyse the effects of different factors in designed experiments. The ANOVA design used is a two-way, fixed effects model with replications. The ANOVA table is shown below.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Appraiser	SSA	a-1	MSA = SSA/(a-1)	F = MSA/MSE
Parts	SSB	b-1	MSB = SSB/(b-1)	F = MSB/MSE
Interaction (Appraiser, Parts)	SSAB	(a-1)(b-1)	MSAB = SSAB/[(a-1)(b-1)]	F = MSAB/MSE
Gage (Error)	SSE	ab(n-1)	MSE = SSE/[ab(n-1)]	
Total	TSS	N-1		

$$SSA = \sum_{i=1}^{a} \frac{(Y_{i..})^2}{bn} - \frac{Y_{\bullet\bullet}^2}{N}$$

$$SSB = \sum_{j=1}^{b} \frac{(Y_{.j.})^2}{an} - \frac{Y_{\bullet\bullet}^2}{N}$$

$$SSAB = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{(Y_{ij.})^2}{n} - \frac{Y_{\bullet\bullet}^2}{N} - SSA - SSB$$

$$TSS = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk}^2 - \frac{Y_{\bullet\bullet}^2}{N}$$

SSE = *TSS* - *SSA* - *SSB* - *SSAB*

a = number of appraisers,
b = number parts,

n = the number of trials, and
N = total number of readings (abn)

When conducting a study, the recommended procedure is to use 10 parts, 3 appraisers and 2 trials, for a total of 60 measurements. The measurement system repeatability is:

$$\text{Repeatability} = 5.15\sqrt{MSE}$$

The measurement system reproducibility is:

$$\text{Reproducibility} = 5.15\sqrt{\frac{MSA - MSAB}{bn}}$$

The interaction between the appraisers and the parts is:

$$\mathrm{I} = 5.15\sqrt{\frac{MSAB - MSE}{n}}$$

The measurement system repeatability and repeatability is:

$$R\&R = \sqrt{\text{Repeatability}^2 + \text{Reproducibility}^2 + I^2}$$

The measurement system part variation is:

$$V_P = 5.15\sqrt{\frac{MSB - MSAB}{an}}$$

The total measurement system variation is:

$$V_T = \sqrt{R\&R^2 + V_P^2}$$

Example

The thickness, in millimetres, of 10 parts have been measured by 3 operators, using the same measurement equipment. Each operator measured each part twice, and the data is given in the table below.

	Operator					
	A		**B**		**C**	
Part	**Trial 1**	**Trial 2**	**Trial 1**	**Trial 2**	**Trial 1**	**Trial 2**
1	65.2	60.1	62.9	56.3	71.6	60.6
2	85.8	86.3	85.7	80.5	92.0	87.4
3	100.2	94.8	100.1	94.5	107.3	104.4
4	85.0	95.1	84.8	90.3	92.3	94.6
5	54.7	65.8	51.7	60.0	58.9	67.2
6	98.7	90.2	92.7	87.2	98.9	93.5
7	94.5	94.5	91.0	93.4	95.4	103.3

8	87.2	82.4	83.9	78.8	93.0	85.8
9	82.4	82.2	80.7	80.3	87.9	88.1
10	100.2	104.9	99.7	103.2	104.3	111.5

Solution

To compute the characteristics of this measurement system, the two-way ANOVA table must be completed. The sum of the 20 readings (10 parts multiplied by 2 trials) for appraiser A is 1710.2. The sum of the 20 readings for appraiser B is 1657.7. The sum of the 20 readings for appraiser C is 1798.0, and the sum of all 60 readings is 5165.9. The sum-of-squares for the appraisers is

$$SSA = \frac{1710.2^2}{10(2)} + \frac{1657.7^2}{10(2)} + \frac{1798.0^2}{10(2)} - \frac{5165.9^2}{60} = 502.5$$

The sum of the 6 readings for each part (3 appraisers multiplied by 2 trials) is given in the table below.

Part	Sum	Sum Squared	(Sum Squared)/6
1	376.7	141,902.9	23,650.5
2	517.7	268,013.3	44,668.9
3	601.3	361,561.7	60,260.3
4	542.1	293,872.4	48,978.7
5	358.3	128,378.9	21,396.5
6	561.2	314,945.4	52,490.9
7	572.1	327,298.4	54,549.7
8	511.1	261,223.2	43,537.2
9	501.6	251,602.6	41,933.8
10	623.8	389,126.4	64,854.4
Total			456,320.9

The sum-of-squares for the parts is

$$SSB = 456{,}320.9 - \frac{5165.9^2}{60} = 11{,}545.5$$

The sum of the 2 trials for each combination of appraiser and part is given in the table below along with the square of this sum, and the square of this sum divided by 2.

Part	Appraiser	Sum	Sum Squared	Sum Squared/2
1	A	125.3	15,700.1	7,850.0
2	A	172.1	29,618.4	14,809.2
3	A	195.0	38,025.0	19,012.5
4	A	180.1	32,436.0	16,218.0

5	A	120.5	14,520.3	7,260.1
6	A	188.9	35,683.2	17,841.6
7	A	189.0	35,721.0	17,860.5
8	A	169.6	28,764.2	14,382.1
9	A	164.6	27,093.2	13,546.6
10	A	205.1	42,066.0	21,033.0
1	B	119.2	14,208.6	7,104.3
2	B	166.2	27,622.4	13,811.2
3	B	194.6	37,869.2	18,934.6
4	B	175.1	30,660.0	15,330.0
5	B	111.7	12,476.9	6,238.4
6	B	179.9	32,364.0	16,182.0
7	B	184.4	34,003.4	17,001.7
8	B	162.7	26,471.3	13,235.6
9	B	161.0	25,921.0	12,960.5
10	B	202.9	41,168.4	20,584.2
1	C	132.2	17,476.8	8,738.4
2	C	179.4	32,184.4	16,092.2
3	C	211.7	44,816.9	22,408.4
4	C	186.9	34,931.6	17,465.8
5	C	126.1	15,901.2	7,950.6
6	C	192.4	37,017.8	18,508.9
7	C	198.7	39,481.7	19,740.8
8	C	178.8	31,969.4	15,984.7
9	C	176.0	30,976.0	15,488.0
10	C	215.8	46,569.6	23,284.8
Total				456,859.0

The sum-of-squares for the interaction between the appraisers and the parts, SSAB, is:

$$SSAB = 456{,}859.0 - \frac{5165.9^2}{60} - 502.5 - 11{,}545.5 = 35.6$$

Squaring all 60 individual reading and summing the values gives 457,405.8. The total sum-of-squares is:

$$TSS = 457{,}405.8 - \frac{5165.9^2}{60} = 12{,}630.4$$

There are 2 degrees-of-freedom for the appraisers, the number of appraisers minus one; 9 degrees-of-freedom for the parts, the number of parts minus one, 18 degrees-of-freedom for the interaction between the appraisers and the parts, the number of appraisers minus one multiplied by the number of parts minus one; 59 total degrees-of-freedom; the total number of readings minus one, and 30 degrees-of-freedom for the gage, total degrees-of-freedom minus the degrees-of-freedom for the appraisers minus the degrees-of-freedom for the parts minus the degrees-of-freedom for the interaction. Since the mean-square-error is the sum-of-square divided by the degrees-of-freedom, the ANOVA table can be completed as shown in the table below.

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**F**	**Significance**
Appraiser	502.5	2	251.3	13.8	0.000057
Parts	11,545.5	9	1,282.8	70.4	0.000000
Interaction	35.6	18	1.98	0.11	0.999996
(Appraisers, Parts)					
Gage (Error)	546.8	30	18.2		
Total	12,630.4	59	214.1		

The significance listed in the table above represents the probability of Type I error. Stated another way, if the statement is made "the appraisers are a significant source of measurement variability", the probability of this statement being incorrect is 0.000057. This significance is the area under the F probability density function to the right of the computed F-statistic. This value can be found by using the function =FINV(F-statistic,d1,d2) in Microsoft Excel, where d1 and d2 are the appropriate degrees of freedom.

Continuing with the example, the repeatability is

$$\text{Repeatability} = 5.15\sqrt{18.2} = 22.0$$

Reproducibility is:

$$\text{Reproducibility} = 5.15\sqrt{\frac{251.3 - 1.98}{10(2)}} = 18.2$$

The interaction between the appraisers and the parts is

$$I = 5.15\sqrt{\frac{1.98 - 18.2}{n}} = \text{cannot be computed}$$

Obviously, the variability due to the interaction cannot be imaginary (the square root of a negative number is an imaginary number); what happened? Each mean square is an estimate subject to sampling error. In some cases, the estimated variance will be negative or imaginary. In these cases, the estimated variance is zero.

The repeatability and reproducibility is

$$R\&R = \sqrt{22^2 + 18.2^2 + 0^2} = 28.6$$

The part variation is:

$$V_P = 5.15\sqrt{\frac{1{,}282.8 - 1.98}{3(2)}} = 75.2$$

The total measurement system variation is:

$$V_T = \sqrt{28.6^2 + 75.2^2} = 80.5$$

Detailed calculations are available in the Microsoft Excel file "**Chapt-4 Measurement Assurance.xlsx**".

Linearity

Measurement system linearity is found by measuring reference part values throughout the operating range of the instrument and plotting the bias against the reference values. The default procedure for determining linearity is to measure 10 parts 5 times each. The percent linearity is equal to the slope, *b*, of the best-fit straight line through the data points, and the linearity is equal the slope multiplied by the process variation.

$L = bV_p$

The bias at any point can be estimate from the slope and the *y*-intercept, y_o of the best-fit line.

$B = y_o + bx$

The estimated slope of the line, *b*, is:

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

where *n* is the number of data points, x_i is the *i*-th reference value, and y_i is the *i*th bias.

The estimated *y*-intercept, y_o, of the line is:

$$y_0 = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n}$$

The coefficient of determination, r^2, is equal to the percentage of variability in the data set explained by the best-fit line.

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{xy}}$$

where *S*xx is the sum of squares of *x*,

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

and *Syy* is the sum of squares of *y,*

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

and *Sxy* is the sum of squares of *x* and *y.*

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

Both Lotus 123 and Microsoft Excel have linear regression routines that automate the task of computing the characteristics of the best-fit straight line. In Lotus 123, the slope is computed using the function @Regression("X-range", "Y-range", 101), the *y*-intercept is computed using the function @Regression("X-range", "Y-range", 1), and the coefficient of determination is computed using the function @Regression("X-range", "Y-range", 3). In Microsoft Excel, the slope is computed using the function =Slope("Y-range", "X-range"), the *y*-intercept is computed using the function =Intercept("Y-range", "X-range"), and the coefficient of determination is computed by squaring the function = Correl("Y-range", "X-range").

Example

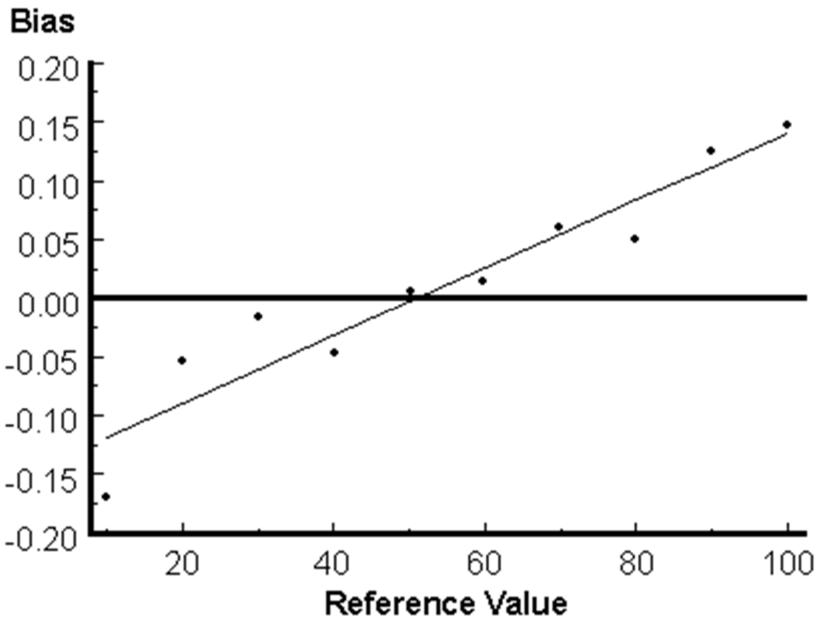
The table below shows the results when 10 parts were measured 5 times each. The last column of this table shows the computed bias. The process standard deviation is 0.03.

Reference	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Bias
10	9.91	9.71	9.96	9.69	9.90	-0.17
20	19.92	20.16	19.88	19.85	19.92	-0.05
30	29.68	30.15	30.46	29.48	30.15	-0.02
40	40.03	39.93	40.08	39.69	40.03	-0.05
50	50.31	49.78	50.04	49.86	50.06	0.00
60	60.58	59.34	59.61	59.88	60.66	0.01
70	70.14	69.93	69.52	69.86	70.85	0.06
80	79.10	79.93	79.83	80.78	80.60	0.05
90	89.40	89.40	90.47	89.96	91.39	0.12
100	100.91	99.49	100.24	100.67	99.42	0.15

Solution

The figure below shows the bias plotted versus the reference value. The slope of the best-fit line is 0.00288, and the y-intercept is -0.14733. The percent linearity is equal to the slope, 0.288%, and the linearity is equal to the slope multiplied by the process variation.

$L = 0.00288(0.03) = 0.0000864$



The bias at any value is estimated from the expression:

$B = -0.14733 + 0.002881x$

where x is the reference value.

Detailed calculations are available in the Microsoft Excel file "**Chapt-4 Measurement Assurance.xlsx**".