Repeatability and Reproducibility

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 Figure 1. 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.
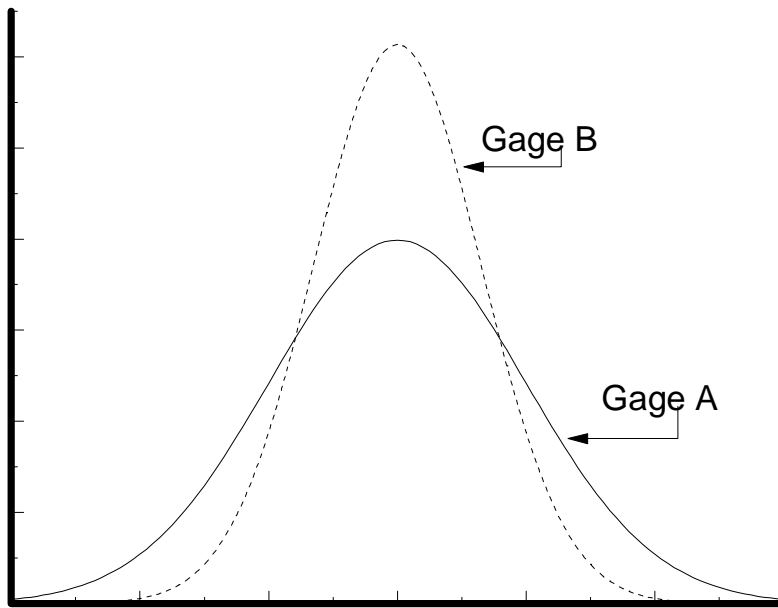
Figure 1. Probability density functions for the thickness of 2 gages.

Reproducibility is the variability of the measurement system caused by differences in operator behavior. Mathematically, it is the variability of the average values obtained by several operators while measuring the same item. Figure 2 displays the probability density functions of the measurements for three operators. The variability of the individual operators are 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. Figure 3 also displays the probability density functions of the measurements for three operators using the same scale as Figure 2. Notice that there is more difference in the means of the measurements shown in Figure 3 than those shown in Figure 2. The reproducibility of the system shown in Figure 3 is higher than the reproducibility of the system shown in Figure 2.
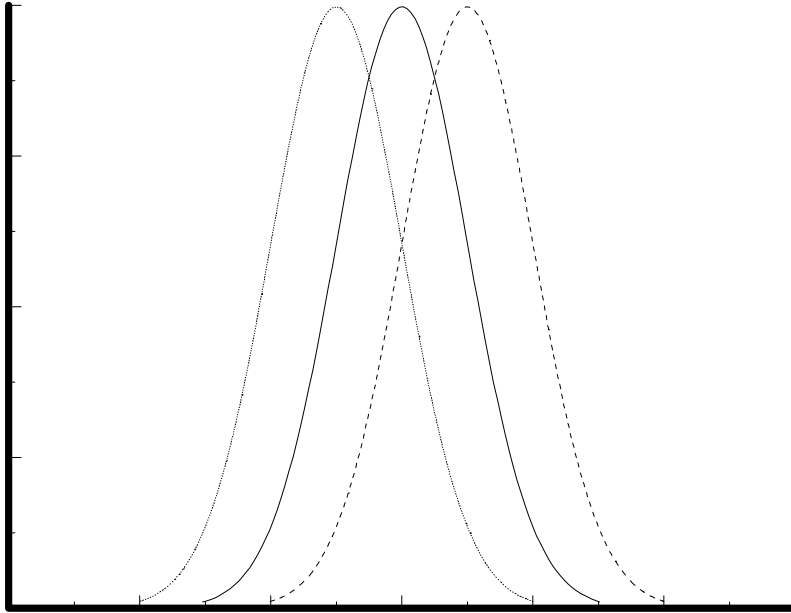
Figure 2. Reproducibility demonstration.
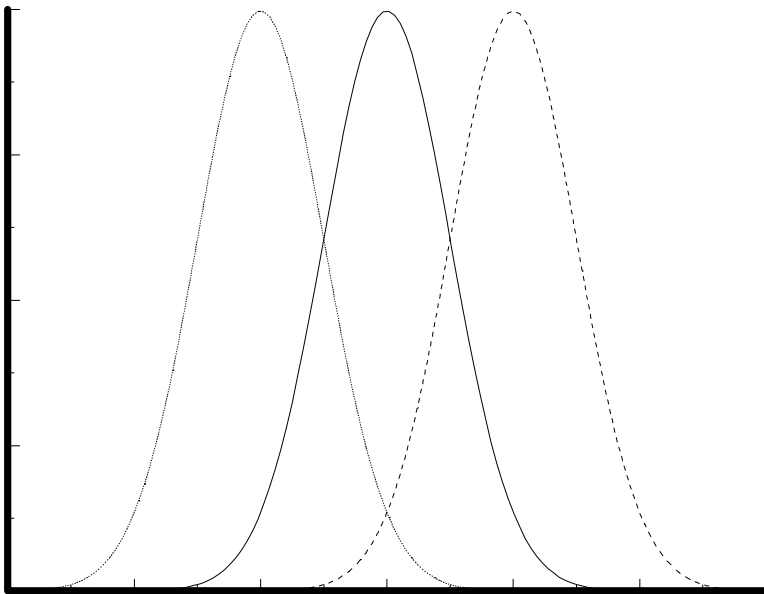
Figure 3. Reproducibility demonstration.

The most commonly used method for computing repeatability and reproducibility is the Range and Average method. The ANOVA method is more accurate, but because of the complex mathematics involved it has been shunned historically. With desktop computers there is no excuse for not using the more accurate ANOVA method.

Range & Average Method

The *Range & Average Method* computes the total measurement system variability, and allows the total measurement system variability to be separated into repeatability, reproducibility, and part variation.

The ANOVA method, discussed in the next section, is preferred to the average range method. The ANOVA method quantifies the interaction between repeatability and reproducibility, and is considered to be more accurate than the average and range method.

To quantify repeatability and reproducibility using average and range method, multiple parts, appraisers, and trials are required. The recommended method is to use 10 parts, 3 appraisers and 2 trials, for a total of 60 measurements. The measurement system repeatability is

$$\text{Repeatability} = \frac{5.15\overline{\overline{R}}}{d_2} \qquad\qquad 1$$

where $\overline{\overline{R}}$ is the average of the ranges for all appraisers and parts, and
d_2 is found in Appendix A with Z = the number of parts times the number of appraisers, and W = the number of trials.

The measurement system reproducibility is

$$\text{Reproducibility} = \sqrt{\left(\frac{5.15\overline{X}_{\text{range}}}{d_2}\right)^2 - \frac{\text{Repeatability}^2}{nr}} \qquad\qquad 2$$

where $\overline{X}_{range}$ is the average of the difference in the average measurements between the appraiser with the highest average measurements, and the appraiser with the lowest average measurements, for all appraisers and parts,
d_2 is found in Appendix A with $Z = 1$ and W = the number of appraisers,
n is the number of parts, and
r is the number of trials.

The measurement system repeatability and reproducibility is

$$R\&R = \sqrt{\text{Repeatability}^2 + \text{Reproducibility}^2} \qquad\qquad 3$$

The part variability is

$$V_P = \frac{5.15R_P}{d_2} \qquad\qquad 4$$

where R_p is the difference between the largest average part measurement and the smallest
average part measurement, where the average is taken for all appraisers and all
trials, and

d_2 is found in Appendix A with $Z = 1$ and $W =$ the number of parts.

The total variability, measurement system variability and part variation combined is

$$V_T = \sqrt{R\&R^2 + V_P^2} \qquad\qquad\qquad 5$$

Example 1

The thickness, in millimeters, 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 Table 1.

Table 1. Range & Average method example data.

Part	Operator					
	A		**B**		**C**	
	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

Repeatability is computed using the average of the ranges for all appraiser and all
parts. This data is given in Table 2.

Table 2. Example problem range calculations.

	Operator								
	A			B			C		
Part	Trial 1	Trial 2	R	Trial 1	Trial 2	R	Trial 1	Trial 2	R
1	65.2	60.1	5.1	62.9	56.3	6.6	71.6	60.6	11.0
2	85.8	86.3	0.5	85.7	80.5	5.2	92.0	87.4	4.6
3	100.2	94.8	5.4	100.1	94.5	5.6	107.3	104.4	2.9
4	85.0	95.1	10.1	84.8	90.3	5.5	92.3	94.6	2.3
5	54.7	65.8	11.1	51.7	60.0	8.3	58.9	67.2	8.3
6	98.7	90.2	8.5	92.7	87.2	5.5	98.9	93.5	5.4
7	94.5	94.5	0.0	91.0	93.4	2.4	95.4	103.3	7.9
8	87.2	82.4	4.8	83.9	78.8	5.1	93.0	85.8	7.2
9	82.4	82.2	0.2	80.7	80.3	0.4	87.9	88.1	0.2
10	100.2	104.9	4.7	99.7	103.2	3.5	104.3	111.5	7.2

The average of the 30 ranges, $\overline{\overline{R}}$, is 5.20. From Appendix A, with $Z = 30$ (10 parts multiplied by 3 appraisers) and $W = 2$ (2 trials), d_2 is 1.128. The repeatability is

$$\text{Repeatability} = \frac{5.15(5.20)}{1.128} = 23.7$$

The average reading for appraiser A is 85.5, the average reading for appraiser B is 82.9, and the average reading for appraiser C is 88.9. To compute reproducibility, the average of the range between the appraiser with the smallest average reading (appraiser B in this example) and the appraiser with the largest average reading (appraiser C in this example) is needed. Table 3 shows this data.

Table 3. Reproducibility example computations.

Part	Trial	Operator B	Operator C	R
1	1	62.9	71.6	8.7
2	1	85.7	92.0	6.3
3	1	100.1	107.3	7.2
4	1	84.8	92.3	7.5
5	1	51.7	58.9	7.2
6	1	92.7	98.9	6.2
7	1	91.0	95.4	4.4
8	1	83.9	93.0	9.1
9	1	80.7	87.9	7.2
10	1	99.7	104.3	4.6
1	2	56.3	60.6	4.3
2	2	80.5	87.4	6.9
3	2	94.5	104.4	9.9
4	2	90.3	94.6	4.3
5	2	60.0	67.2	7.2
6	2	87.2	93.5	6.3
7	2	93.4	103.3	9.9
8	2	78.8	85.8	7.0
9	2	80.3	88.1	7.8
10	2	103.2	111.5	8.3

The average of the ranges, $\overline{X}_{range}$, is 7.015. From Appendix A, using $Z = 1$ and $W = 3$ for 3 appraisers, is 1.91. The reproducibility is

$$\text{Reproducibility} = \sqrt{\left(\frac{5.15(7.015)}{1.91}\right)^2 - \frac{23.7^2}{10(2)}} = 18.2$$

The repeatability and reproducibility is

$$R\&R = \sqrt{23.7^2 + 18.2^2} = 29.9$$

The part variability is computed using the difference between the largest and smallest part measurement, where the average is taken for all parts and appraisers. This data is shown in Table 4.

Table 4. Example part variability computations.

	Operator						
	A		B		C		
Part	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2	Avg
1	65.2	60.1	62.9	56.3	71.6	60.6	62.78
2	85.8	86.3	85.7	80.5	92.0	87.4	86.28
3	100.2	94.8	100.1	94.5	107.3	104.4	100.22
4	85.0	95.1	84.8	90.3	92.3	94.6	90.35
5	54.7	65.8	51.7	60.0	58.9	67.2	59.72
6	98.7	90.2	92.7	87.2	98.9	93.5	93.53
7	94.5	94.5	91.0	93.4	95.4	103.3	95.35
8	87.2	82.4	83.9	78.8	93.0	85.8	85.18
9	82.4	82.2	80.7	80.3	87.9	88.1	83.60
10	100.2	104.9	99.7	103.2	104.3	111.5	103.97

The part with the largest average belongs to part 10, 103.97. The lowest average belongs to part 5, 59.72. This difference , 44.25, is V_p. From Appendix A, using $Z = 1$ and $W = 10$ for 10 parts, $d_2 = 3.18$. The part variability is

$$V_P = \frac{5.15(44.25)}{3.18} = 71.7$$

The total measurement system variability is

$$V_T = \sqrt{29.9^2 + 71.7^2} = 77.7$$

Analysis of Variance Method

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 analyze 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 in Table 5.

Table 5. Two-Way ANOVA Table.

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

$$SSA = \sum_{i=1}^{a} \frac{(Y_{i..})^2}{bn} - \frac{Y_{..}^2}{N} \qquad\qquad 6$$

$$SSB = \sum_{j=1}^{b} \frac{(Y_{.j.})^2}{an} - \frac{Y_{..}^2}{N} \qquad\qquad 7$$

$$SSAB = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(Y_{ij.})^2}{n} - \frac{Y_{..}^2}{N} - SSA - SSB \qquad\qquad 8$$

$$TSS = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk}^2 - \frac{Y_{..}^2}{N} \qquad\qquad 9$$

$$SSE = TSS - SSA - SSB - SSAB \qquad\qquad 10$$

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

$$Repeatability = 5.15\sqrt{MSE} \qquad\qquad 11$$

The measurement system reproducibility is

$$Reproducibility = 5.15\sqrt{\frac{MSA - MSAB}{bn}} \qquad\qquad 12$$

The interaction between the appraisers and the parts is

$$I = 5.15\sqrt{\frac{MSAB - MSE}{n}} \qquad\qquad 13$$

The measurement system repeatability and repeatability is

$$R\&R = \sqrt{\text{Repeatability}^2 + \text{Reproducibility}^2 + I^2} \qquad\qquad 14$$

The measurement system part variation is

$$V_P = 5.15\sqrt{\frac{MSB - MSAB}{an}} \qquad\qquad 15$$

The total measurement system variation is

$$V_T = \sqrt{R\&R^2 + V_P^2} \qquad\qquad 16$$

Example 2
The thickness, in millimeters, 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 Table 6.

Table 6. ANOVA method example data.

	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

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 Table 7 along with the square of this sum, and the square of this sum divided by 6.

Table 7. Part sum-of-squares computations.

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 Table 8 along with the square of this sum, and the square of this sum divided by 2.

Table 8. Interaction sum-of-square computations.

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

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

The sum-of-squares for the gage or error is

$$SSE = 12,630.4 - 502.5 - 11,545.5 - 35.6 = 546.8$$

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 Table 9.

Table 9. Example ANOVA table.

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 (Appraisers, Parts)	35.6	18	1.98	0.11	0.999996
Gage (Error)	546.8	30	18.2		
Total	12,630.4	59	214.1		

The significance listed in Table 9 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 using the function FDIST(F-statistic,d_1,d_2) in Microsoft Excel or Lotus 123, where d_1 and d_2 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

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

Appendix A - Values of d$_2$

z	w													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.41	1.91	2.24	2.48	2.67	2.83	2.96	3.08	3.18	3.27	3.35	3.42	3.49	3.55
2	1.28	1.81	2.15	2.40	2.60	2.77	2.91	3.02	3.13	3.22	3.30	3.38	3.45	3.51
3	1.23	1.77	2.12	2.38	2.58	2.75	2.89	3.01	3.11	3.21	3.29	3.37	3.43	3.50
4	1.21	1.75	2.11	2.37	2.57	2.74	2.88	3.00	3.10	3.20	3.28	3.36	3.43	3.49
5	1.19	1.74	2.10	2.36	2.56	2.78	2.87	2.99	3.10	3.19	3.28	3.36	3.42	3.49
6	1.18	1.73	2.09	2.35	2.56	2.73	2.87	2.99	3.10	3.19	3.27	3.35	3.42	3.49
7	1.17	1.73	2.09	2.35	2.55	2.72	2.87	2.99	3.10	3.19	3.27	3.35	3.42	3.48
8	1.17	1.72	2.08	2.35	2.55	2.72	2.87	2.98	3.09	3.19	3.27	3.35	3.42	3.48
9	1.16	1.72	2.08	2.34	2.55	2.72	2.86	2.98	3.09	3.19	3.27	3.35	3.42	3.48
10	1.16	1.72	2.08	2.34	2.55	2.72	2.86	2.98	3.09	3.18	3.27	3.34	3.42	3.48
11	1.15	1.71	2.08	2.34	2.55	2.72	2.86	2.98	3.09	3.18	3.27	3.34	3.41	3.48
12	1.15	1.71	2.07	2.34	2.55	2.72	2.85	2.98	3.09	3.18	3.27	3.34	3.41	3.48
13	1.15	1.71	2.07	2.34	2.55	2.71	2.85	2.98	3.09	3.18	3.27	3.34	3.41	3.48
14	1.15	1.71	2.07	2.34	2.54	2.71	2.85	2.98	3.09	3.18	3.27	3.34	3.41	3.48
15	1.15	1.71	2.07	2.34	2.54	2.71	2.85	2.98	3.08	3.18	3.26	3.34	3.41	3.48
>15	1.128	1.693	2.059	2.326	2.534	2.704	2.847	2.97	3.078	3.173	3.258	3.336	3.407	3.472