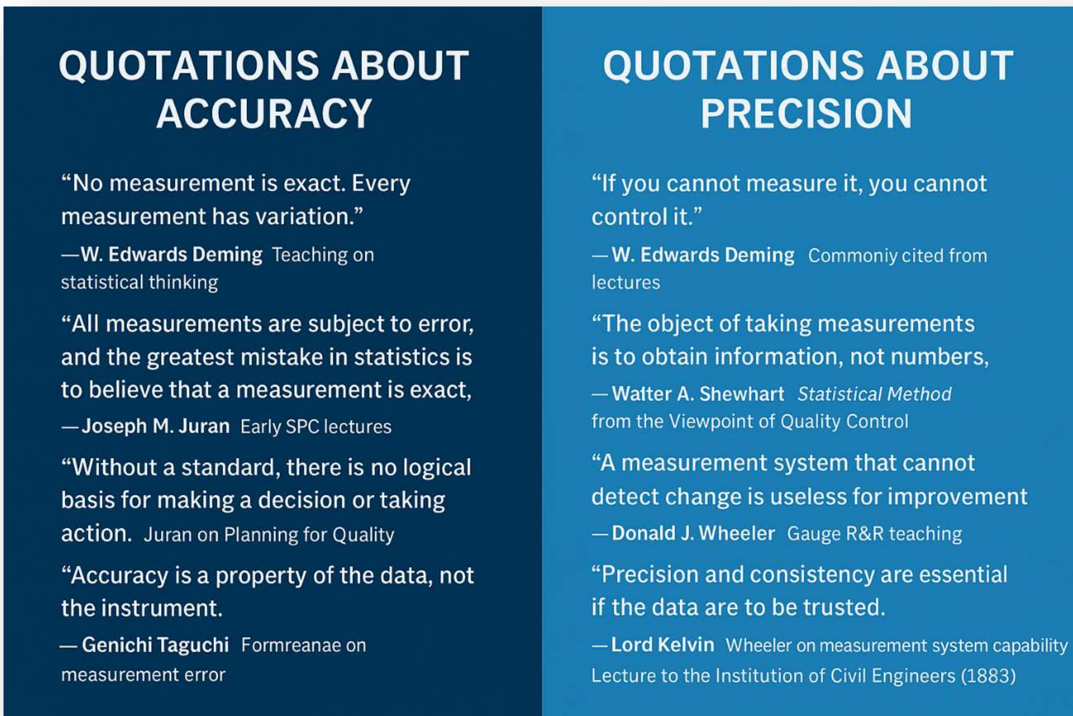
Understanding and Conducting Measurement System Analysis (MSA)

Dr David Scrimshire FRAeS, FCME, BSc (1st Hons), PhD

Managing Director – TEC Transnational Ltd



TEC Transnational Ltd

https://tectransnational.com/

Introduction

Just as the processes that create products may vary, the process of obtaining measurements and data also have variation and may produce incorrect results.

"*No measurement is exact. Every measurement has variation*" – Deming

Measurement System Analysis (**MSA**) is a statistical method used to assessing the quality of a measurement system for making quality control decisions, by determining the amount of variation in the data it collects. It helps distinguish between variation caused by the actual process and variation caused by the measurement system itself (i.e. gauge, operator, procedure). By ensuring data is accurate and consistent, MSA prevents wasted effort on fixing the wrong problem and builds confidence in data-driven improvements.

"*All measurements are subject to error, and the greatest mistake in statistics is to believe that a measurement is exact*" – Shewhart

MSA should be conducted as part of New Product Introduction (NPI) to validate the measurement system prior to production. Other situations where MSA should be repeated, include: changes to gauge design, refurbishment/repair, environment, product design change to the feature being measured, etc. In any event it increasingly becoming a customer-specific requirement!

"*Inadequate metrology leads to wrong engineering decisions*" – Taguchi

So, let's get started by identifying the product **characteristic** to be measured and selecting the **gauge** we wish to use .. .. ..

Characteristic identification

The ***type*** of **characteristic** to be measured must be identified and its associate '***requirements***' confirmed –

- **product control**: the ***criticality*** and ***specification*** (i.e. nominal, LSL and USL) used to determine conformance or nonconformance – assessment of the measurement system based on the **tolerance** [%GRR to TOLERANCE]

- **process control**: the ***control specifications*** (i.e. LCL and UCL) for SPC (e.g. X_{bar} & R Chart) – assessment of the measurement system based on **process variation** [%GRR to PROCESS VARIATION]

Also, the **characteristic *classification*** must be established (e.g. Key Characteristic, Critical, Major or Minor) – usually specified by the customer or the design authority.

Gauge selection

When selecting a gauge to measure a product characteristic, you must first consider compatibility with the part geometry and resolution (discrimination).

part geometry –

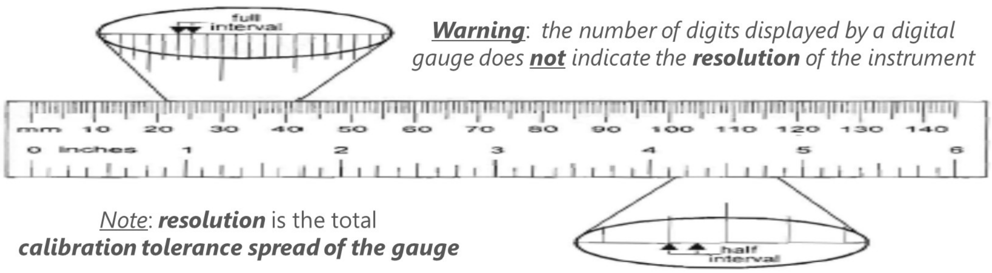
- Size and shape of the part: *Ensure the gauge can physically access the feature*
- Surface condition: *Rough or reflective surfaces may require non-contact gauging like optical or air gauges*
- Internal vs. external features: *Bore gauges for internal diameters, calipers or micrometres for external ones*

resolution –

Tolerance of the feature: *Choose a gauge that is at least 10 times more precise than the tolerance of the feature being measured (10:1 rule)*

Type of measurement: *Decide whether you need to measure dimensions (length, diameter), form (flatness, roundness), or surface finish*

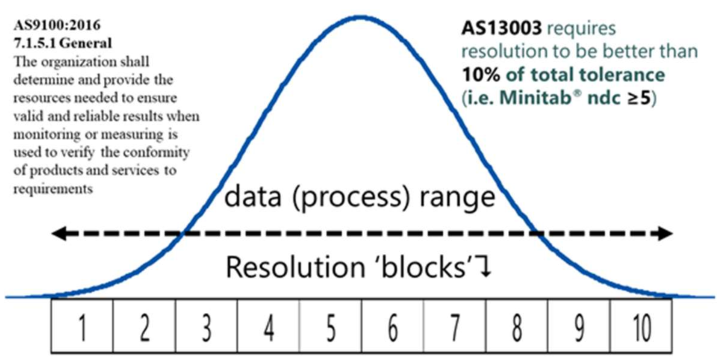
The ***resolution*** (***discrimination***) of a gauge is its ability to detect and faithfully indicate small changes in a product characteristic.



Typically, **resolution** is the value of the ***smallest* graduation** on the scale of the gauge (also known as '***gauge tolerance***').

If the gauge has 'coarse' graduations, then a half-graduation can be used (also known as '***interpolation***').

Related standards (e.g. **AS13003**) typically require resolution to be better than 10% of the product specification (i.e. total tolerance: USL – LSL).



Strictly, this 10 to 1 'rule-of-thumb' should be defined as one-tenth of the range to be measured. For SPC purpose this would be one-tenth of the process control limits (i.e. range: UCL – LCL).

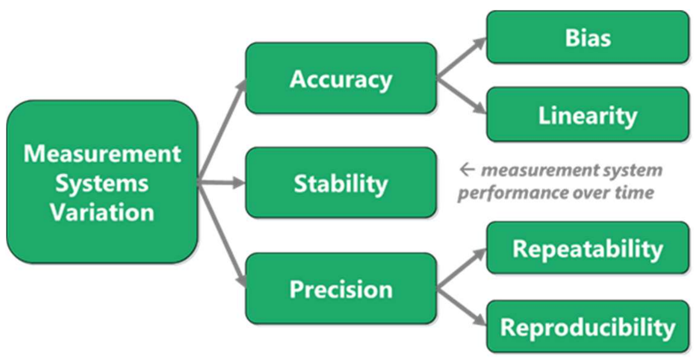
Note: **Minitab®** refers to number of distinct categories – same concept, but different calculations (the Minitab® criteria for acceptability is **ndc ≥5** – not **10**)

Use of a statistical app

Calculations and graphics are best undertaken using a suitable **statistical *app*** – here we have adopted **Minitab®** that is a recognized industry standard, and its results are acceptable to most customers in the automotive and aerospace sectors.

Types of MSA studies

In manufacturing and assembly operations, **variability** within both the process and product is inevitable. All measurements systems are also subject to **variability**, which can be ***categorized*** as –



Accuracy = Linearity + Bias

Linearity – a measure of how the size of the part affects the accuracy of the measurement system (i.e. the difference in the observed accuracy values through the expected range of measurements)

Bias – a measure of the impartiality in the measurement system (i.e. the difference between the observed average part measurement and a master value – oversize *or* undersize)

Precision = **Repeatability** + **Reproducibility**

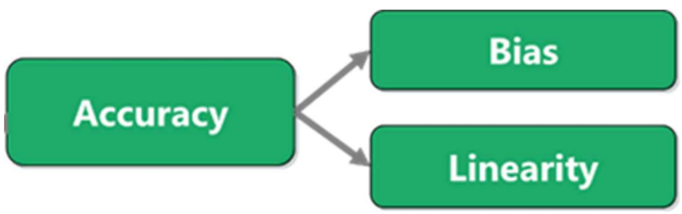
Repeatability – the variation due to the measuring device (the Gauge) i.e. the variation observed when the same operator (the Appraiser) measures the same part repeatedly with the same device.

Reproducibility – the variation due to the measurement system i.e. the variation observed when different operators (the Appraisers) measure the same parts using the same device and same work instruction.

Stability – a measure of how accurately the system performs over time i.e. the total variation obtained with a particular device, on the same part, when measuring a single characteristic over time (e.g. minutes, hours) – environment/human factors?

Accuracy (Linearity & Bias) study

To determine the **accuracy** (**linearity** + **bias**) of a measurement system, first select five parts and measure each; by layout inspection (CMM); to determine their ***reference*** values.



Case Study

In this Case Study, the 'parts' were represented by calibrated **gauge blocks** of known values: 2.00, 4.00, 6.00, 8.00 and 10.00 respectively.

A single **operator** then measured each 'part' **12-times** using a Micrometer – total of **60 measurements** – using the ***same* gauge**.

These are the results as recorded …

Five 'reference' parts measured 12-times each
same operator ~ same gauge

Part	**1**	**2**	**3**	**4**	**5**
Reference value	**2.00**	**4.00**	**6.00**	**8.00**	**10.00**
Trails	2.7	5.10	5.80	7.60	9.10
	2.5	3.90	5.70	7.70	9.30
	2.4	4.20	5.90	7.80	9.50
	2.5	5.00	5.90	7.70	9.30
	2.7	3.80	6.00	7.80	9.40
	2.3	3.90	6.10	7.80	9.50
	2.5	3.90	6.00	7.80	9.50
	2.5	3.90	6.10	7.70	9.50
	2.4	3.90	6.40	7.80	9.60
	2.4	4.00	6.30	7.50	9.20
	2.6	4.10	6.00	7.60	9.30
	2.4	3.80	6.10	7.70	9.40

Study results

Using the data, a Gauge Linearity and Bias Study was performed using Minitab®–

Choose Stat > Quality Tools > Gage Study > Gage Linearity and Bias Study

in Part numbers, enter Part Number

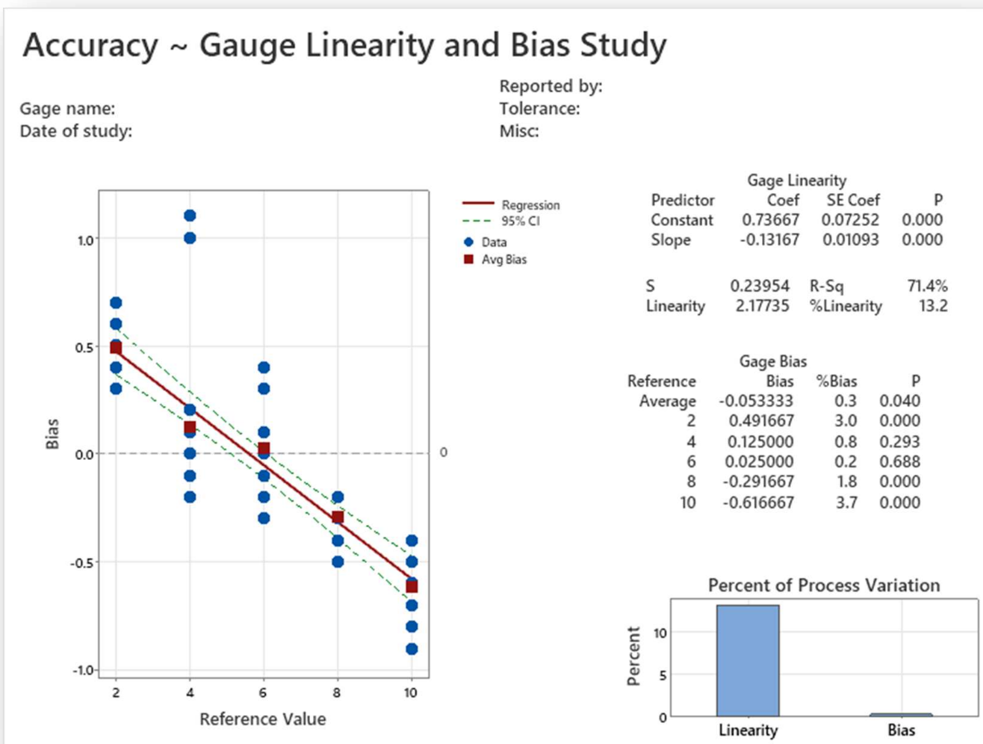
in Reference values, enter Master Value

in Measurement data, enter Measurement Value

in Process variation, optionally enter 16.5368 (optional)

Click OK .. .. ..

The following analysis is displayed –



The **blue circles** represent the ***difference*** between a part's reference value and the operator's measurements of the part, and the **red squares** represent the *average* values. A **red line** is fitted through these values, using ordinary least squares regression.

Ideally, the plotted **red line** should be close to the horizontal, indicating that the average bias is relatively constant and does not depend on the reference value.

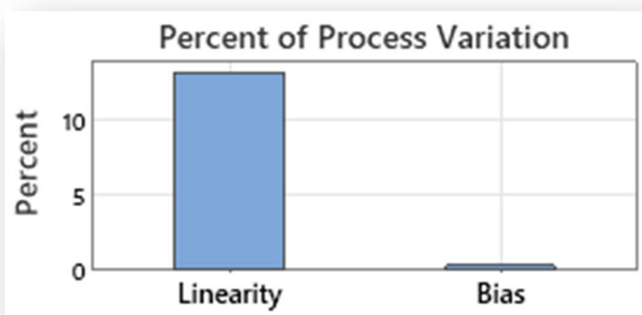
When the **red line** is close to 0 (zero), the difference between the observed average measurement and the reference value is very small, which also indicates that the system does not contain significant bias.

Linearity assesses the difference in average bias through the expected operating range of the measurement system.

In this example, measurements for smaller parts are higher than their corresponding reference part values. Measurements for larger parts tend to be lower than their corresponding reference part values. The **%Linearity** (absolute value of the slope x 100) is **13.2** – therefore, **linearity** accounts for approximately **13%** of the overall process variation.

Bias is the *difference* between a part's reference value and the operator's measurements of the part.

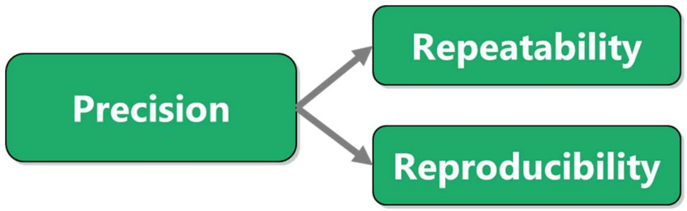
Here, the **%Bias** (over/under reading) for the reference average is only **0.3**, which means that the gauge bias accounts for less than **0.3%** of the overall process variation –



.Consequently, **Linearity** is the predominant problem!

Precision (Repeatability & Reproducibility) study

To determine the **precision** of a measurement system, a selection of products are measured several times – with different people (appraisers) – using the same gauge.

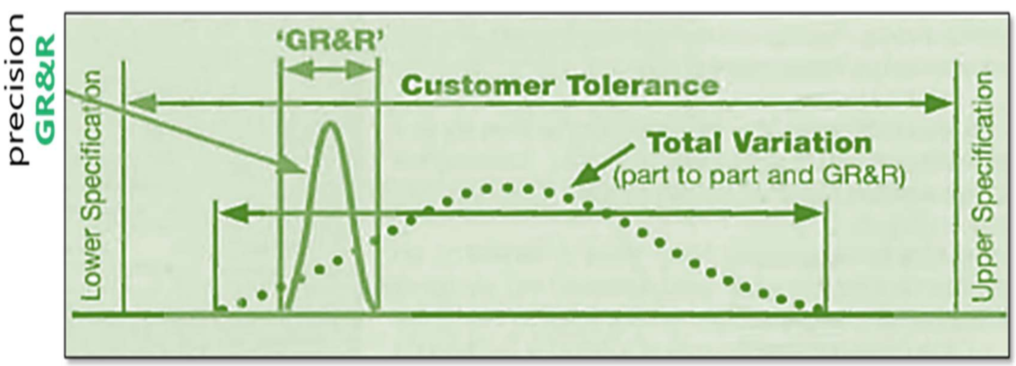


As the components are not changing – any variation in results must represent the ***Repeatability*** of the gauge (equipment) and the ***Reproducibility*** of measurements by different people (appraisers).

A **GR&R Study *repeats*** this approach on several parts to assess the results in a single study.

The following schematic illustrates the **GR&R *acceptability criteria*** –

GR&R as a percentage of the **Total Variation**
= GR&R ÷ Total Variation x 100%



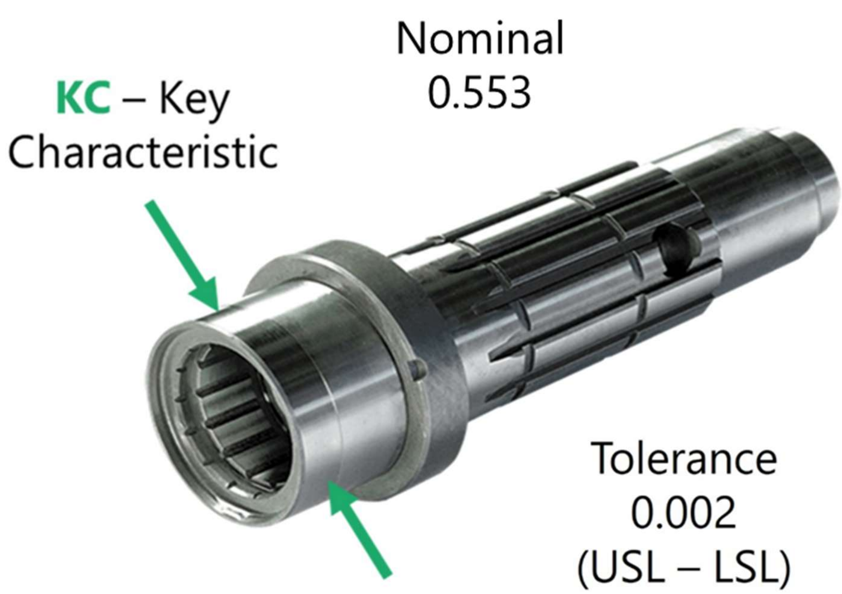
GR&R as a percentage of the **Customer Tolerance**
= GR&R ÷ Customer Tolerance x 100%

Case Study

Choosing **10 parts**, **3 operators** and measuring **3 times** is a common **GR&R Study** 'standard practice'. Always ensure that the operators (appraisers) measure the parts under 'typical' conditions and in *random order* –

1. Obtain a sample of **<u>10</u>** parts that represent the *full operating range* of the process
2. Select the operators (**A**, **B**, **C**, ...) and number the parts **1** through **10** so that the numbers are **<u>not</u> visible** to the appraisers
3. Identify and calibrate the gauge!
4. Have the appraisers measure the **<u>10</u>** parts **separately** and in a **random order** – ***trial #1***
5. Repeat the cycle using a different random order of measurement – ***trial #2***
6. Repeat the cycle again using a ***different*** **random order** of measurement – ***trial #3***

This Case Study is a 'real one' conducted during a typical **TEC** class workshop. The focus is the measurement of a shaft diameter designated as a **Key Characteristic** (**KC**) by the customer –



The diameter has a **nominal** of **0.553** and a **tolerance** of **0.002** (USL – LSL).

The results as recorded –

RunOrder	Parts	Operators	Measurement	RunOrder	Parts	Operators	Measurement	RunOrder	Parts	Operators	Measurement
1	9	Daren	0.55380	31	6	Daren	0.55395	61	2	Daren	0.55295
2	7	Daren	0.55360	32	2	Daren	0.55300	62	10	Daren	0.55380
3	6	Daren	0.55400	33	8	Daren	0.55370	63	5	Daren	0.55330
4	4	Daren	0.55345	34	10	Daren	0.55380	64	7	Daren	0.55355
5	1	Daren	0.55295	35	9	Daren	0.55380	65	8	Daren	0.55370
6	2	Daren	0.55305	36	3	Daren	0.55370	66	9	Daren	0.55380
7	3	Daren	0.55380	37	5	Daren	0.55325	67	6	Daren	0.55395
8	5	Daren	0.55335	38	4	Daren	0.55345	68	3	Daren	0.55365
9	8	Daren	0.55375	39	7	Daren	0.55355	69	4	Daren	0.55340
10	10	Daren	0.55385	40	1	Daren	0.55290	70	1	Daren	0.55285
11	5	Tim	0.55325	41	7	Tim	0.55355	71	9	Tim	0.55380
12	9	Tim	0.55385	42	8	Tim	0.55370	72	5	Tim	0.55325
13	1	Tim	0.55290	43	10	Tim	0.55380	73	4	Tim	0.55340
14	4	Tim	0.55340	44	3	Tim	0.55370	74	6	Tim	0.55395
15	3	Tim	0.55375	45	1	Tim	0.55285	75	8	Tim	0.55370
16	10	Tim	0.55385	46	5	Tim	0.55320	76	3	Tim	0.55365
17	8	Tim	0.55375	47	9	Tim	0.55380	77	1	Tim	0.55290
18	2	Tim	0.55300	48	4	Tim	0.55335	78	7	Tim	0.55355
19	7	Tim	0.55355	49	2	Tim	0.55295	79	2	Tim	0.55295
20	6	Tim	0.55395	50	6	Tim	0.55395	80	10	Tim	0.55380
21	3	Gurpreet	0.55370	51	4	Gurpreet	0.55335	81	7	Gurpreet	0.55360
22	9	Gurpreet	0.55375	52	10	Gurpreet	0.55375	82	6	Gurpreet	0.55400
23	2	Gurpreet	0.55295	53	2	Gurpreet	0.55290	83	8	Gurpreet	0.55375
24	4	Gurpreet	0.55335	54	8	Gurpreet	0.55365	84	1	Gurpreet	0.55290
25	5	Gurpreet	0.55330	55	6	Gurpreet	0.55395	85	5	Gurpreet	0.55335
26	6	Gurpreet	0.55400	56	5	Gurpreet	0.55330	86	4	Gurpreet	0.55340
27	10	Gurpreet	0.55380	57	7	Gurpreet	0.55355	87	2	Gurpreet	0.55300
28	7	Gurpreet	0.55355	58	3	Gurpreet	0.55370	88	3	Gurpreet	0.55370
29	8	Gurpreet	0.55370	59	9	Gurpreet	0.55375	89	9	Gurpreet	0.55385
30	1	Gurpreet	0.55285	60	1	Gurpreet	0.55285	90	10	Gurpreet	0.55385

Study results

Using Minitab®, a Gauge Linearity and Bias Study was performed using Minitab®–

Choose Stat > Quality Tools > Gage Study > Gage R&R Study (Crossed)

in Part numbers, enter Parts

in Operators, enter Operators

in Measurement data, enter Measurement

in Method of Analysis, select ANOVA

Click OK .. .. ..

The following analysis is displayed –

The Two-Way **ANOVA** Table includes terms for **parts**, **operators**, and the **parts-operators** interaction –

Two-Way ANOVA Table With Interaction

Source	DF	SS	MS	F	P
Parts	9	0.0000109	0.0000012	964.236	0.000
Operators	2	0.0000000	0.0000000	2.941	0.078
Parts * Operators	18	0.0000000	0.0000000	0.923	0.555
Repeatability	60	0.0000001	0.0000000		
Total	89	0.0000110			

α to remove interaction term = 0.05

If the **p-value** for the **Parts*Operators** interaction is **≥0.05**, **Minitab®** omits the interaction from the full model because it is not significant.

In the Case Study, the p-value is **0.555**, so Minitab generates a second two-way ANOVA table that omits the interaction from the final model.

Two-Way ANOVA Table Without Interaction

Source	DF	SS	MS	F	P
Parts	9	0.0000109	0.0000012	906.015	0.000
Operators	2	0.0000000	0.0000000	2.763	0.069
Repeatability	78	0.0000001	0.0000000		
Total	89	0.0000110			

The **Variance Components** (**VarComp**) is used to compare the variation from each source of measurement error to the total variation –

Variance Components

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.0000000	1.04
Repeatability	0.0000000	0.98
Reproducibility	0.0000000	0.06
Operators	0.0000000	0.06
Part-To-Part	0.0000001	98.96
Total Variation	0.0000001	100.00

In the **Case Study**, the **%Contribution** (of VarComp) column in the Gauge R&R table shows that the variation from **Part-To-Part** is **98.96%**.

This value is much larger than **Total Gauge R&R**, which is just **1.04%** consequently, much of the variation is due to ***differences*** between **parts**.

The **Gauge Evaluation** is used to compare the measurement system variation (**%Study Var**) to the total variation.

Gage Evaluation

Source	StdDev (SD)	Study Var (6 × SD)	%Study Var (%SV)	%Tolerance (SV/Toler)
Total Gage R&R	0.0000376	0.0002257	10.21	5.64
Repeatability	0.0000366	0.0002194	9.92	5.48
Reproducibility	0.0000089	0.0000532	2.41	1.33
Operators	0.0000089	0.0000532	2.41	1.33
Part-To-Part	0.0003667	0.0021999	99.48	55.00
Total Variation	0.0003686	0.0022115	100.00	55.29

Number of Distinct Categories = 13

The **Total Gauge R&R** equals **10.21%** of the study variation. This **Total Gage R&R** percentage contribution is very likely to be acceptable depending on the application.

For this data, the **Number of Distinct Categories** is **13**. According to the AIAG, you need at least **5** distinct categories to have an adequate measuring system – so this is again okay!

The same data is also quoted in terms of a percentage of tolerance (**%Tolerance**) which is **5.64%** – likely to be of greater importance to the customer.

Individual values of **Repeatability** (**9.92%** and **5.48%** respectively) and **Reproducibility**/**Operators** (**2.41%** and **1.33%** respectively).

Okay, but are these results acceptable?

We have three options –

- **percentage of variation**
- **percentage of tolerance**
- **Number of Distinct Categories**

Requirements for GR&R acceptability

The acceptability of a measurement system is usually expressed in terms of the percentage of a characteristic's tolerance (USL – LSL) and default to sector-specific limits or as specified by the customer concerned.

Automotive (AIAG)

Acceptance limits for characteristics identified in the control plan –

<10%: Measurement system is generally considered acceptable

10% - 30%: May be acceptable depending on the application, tolerance, and risk

>30%: Unacceptable — the system needs improvement or replacement

Aero-Engine (AESQ) & AS13003

Separate acceptance limits for different key characteristic categories are given for both repeatability and GR&R –

Critical	Major	Minor
≤10% of tolerance	≤20% of tolerance	≤30% of tolerance

Returning to our Case Study results, for both the Automotive and Aero-Engine sectors –

Total Gauge R&R of **10.21%** (**study**) is marginally above the "≤10% requirement" and **5.64%** (**tolerance**) is well below the "≤10% requirement".

The measurement system is very likely to be deemed acceptable
– subject to approval by the customer concerned

Other Minitab® Graphs

Other graphs also provide the useful information about the measurement system:

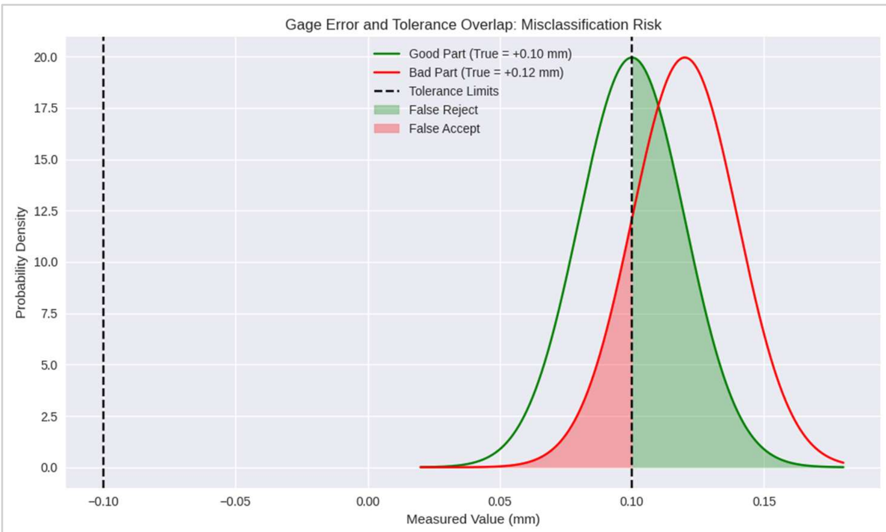
- **Components of Variation** graph displays the %Contribution from Part-To-Part is larger than that of Total Gauge R&R. Thus, much of the variation is due to differences between parts
- **R Chart by Operators** shows that Daren measures parts more consistently
- **X_{bar} Chart by Operators** shows that the majority of the points are ***outside*** the control limits – thus, much of the variation is due to differences between parts
- **Measurement by Part** shows that the differences between parts are large
- **Measurement by Operator** graph, the differences between operators are small compared to the differences between parts. Daren's measurements are slightly higher than the measurements of the other operators and Tim's shows slightly larger variation
- **Operator * Part Interaction** lines are ***parallel*** and the **p-value** for the Operator*Part interaction found in the table is **0.555**. These results indicate that no significant interaction between each Part and Operator exists.

Probabilities of misclassification

In any measurement system there are always risks of **misclassification** because of the gauge variation consequently, the measured value of the part does not always equal the true value of the part.

This discrepancy between the measured value and the actual value creates the potential for 'misclassifying' the part.

Consider this visualization shows how **gauge error** overlaps with **tolerance limits** (in this example it's the **USL**) and where misclassification risk arises .. .. ..



- Green curve: Distribution of measured values for a good part at +0.10 mm (exactly at the upper tolerance limit).
- Red curve: Distribution of measured values for a bad part at +0.12 mm (just outside tolerance).
- Vertical dashed lines: Engineering tolerance limits (±0.10 mm).
- Shaded regions:
  - False Reject (green area): Good parts measured above +0.10 mm.
  - False Accept (red area): Bad parts measured below +0.10 mm.

In this example .. .. ..

- False Accepts: About **16%** of **bad parts** *just outside tolerance* may slip through
- False Rejects: About **50%** of *borderline* **good parts** risk being rejected
- The AIAG Gauge R&R would only say ">30% variation is <u>not</u> acceptable"

Probabilities of misclassification with Minitab®

Because we have the specification limits **0.551** – **0.555** mm quoted for our Case Study, Minitab® calculates the probabilities of ***misclassifying*** product.

Minitab® calculates both the **joint probabilities** and the **conditional probabilities** of misclassification .. .. ..

Probabilities of Misclassification

Joint Probability

Description	Probability
A randomly selected part is bad but accepted	0.000
A randomly selected part is good but rejected	0.000

Conditional Probability

Description	Probability
A part from a group of bad products is accepted	0.138
A part from a group of good products is rejected	0.000

Joint probability

Use the joint probability when prior knowledge about the acceptability of the parts is unknown. For example, when sampling from the line and it is not known whether each particular part is good or bad.

There are two misclassifications that can be made –

- The probability that the part is **bad**, and you ***accept*** it (**0.000** = **zero%**)
- The probability that the part is **good**, and you ***reject*** it (**0.000** = **zero%**)

Conditional probability

Use the conditional probability when prior knowledge about the acceptability of the parts is known. For example, when sampling from a pile of rework or from a pile of product that will soon be shipped as good.

There are two misclassifications that can be me made –

- The probability that you ***accept*** a part that was sampled from a pile of **bad** product that needs to be reworked – also called false accept (**0.138** = **13.8%**)
- The probability that you ***reject*** a part that was sampled from a pile of **good** product that is about to be shipped – also called false reject (**0.000** = **zero%**)

What are acceptable figures?

Acceptable thresholds for both types of 'misclassification' are not formally quoted by Automotive (AIAG), Aero-Engine (AESQ) or Wheeler. But for the **joint probability** figure general guidelines for include –

< 0.01 (**1%**) is ideal for critical applications (e.g. automotive, aero-engine, aerospace)

< 0.05 (**5%**) is acceptable for most manufacturing environments

> 0.10 (**10%**) may indicate a poor measurement system needing improvement

For this Case Study: considering these **probabilities of misclassification *figures*** and the previous **GR&R results**, we can say that the measurement system would be **acceptable** by both Automotive (AIAG) and Aero-Engine (AESQ) customers.

Unsatisfactory GR&R results

Aero-engine example

In the Case Study shown in AESQ's Reference Manual: **RM13003** (5.19.3), an aero engine organization is manufacturing machined structures. An inspection device is used to determine a **critical feature** on one of the parts. To evaluate the measurement system and determine if it is fit for its intended purpose an MSA (GR&R) is conducted.

The critical feature is an outer diameter with specification limits **838.60** - **838.80** mm (total tolerance = **0.2** mm). The inspection device is a dial gauge comparator together with a master gauge.

Ten parts were selected that represent the expected range of the process variation. Three operators measured the ten parts, three times per part, in a random order without seeing each other's readings.

Here is the data taken from **RM13003** (5.19.3) –

	Operator A			***Operator B***			***Operator C***		
Part	***Trial 1***	***Trial 2***	***Trial 3***	***Trial 1***	***Trial 2***	***Trial 3***	***Trial 1***	***Trial 2***	***Trial 3***
1	838.79	838.77	838.80	838.78	838.77	838.79	838.78	838.80	838.79
2	838.69	838.68	838.70	838.69	838.70	838.72	838.72	838.69	838.73
3	838.72	838.69	838.71	838.70	838.71	838.73	838.72	838.74	838.71
4	838.75	838.74	838.73	838.73	838.75	838.73	838.76	838.76	838.72
5	838.73	838.72	838.70	838.71	838.73	838.72	838.73	838.73	838.75
6	838.77	838.79	838.79	838.77	838.79	838.77	838.78	838.79	838.78
7	838.67	838.68	838.69	838.70	838.69	838.66	838.68	838.67	838.70
8	838.60	838.61	838.62	838.61	838.64	838.60	838.62	838.60	838.62
9	838.63	838.65	838.66	838.66	838.63	838.65	838.66	838.65	838.64
10	838.78	838.78	838.77	838.77	838.78	838.75	838.77	838.75	838.76

The **Sigma Value** was set at **6.00**.

The **Gauge Evaluation** is used to compare the measurement system variation (**%Study Var**) to the total variation and as a percentage of tolerance (**%Tolerance**) –

Gage Evaluation

Source	StdDev (SD)	Study Var (6 × SD)	%Study Var (%SV)	%Tolerance (SV/Toler)
Total Gage R&R	0.0135053	0.081032	23.22	40.52
Repeatability	0.0132723	0.079634	22.82	39.82
Reproducibility	0.0024979	0.014987	4.29	7.49
Operator	0.0024979	0.014987	4.29	7.49
Part-To-Part	0.0565803	0.339482	97.27	169.74
Total Variation	0.0581698	0.349019	100.00	174.51

In this example, the Gauge R&R **%Tolerance** figure was **40.52%**, that is greater than the permitted maximum of **10%** for a **critical feature**.

The measurement system is therefore deemed to be **unacceptable**

As we have the specification limits **838.6** – **838.8** quoted for this Case Study, Minitab® calculates the probabilities of ***misclassifying*** product.

Minitab® calculates both the **joint probabilities** and the **conditional probabilities** of misclassification .. .. ..

Joint Probability

Description	Probability
A randomly selected part is bad but accepted	0.014
A randomly selected part is good but rejected	0.022

Conditional Probability

Description	Probability
A part from a group of bad products is accepted	0.153
A part from a group of good products is rejected	0.024

Automotive example

In the Case Study shown in IAQG's Reference Manual: **MEASUREMENT SYSTEMS ANALYSIS** (4th Edition), ten parts were selected that represent the expected range of the process variation. Three operators measured the ten parts, three times per part, in a random order without seeing each other's readings.

Here is the Data Collection Sheet –

Appraiser /Trial #		PART									
		1	2	3	4	5	6	7	8	9	10
A	1	0.29	-0.56	1.34	0.47	-0.80	0.02	0.59	-0.31	2.26	-1.36
	2	0.41	-0.68	1.17	0.50	-0.92	-0.11	0.75	-0.20	1.99	-1.25
	3	0.64	-0.58	1.27	0.64	-0.84	-0.21	0.66	-0.17	2.01	-1.31
	Average										
	Range										
B	1	0.08	-0.47	1.19	0.01	-0.56	-0.20	0.47	-0.63	1.80	-1.68
	2	0.25	-1.22	0.94	1.03	-1.20	0.22	0.55	0.08	2.12	-1.62
	3	0.07	-0.68	1.34	0.20	-1.28	0.06	0.83	-0.34	2.19	-1.50
	Average										
	Range										
C	1	0.04	-1.38	0.88	0.14	-1.46	-0.29	0.02	-0.46	1.77	-1.49
	2	-0.11	-1.13	1.09	0.20	-1.07	-0.67	0.01	-0.56	1.45	-1.77
	3	-0.15	-0.96	0.67	0.11	-1.45	-0.49	0.21	-0.49	1.87	-2.16
	Average										
	Range										
Part Average											

To match the AIAG calculations, the **Sigma Value** was set at **5.15**.

The **Gauge Evaluation** is used to compare the measurement system variation (**%Study Var**) to the total variation and as a percentage of tolerance (**%Tolerance**) –

Gage Evaluation

Source	StdDev (SD)	Study Var (5.15 × SD)	%Study Var (%SV)
Total Gage R&R	0.30237	1.55721	27.86
Repeatability	0.19993	1.02966	18.42
Reproducibility	0.22684	1.16821	20.90
Operator	0.22684	1.16821	20.90
Part-To-Part	1.04233	5.36799	96.04
Total Variation	1.08530	5.58929	100.00

In this example, the Gauge R&R **%Study Var** figure was **27.86%**, that is just within the 10% - 30% band (and may be acceptable depending on the application, tolerance, and risk).

The measurement system ***may*** therefore be deemed to be **acceptable**

Because we have <u>not</u> provided the **specification limits**, Minitab® cannot calculate the probabilities of ***misclassifying*** product.

Dealing with unacceptable GR&R results

If the results of a GR&R Study yield **unacceptable** results, sources of variation in the measurement system must identified and reduced – starting with the gauge, operators, and procedures.

A structured approach is recommended .. .. ..

---

1. Analyze the GR&R Results Thoroughly

- **Use the ANOVA table** to pinpoint significant sources of variation – whether it is the part, operator, or interaction between them

- **Review graphs and variance components** to visualize where inconsistencies arise (e.g. operator bias or inconsistent readings)

---

2. Troubleshoot Key Components

- **Gauge/Instrument Issues**:
  - Check calibration and maintenance records
  - Replace or upgrade the gauge if it is inherently imprecise
  - Ensure the gauge is suitable for the tolerance range of the parts being measured
- **Operator Inconsistency**:
  - Standardize procedures to reduce subjective judgment
  - Provide training to ensure consistent measurement techniques
  - Limit the number of operators if variability is high
- **Measurement Procedure**:
  - Clarify instructions and reduce ambiguity
  - Automate measurements where possible to reduce human error
  - Ensure environmental conditions (temperature, lighting, vibration) are stable and controlled

---

3. Refine the Study Design

- Increase the number of parts, trials, or operators to get more reliable data
- Ensure parts used in the study represent the full range of variation in production –

  "You need some 'BAD' parts to conduct a 'GOOD' Gauge R&R Study!"
- Randomize the order of measurements to reduce bias

---

4. Repeat the GR&R Study

- After implementing improvements, re-run the GR&R study to validate changes

- Aim for a %GRR (Gauge R&R) of less than 10% for high-precision systems, or under 20% for less critical applications
    - "Always confirm with the customer!"

---

5. Document and Monitor

- Record all changes made to the measurement system
- Establish ongoing monitoring to catch drift or degradation in measurement quality

---

Wheeler's 'Honest' EMP Study – ANOVA Method

An alternative Gauge R&R study has been proposed by Dr Donald Wheeler in the form of his 'Honest' EMP Studies (2006) that describes the calculations, output, and classifications for the EMP crossed study. Also known as Wheeler's Method, Honest Gauge R&R Study or Wheeler's 'Honest' EMP Study.

In many respects, AIAG Gauge R&R studies and Wheeler's EMP studies are similar. The collection of subgroups of measurements by part and operator are common to both studies. The results for both studies include assessments of the stability of the measurements.

The principal difference in the methods is how they determine whether a measurement system is acceptable.

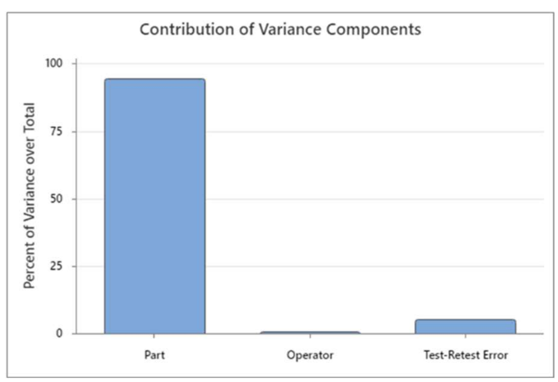
The development of the GR&R Study by the automotive/aero-engine sectors is from a tradition of processes that require high precision from the measurements to meet tight tolerances of characteristics on safety-critical products. The development of Wheeler's EMP criteria comes from a tradition that uses the measurement system to detect shifts in the process average for process improvement activities – and are less strict.

Bearing in mind that the Case Study shown in AESQ's Reference Manual: **RM13003** has been shown to be **unacceptable** for any characteristic category, let us apply the EMP (Crossed) Study – ANOVA Method criteria to the data .. .. ..

The **Variance Components**, Test-Retest Error – Repeatability (**5.206%**) and Operator – Reproducibility (**0.184%**), are very small in comparison to the Part – Product variation (**94.619%**) –

Variance Components

Source	Variance	%Total	StdDev
Test-Retest Error (Repeatability)	0.0001762	5.206	0.0132723
Operator (Reproducibility)	0.0000062	0.184	0.0024979
Part (Product variation)	0.0032013	94.610	0.0565803
Total	0.0033837	100.000	0.0581698



The **EMP Statistics** demonstrate that the measurement system is assessed to be '**First Class**' using the **Classification Guidelines** –

EMP Statistics

Statistic	Value	Classification
Test-Retest Error	0.0133	
Degrees of Freedom	78.0000	
Probable Error	0.0090	
Intraclass Correlation (no bias)	0.9478	First Class
Intraclass Correlation (with bias)	0.9461	First Class
Bias Impact	0.0017	

Classification Guidelines

Classification	Intraclass Correlation	Attenuation of Process Signals	Probability of Warning, Test 1*	Probability of Warning, Tests*
First Class	0.80 - 1.00	Less than 11%	0.99 - 1.00	1.00
Second Class	0.50 - 0.80	11 - 29%	0.88 - 0.99	1.00
Third Class	0.20 - 0.50	29 - 55%	0.40 - 0.88	0.92 - 1.00
Fourth Class	0.00 - 0.20	More than 55%	0.03 - 0.40	0.08 - 0.92

**Probability of detecting a three-standard-deviation shift within 10 subgroups using test 1 or tests 1, 5, 6, and 8.*

Reminder: AIAG Gauge R&R study **rejected** this gauge based on **% contribution** –

Gage Evaluation

Source	StdDev (SD)	Study Var (6 × SD)	%Study Var (%SV)	%Tolerance (SV/Toler)
Total Gage R&R	0.0135053	0.081032	23.22	40.52
Repeatability	0.0132723	0.079634	22.82	39.82
Reproducibility	0.0024979	0.014987	4.29	7.49
Operator	0.0024979	0.014987	4.29	7.49
Part-To-Part	0.0565803	0.339482	97.27	169.74
Total Variation	0.0581698	0.349019	100.00	174.51

In the **RM13003** example, the Gauge R&R **%Tolerance** figure was **40.52%**, that is greater than the permitted maximum of **10%** for a **<u>critical</u> feature**.

The measurement system was therefore deemed to be **unacceptable**

As we have seen, when you specify at least one **specification limit**, Minitab® calculates the probabilities of ***misclassifying*** product – regardless of whether you are using **AIAG Gauge R&R** or **Wheeler's EMP** – and calculates both the **joint probabilities** and the **conditional probabilities** of such 'misclassification'

In this example, the same risks of 'misclassification' remain –

Joint Probability

Description	Probability
A randomly selected part is bad but accepted	0.014
A randomly selected part is good but rejected	0.022

Conditional Probability

Description	Probability
A part from a group of bad products is accepted	0.153
A part from a group of good products is rejected	0.024

Wheeler's 'Honest' EMP Study has a particular benefit when serial data is not available (i.e. in low volume production).

When used with serial data Wheeler's EMP can detect a shift in the process by selectively applying the industry control chart 'detection rules'. Therefore, a measurement system that might, by the AIAG Gauge R&R methods be condemned, can be shown to be entirely capable of monitoring and warning of a process shift.

For quality control purposes, Wheeler's EMP can be said to give a more 'actionable' insight.

CONCLUSION

A **GR&R Study** quantifies how much of the total variation is due to the measurement system – characterised by variation due to the gauge itself and to the operators. The resulting metrics help determine if the measurement system is precise enough to distinguish between good and bad parts.

While **Wheeler's 'Honest' EMP Study** may be used to evaluate the usefulness of a measurement system for process control and improvement applications, automotive and aero-engine customers still demand the use of the **AIAG Gauge R&R Study**.

When to still use AIAG Gauge R&R .. .. ..

- Compliance: If your customers require an MSA report, run AIAG Gauge R&R and keep it on file
- Benchmarking: Compare the results obtained with the AIAG Gauge R&R Study with Wheeler's 'Honest' EMP Study
- In short, do both – it's easy with Minitab® – but consider the impact of the "Probabilities of Misclassification" analyses

Remember the **Golden Rule**

"*Those with the gold make the rule!*"

Stability

Measurement **stability** is the change in **bias** over time. It represents the total variation in measurements of the same part measured over time. This variation over time is also referred to as **consistency** or **drift**.

A **Stability** study measures of how accurately the measurement system performs over time. That is the total variation obtained with a particular device, on the same part, when measuring a single characteristic over time (e.g. minutes, hours, days, etc.).

Establish a master sample

- Sample or Production part (in mid-range)
- Establish its reference value – to traceable standard
- Designate as the master sample

Measure with gauge (*experienced* appraiser)

- Periodically (hourly, per shift, daily) measure master sample **3-5** times
- Take readings at different times of the day

A **control chart** may be used to monitor the stability of a measurement system by measuring a master sample using the same system over time. As measurements are taken, points within the limits indicate that the process has <u>not</u> changed, and points outside the limits indicate that the process has changed.

Plot X_{bar} & R (or X_{bar} & S) control chart (time order)

- Establish control limits
- Evaluate for out-of-control or unstable conditions
- Identify any patterns (with special attention to X_{bar})

Knowledge of the equipment and measurement conditions help identify special causes when the system is found to be out-of-control or unstable.

Case Study

In this Case Study, an experience operator measured a nominated characteristic 5 times on each shift, over a total of 20-shifts.

The results as recorded –

1	2	3	4	5	6	7	8		20
6.005	6.122	6.03	5.884	5.884	6.076	6.092	6.014		6.013
6.027	6.125	5.972	5.968	5.968	6.068	6.005	5.989		5.966
6.167	6.127	6.041	6.019	6.019	5.977	5.995	6.044		6.122
5.811	5.994	6.003	6.057	6.057	5.916	5.884	5.95		6.174
6.038	5.927	6.198	6.075	6.075	5.955	5.992	5.963		6.012

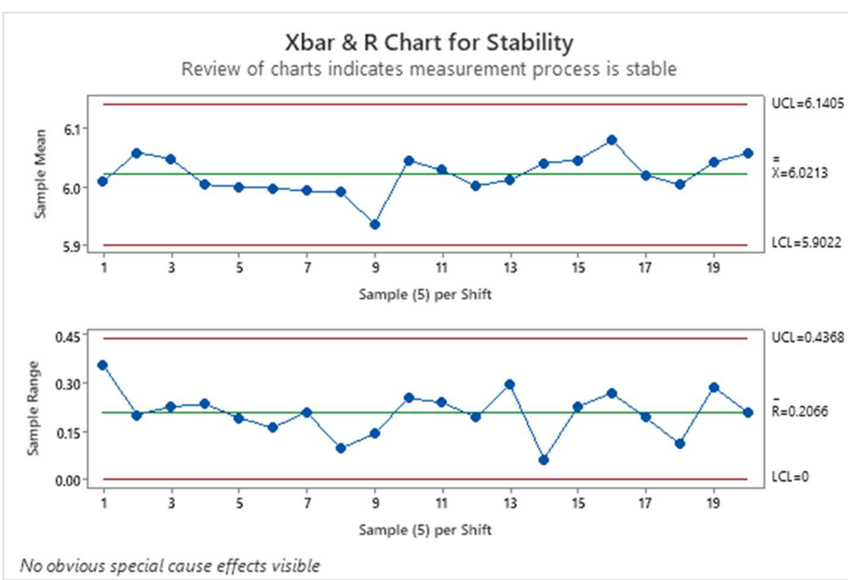
Study results

Using the data, an X_{bar} & R Chart was created using Minitab®–

Choose Stat > Control Charts > Variable Charts for Subgroups > Xbar-R

Subgroup sizes: 5

The following analysis is displayed –



Before interpreting the X_{bar} Chart, examine the R Chart to determine whether the process variation is in control. If the R Chart is not in control, then the control limits on the X_{bar} chart are not accurate.

The R Chart plots the subgroup ***ranges***. As the subgroup size (5) is constant, then the centre line on the R Chart is the average of the subgroup ranges.

As there are no out-of-control points, the control limits on the X_{bar} Chart will be accurate.

The X_{bar} Chart plots the ***average*** of the measurements within each subgroup. The centre line is the average of all subgroup averages. The control limits on the X_{bar} Chart, which are set at a distance of 3 standard deviations above and below the centre line, show the amount of variation that is expected in the subgroup averages.

As there are no out-of-control points Case Study, the measurement process is considered to be '***in statistical control***' and therefore **stable**.

Conclusion

The purpose of this **article** has been to present pragmatic guidelines for assessing measuring systems used for quality control purposes. The primary focus is on MSA methods acceptable within the automotive, aero-engine and aerospace sectors.

Measurement systems can include measurement devices (e.g. gauges), operators and measurement procedures all of which can affect the measurement of a characteristic.

As we have seen, measurement error can be classified into two major categories of **Accuracy** (Linearity & Bias) and **Precision** (Repeatability & Reproducibility). The analytical methods you use will depend on customer-specific demands. We have also added **Stability** that assesses how accurately the measurement system performs over time.

MSA is a method for determining whether a measurement system is acceptable. In broad terms, in an acceptable measurement system most of the variation must be part-to-part, meaning that the measurement system can effectively distinguish differences between parts.

In the words of David Crosby:

"*If you don't know the capability of your measurement system, you don't know if your measurements or your products are good or bad*"

In the words of W. Edwards Deming:

"*Any technique can be useful if its limitations are understood and observed*"

Remember, **MSA** does *not* evaluate the product – it evaluates the ability to measure the part accurately and consistently. If you don't know the capability of your measurement system, you don't know if your measurements or your parts are good or bad.

If measurement variation can be reduced and gauge repeatability and reproducibility ratios improved, it is easier to differentiate between parts that are in or out of specification, allowing parts to be accepted or rejected with greater confidence.

A final thought: When the measurement system is being used with Control Charts (i.e. X_{bar} & R, I MR, etc.) is it still acceptable? If so, its's advisable to re-check substituting LCL for LSL and UCL for LSL.

Improving the quality of a measurement system for making quality control decisions is a cycle of investigation, correction, and validation.

TEC's Training Courses

✳️ F-to-F Course: https://tectransnational.com/courses/measurement-systems-analysis-as13003-face-to-face ~ Bring your own products & gauges for the Workshop

✴️ e-Learning Course: ~ https://tectransnational.com/courses/as13003-measurement-systems-analysis-msa-elearning ~ Download Excel data to test you skills