Measurement Systems Analysis
January 2015 Meeting

Steve Cox

ASQ
The Global Voice of Quality™

Steve Cox

- Currently retired
- 33 Years with 3M Mostly quality related: 37 total in Quality
- ASQ Certified Quality Engineer
- Certified Black Belt Coach, Supported Healthcare and new engineer Curriculum.
- Experience in plant, plant support, and “Big” B support
- Photo Products, Printing Products, Tape, Healthcare.
- Quality Engineering
- Quality Supervisor
- Tech Service
- Reliability Engineering
- Interests
    - DOE
    - Modeling
    - Monte Carlo Simulation
    - MSA

MSA Definition

- Measurement system analysis (MSA) is an experimental and mathematical method of determining how much the variation within the measurement process contributes to overall process variability.

MSA

• Why MSA?

- **I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of *science*, whatever the matter may be.**
  Lord Kelvin
  - Lecture on "Electrical Units of Measurement" (3 May 1883), published in *Popular Lectures* Vol. I, p. 73

Why MSA?

- Improvement cannot take place without metrics.
  Often an improvement in the metrics of a system improves the system.

Definitions

- Accuracy
  Closeness to target value

- Precision
  Grouping of values

- Bias

  Offset from “real” value

  Linearity
  Constant bias

Definitions

- Stability
  The ability of a standard to exhibit the same value under varying environments.

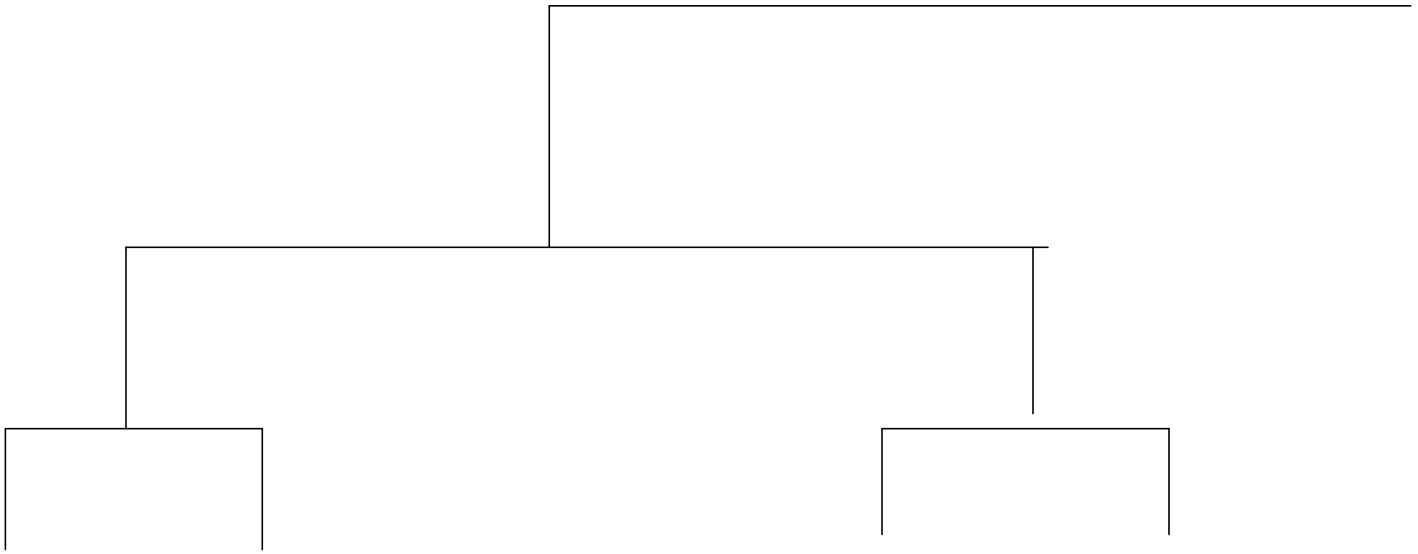
- Repeatability
  The variation within an appraiser

- Reproducibility
  The variation between appraisers.

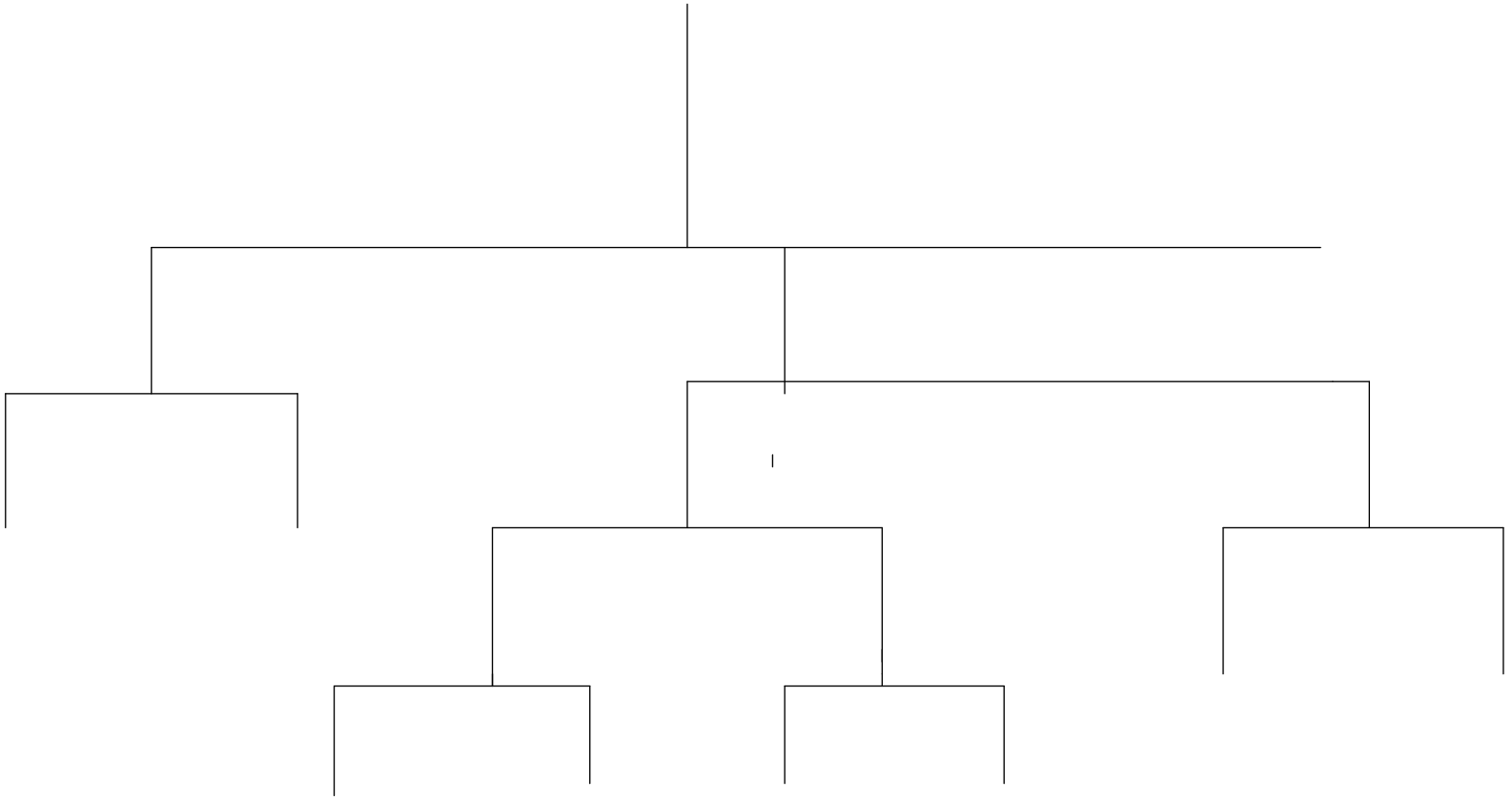
Types of Tests

- Destructive
  Test cannot be repeated. Test destroys or compromises sample.
  The test has inherent repeatability issues
- Non-Destructive
  Test does not materially affect the sample

Appraiser Matrix



Instruments as Appraiser



Sample MSA

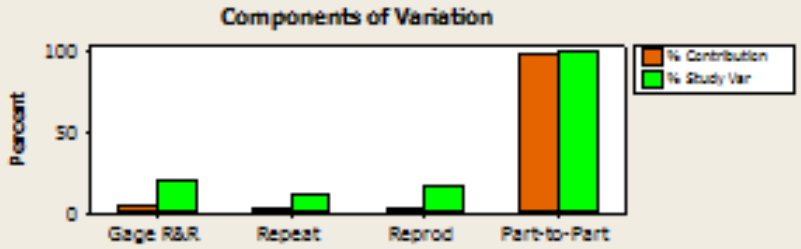
- Four Appraisers
- Five samples
- 2 replicates
- Destructive test.

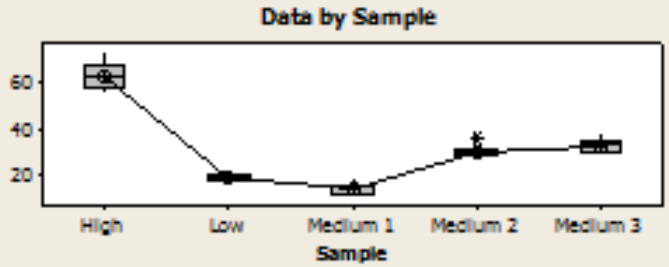
Sample MSA

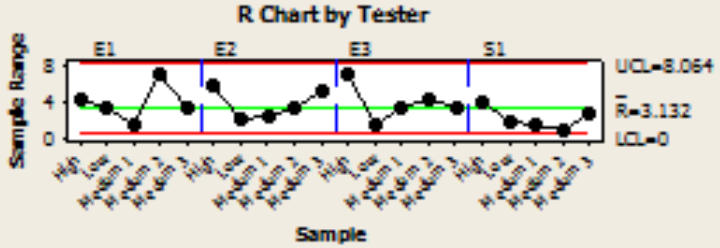
Gage R&R (ANOVA) for Data

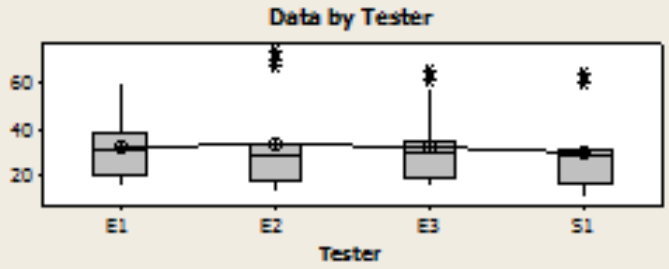
Reported by:
Tolerance:
Misc:

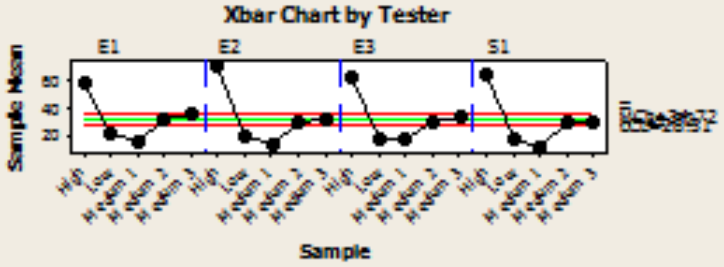
Gage name:
Date of study:

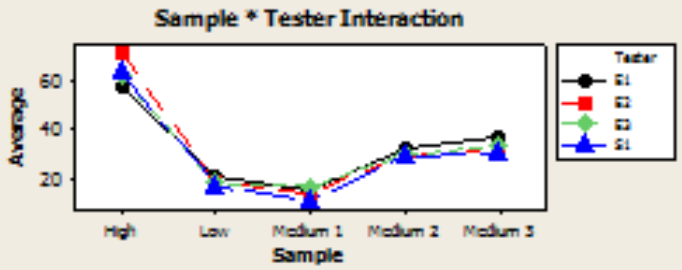












Two-Way ANOVA Table With Interaction

Gage R&R Study - ANOVA Method

Source	DF	SS	MS	F	P
Sample	4	18192.1	4548.03	142.451	0.000
Tester	3	74.0	24.67	0.773	0.531
Sample * Tester	12	383.1	31.93	8.918	0.000
Repeatability	40	143.2	3.58		
Total	59	18792.4			

Alpha to remove interaction term = 0.25

Gage R&R

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	13.029	3.35
Repeatability	3.580	0.92
Reproducibility	9.449	2.43
Tester	0.000	0.00
Tester*Sample	9.449	2.43
Part-To-Part	376.342	96.65
Total Variation	389.371	100.00

Source	StdDev (SD)	Study Var (6 * SD)	%Study Var (%SV)
Total Gage R&R	3.6096	21.658	18.29
Repeatability	1.8921	11.353	9.59
Reproducibility	3.0739	18.443	15.58
Tester	0.0000	0.000	0.00
Tester*Sample	3.0739	18.443	15.58
Part-To-Part	19.3995	116.397	98.31
Total Variation	19.7325	118.395	100.00

Number of Distinct Categories = 7

Appraisers

- Must be representative of the population.
- Include best, worst and “tweens.
- How many?

not too many
not too few

Samples

- Representative of the range of interest.
- Samples would ideally be blind.
- Generally the spec range plus a “skosh”
- It is tempting to take “available” samples. “there be dragons here”

Replicates

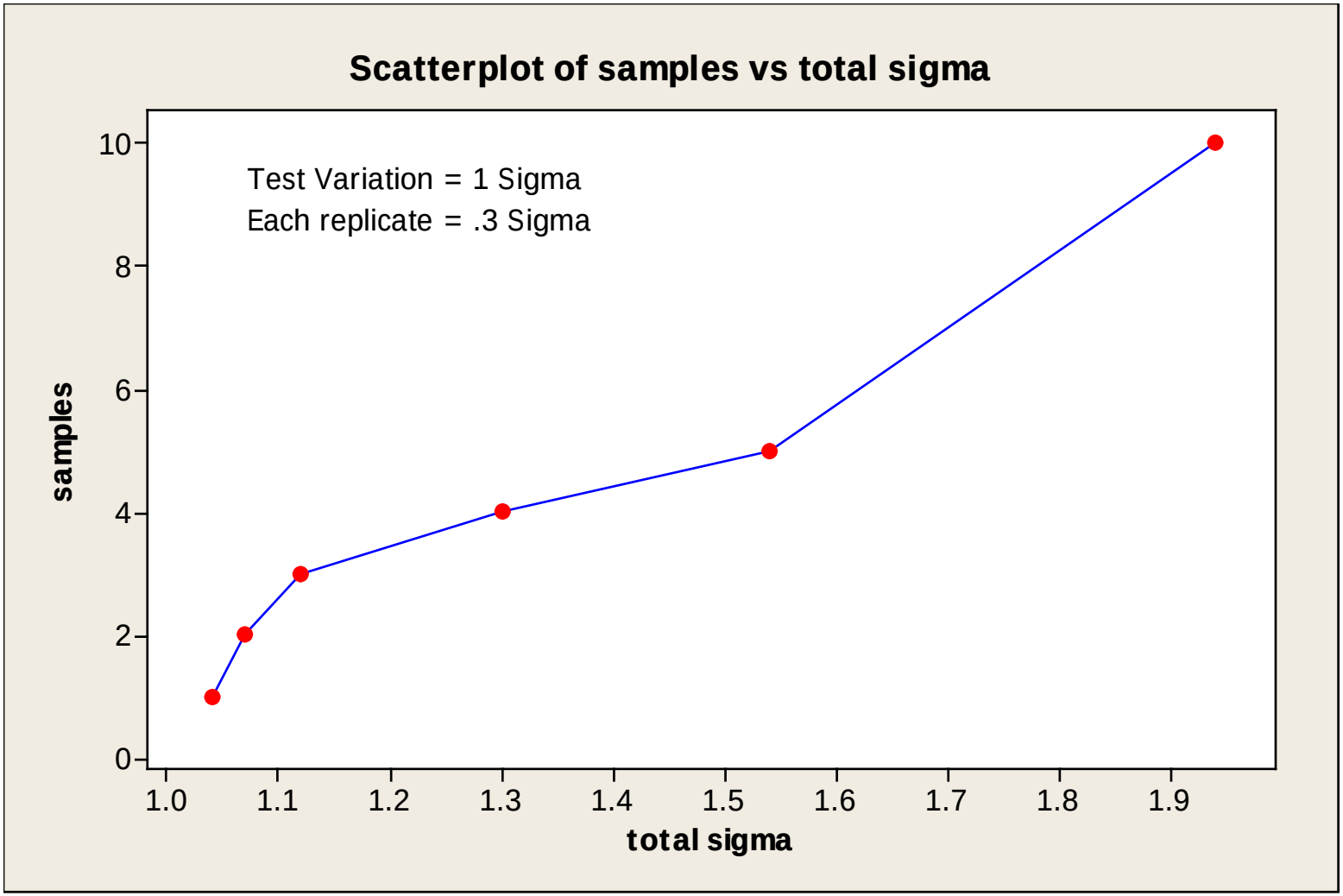
- Pooled to get test variation, so important.
- Too few poor estimate (low)
- Too many poor estimate (high)
- Like earth you need to be in the Goldilocks zone.

- Always ,it is ***critical*** with destructive tests

Destructive Samples

- Samples always contain “extra” variation.
- Randomized “sister samples”.
- Sister samples limits the scope of the study.

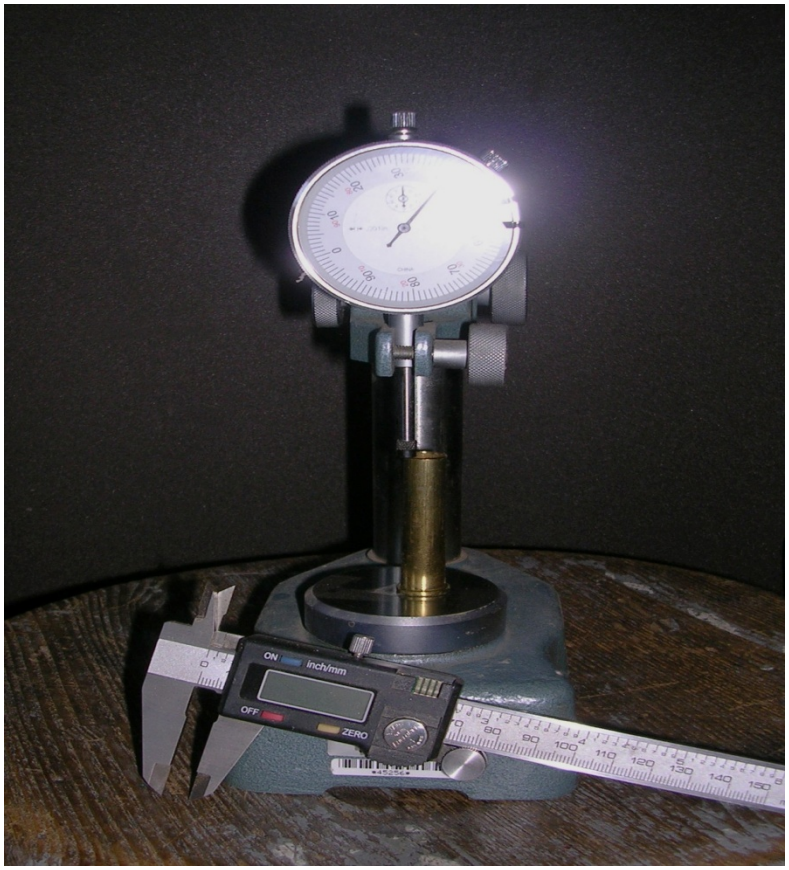
Destructive Test Variation



45-70 Brass Example

- Actual example from my shop.
- How much variation is there in measuring 45-70 brass.
- Need to know when to trim.
- Calipers used.
- 1 Trained operator, 2 noobies.
- 10 samples, 3 replicates, non-destructive.
- Range of samples from within received sample, not possible to bracket range of interest

45-70 Brass Length



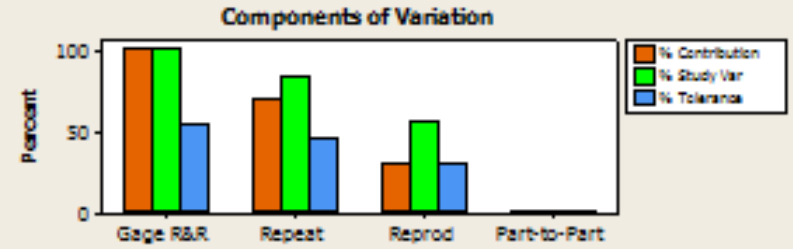
Measuring trim length of 45-70 shell casing with caliper or height gage and indicator

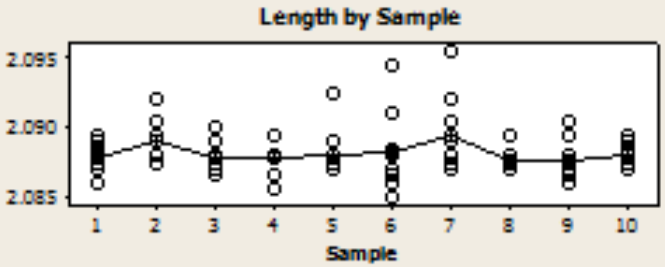
ASQ

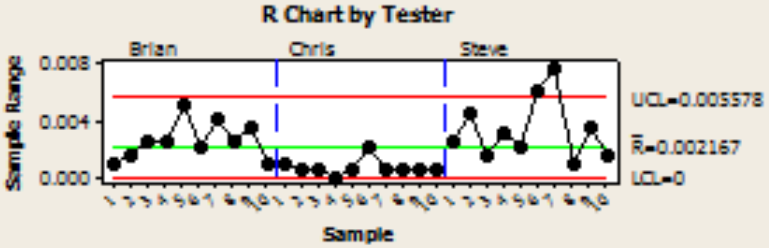
45-70 Case Length

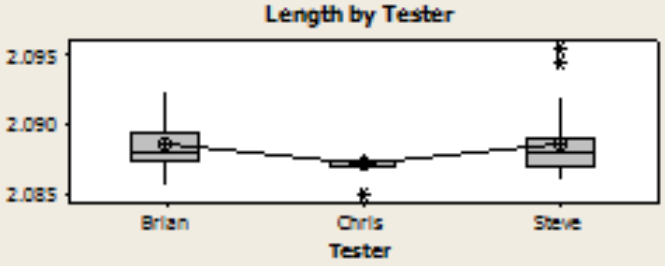
Gage name: Caliper
Date of study:

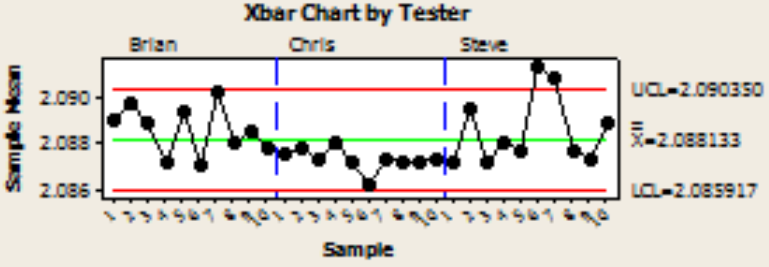
Reported by: Steve Cox
Tolerance:
Misc: 1204 MSA Talk

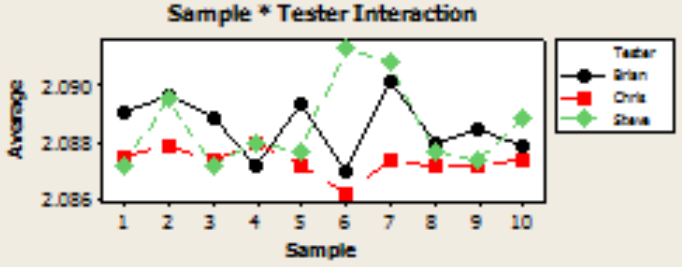












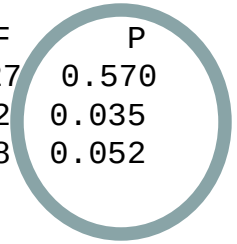
ASQ

Gage R&R Study - ANOVA Method

Gage R&R for Length

Gage name: Caliper
Date of study:
Reported by: Steve Cox
Tolerance:
Misc: 1204 MSA Talk

Source	DF	SS	MS	F	P
•Sample	9	0.0000301	0.0000033	0.86627	0.570
Tester	2	0.0000312	0.0000156	4.05162	0.035
Sample * Tester	18	0.0000694	0.0000039	1.76408	0.052
Repeatability	60	0.0001312	0.0000022		
Total	89	0.0002619			



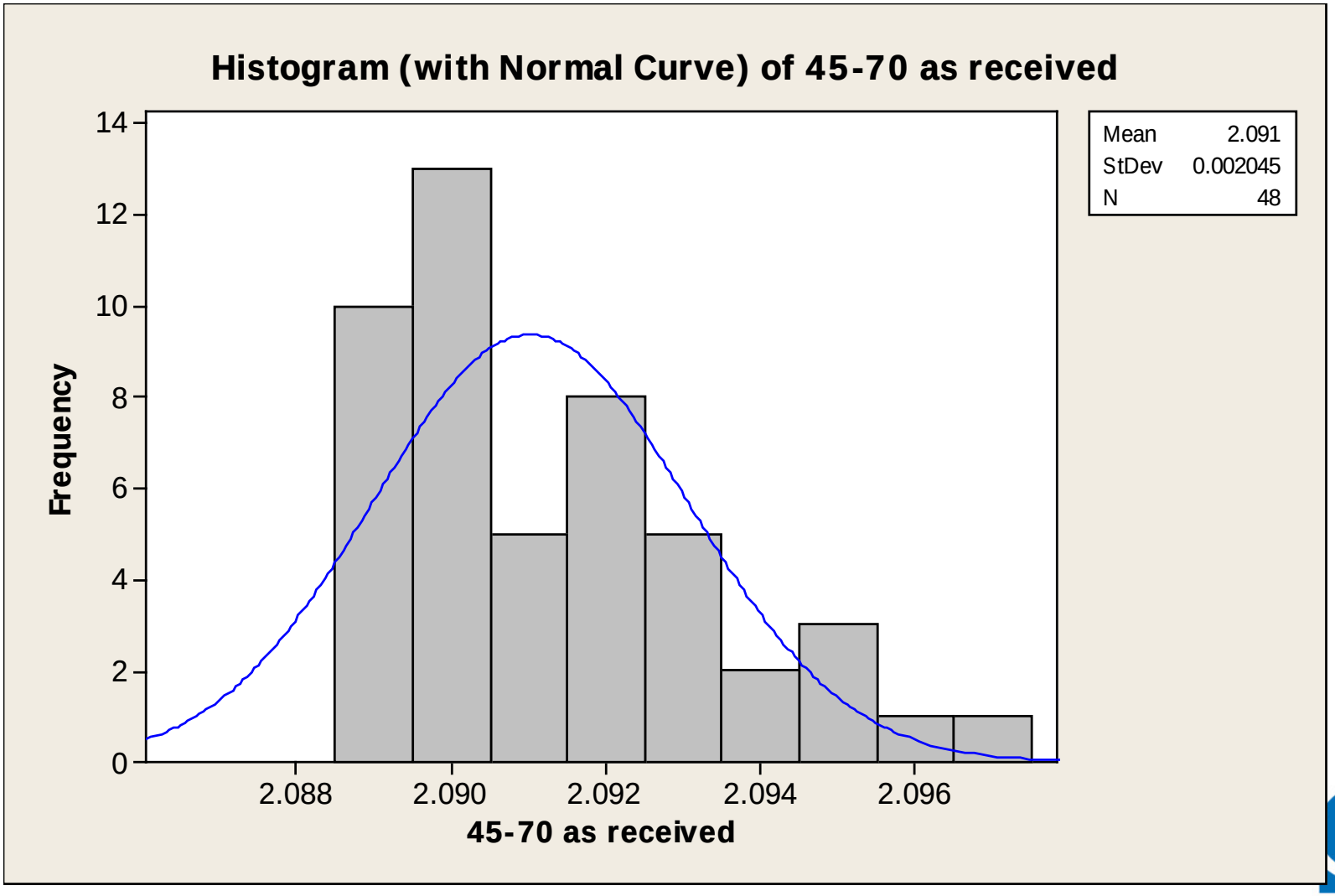
Gage R&R

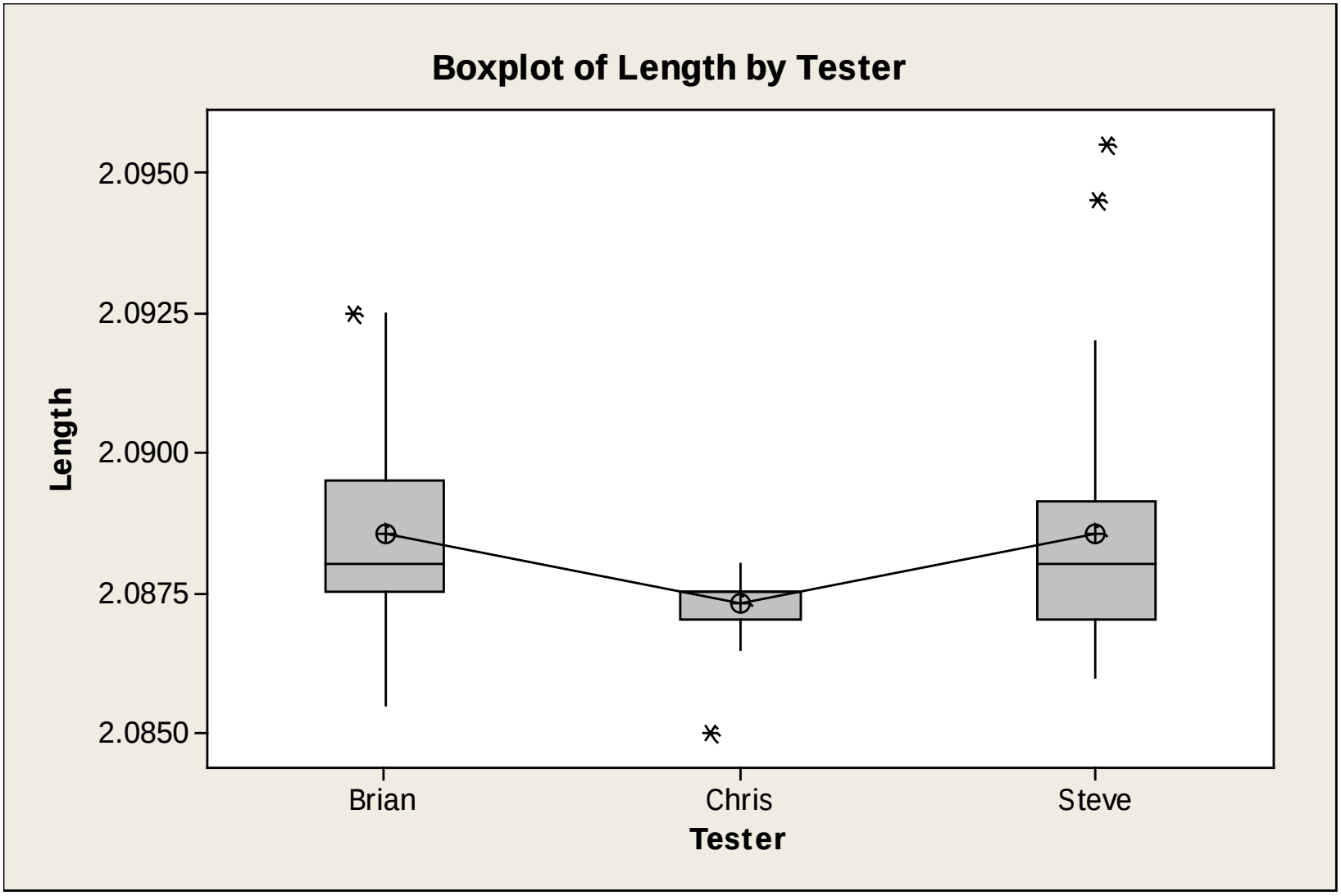
Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.0000031	100.00
Repeatability	0.0000022	69.73
Reproducibility	0.0000009	30.27
Tester	0.0000004	12.51
Tester*Sample	0.0000006	17.76
Part-To-Part	0.0000000	0.00
Total Variation	0.0000031	100.00

Process tolerance = 0.02

Source	Study Var StdDev (SD)	%Study Var (6 * SD)	%Tolerance (%SV)	(SV/Toler)
Total Gage R&R	0.0017706	0.0106239	100.00	53.12
Repeatability	0.0014786	0.0088713	83.50	44.36
Reproducibility	0.0009742	0.0058452	55.02	29.23
Tester	0.0006263	0.0037580	35.37	18.79
Tester*Sample	0.0007462	0.0044771	42.14	22.39
Part-To-Part	0.0000000	0.0000000	0.00	0.00
Total Variation	0.0017706	0.0106239	100.00	53.12

Number of Distinct Categories = 1


Histogram (with Normal Curve) of 45-70 as received
Mean 2.091
StDev 0.002045
N 48
Frequency
0
2
4
6
8
10
12
14
2.088
2.090
2.092
2.094
2.096
45-70 as received

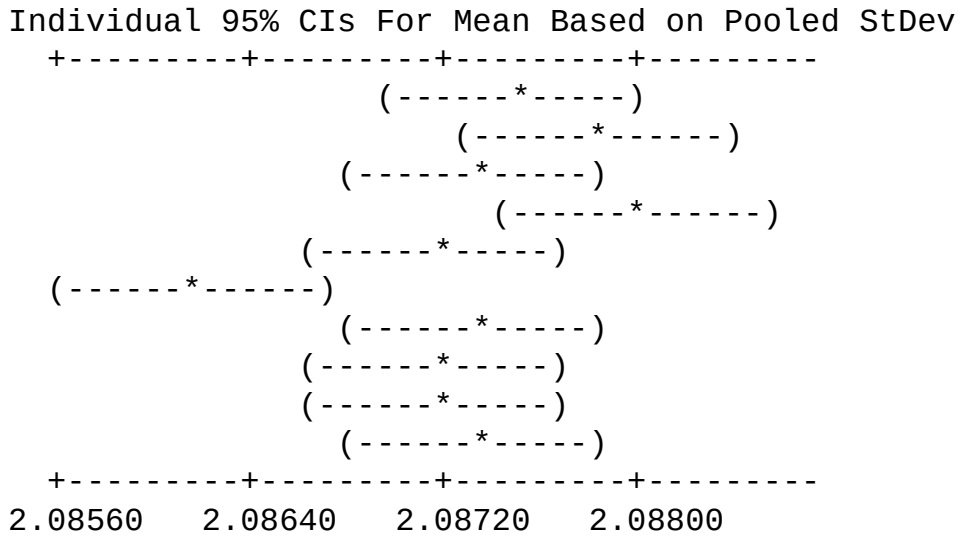

Boxplot of Length by Tester
Length
2.0950
2.0925
2.0900
2.0875
2.0850
Brian
Chris
Steve
Tester

One-way ANOVA: len versus case Chris

Source	DF	SS	MS	F	P
case	9	0.0000065	0.0000007	3.75	0.007
Error	20	0.0000038	0.0000002		
Total	29	0.0000103			

S = 0.0004378 R-Sq = 62.78% R-Sq(adj) = 46.04%

Level	N	Mean	StDev
1	3	2.08750	0.00050
2	3	2.08783	0.00029
3	3	2.08733	0.00029
4	3	2.08800	0.00000
5	3	2.08717	0.00029
6	3	2.08617	0.00104
7	3	2.08733	0.00029
8	3	2.08717	0.00029
9	3	2.08717	0.00029
10	3	2.08733	0.00029

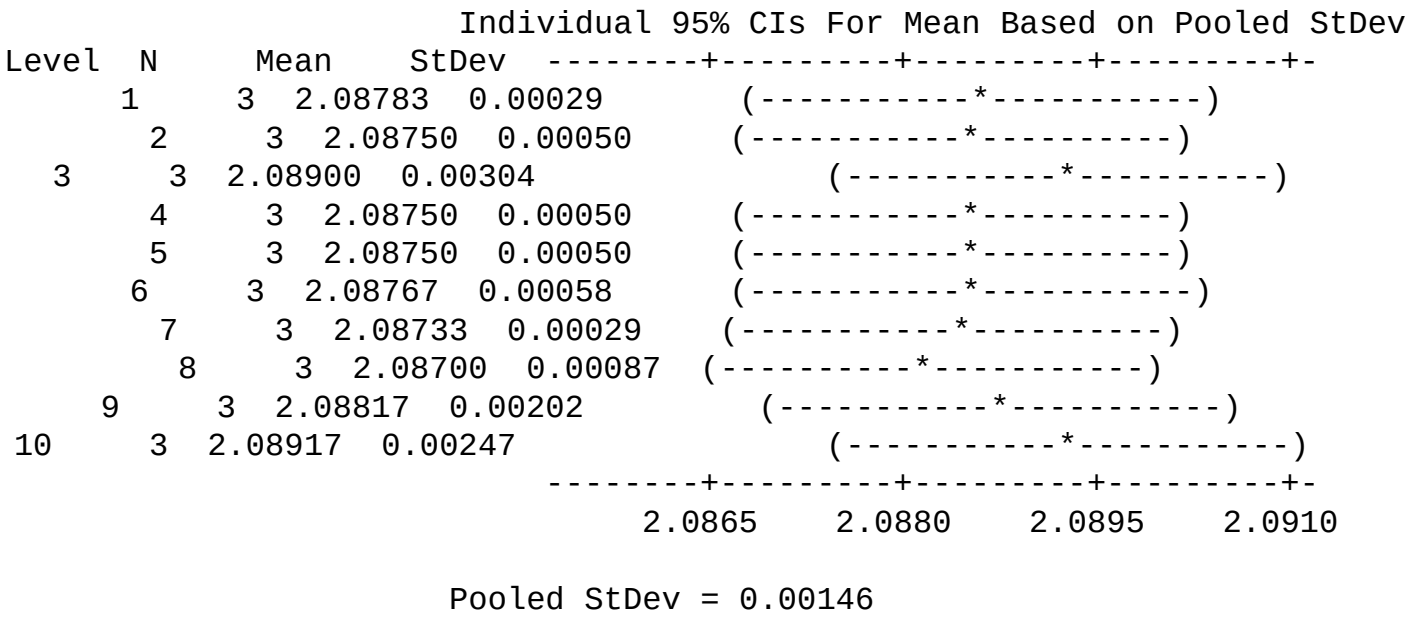


Pooled StDev = 0.00044

One-way ANOVA: Length versus Sample Brian

Source	DF	SS	MS	F	P
Sample	9	0.0000136	0.0000015	0.71	0.696
Error	20	0.0000428	0.0000021		
Total	29	0.0000565			

S = 0.001463 R-Sq = 24.14% R-Sq(adj) = 0.00%



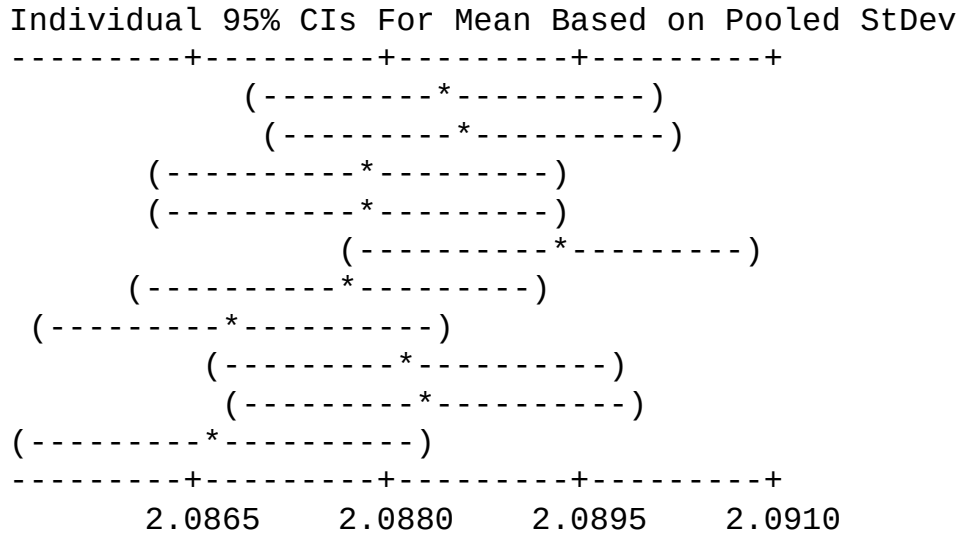
Results for: Sort Tester & Sample(Tester = Steve

Results for: Sort Tester & Sample(Tester = Steve)

Source	DF	SS	MS	F	P
Sample	9	0.0000177	0.0000020	1.16	0.370
Error	20	0.0000340	0.0000017		
Total	29	0.0000517			

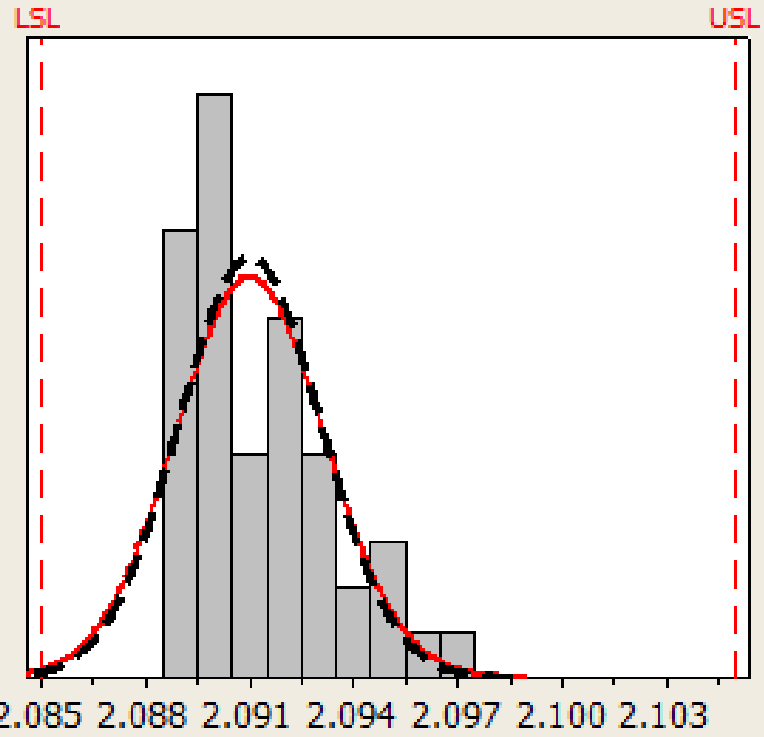
S = 0.001304 R-Sq = 34.29% R-Sq(adj) = 4.72%

Level	N	Mean	StDev
1	3	2.08850	0.00087
2	3	2.08867	0.00076
3	3	2.08783	0.00104
4	3	2.08783	0.00104
5	3	2.08933	0.00202
6	3	2.08767	0.00161
7	3	2.08683	0.00161
8	3	2.08817	0.00104
9	3	2.08833	0.00144
10	3	2.08667	0.00104



Pooled StDev = 0.00130

Process Capability of 45-70 as received



Process Data	
LSL	2.085
Target	*
USL	2.105
Sample Mean	2.09102
Sample N	48
StDev(Within)	0.00215029
StDev(Overall)	0.0020446

Within
Overall

Potential (Within) Capability	
Cp	1.55
CPL	0.93
CPU	2.17
Cpk	0.93

Overall Capability	
Pp	1.63
PPL	0.98
PPU	2.28
Ppk	0.98
Cpm	*

Observed Performance	
PPM < LSL	0.00
PPM > USL	0.00
PPM Total	0.00

Exp. Within Performance	
PPM < LSL	2555.10
PPM > USL	0.00
PPM Total	2555.10

Exp. Overall Performance	
PPM < LSL	1616.12
PPM > USL	0.00
PPM Total	1616.12

ASQ

Test Estimate from Gage R&R .00177
Test Estimate from IM-R range .0013
Population Standard Deviation .00177

Power and Sample Size
Chris
1-Sample t Test
Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.00044
Sample

Difference	Size	Power
0.004	2	0.955695

Power and Sample Size
Brian
1-Sample t Test
Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.00146
Sample

Difference	Size	Power
0.004	2	0.455563

Power and Sample Size
Steve
1-Sample t Test
Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.0013
Sample

Difference	Size	Power
0.004	2	0.503947

ASQ

45 Brass Take Away

- Measuring brass tubes even with digital gauging is not simple.
- Height gauge probably a better method but would need verification.
- Upper spec would need to be reduced about .005 (worst operator) to be absolutely safe. This means trimming early.
- Just because numbers are “bad” does not mean system is not useful.
- In this case the “worst” operator is the limiting factor.

Example 2 Blood Pressure.

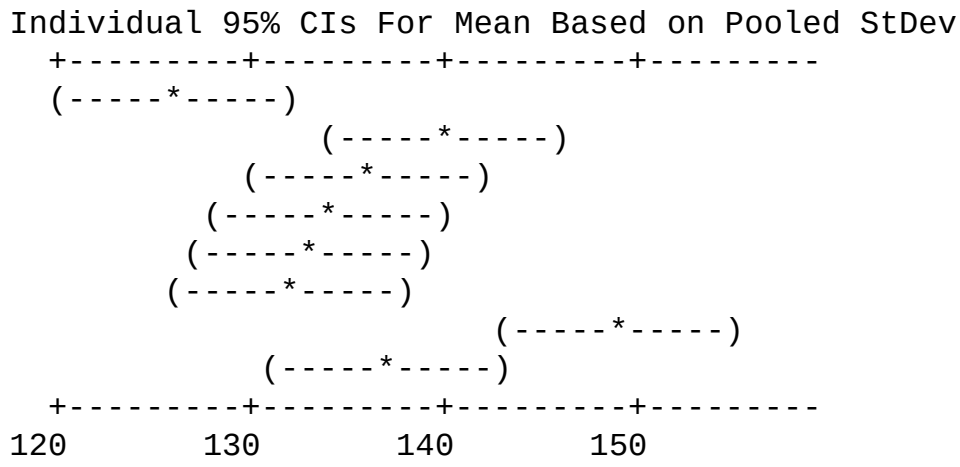
- Blood pressure varies throughout the day. What is the variation of the device?

One-way ANOVA: systolic versus Reading

Source	DF	SS	MS	F	P
Reading	7	914.0	130.6	5.53	0.002
Error	16	378.0	23.6		
Total	23	1292.0			

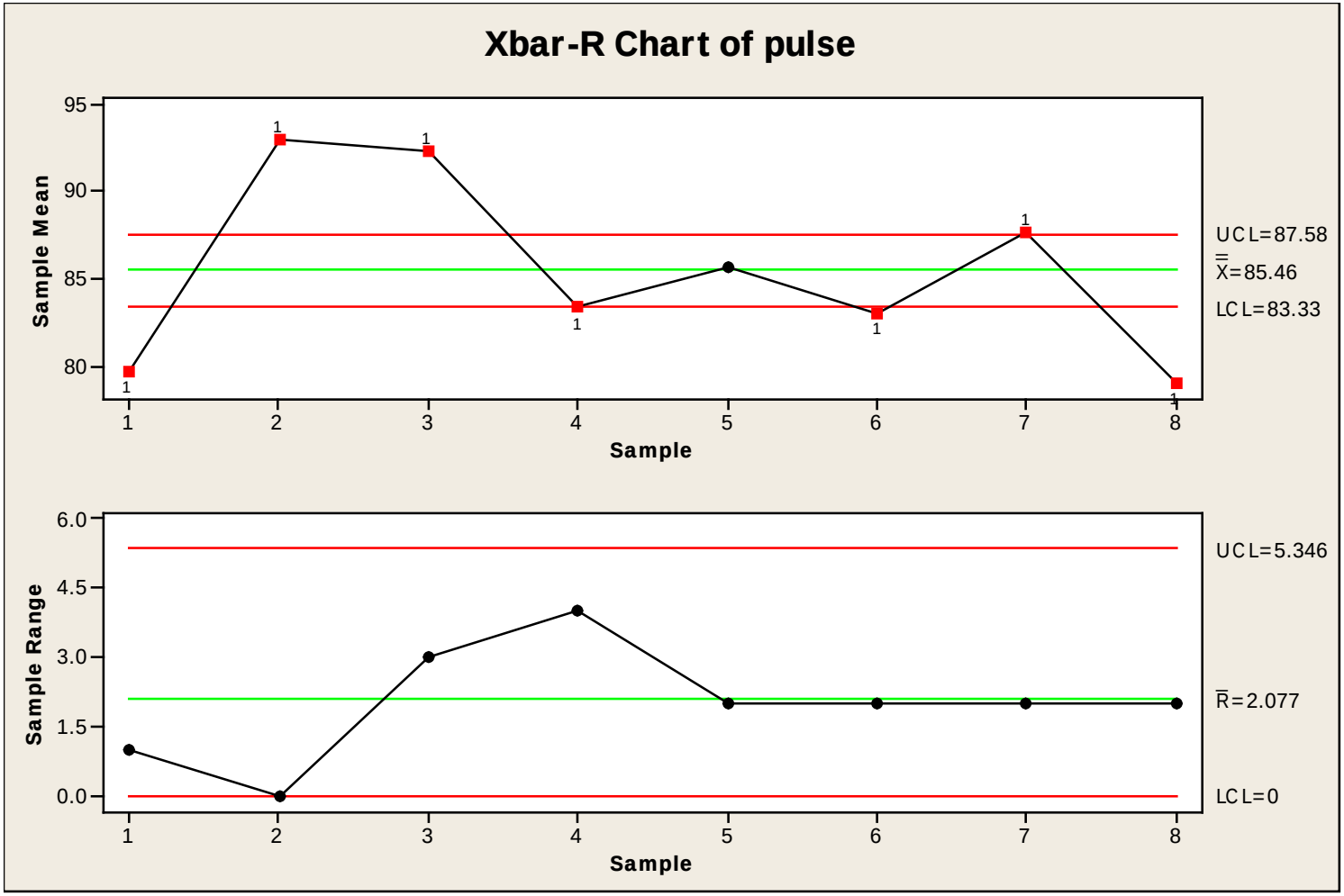
S = 4.861 R-Sq = 70.74% R-Sq(adj) = 57.94%

Level	N	Mean	StDev
1	3	126.33	5.86
2	3	140.00	4.58
3	3	136.00	0.00
4	3	134.33	3.21
5	3	133.00	3.61
6	3	132.33	5.51
7	3	149.00	4.36
8	3	137.00	7.81



Pooled StDev = 4.86

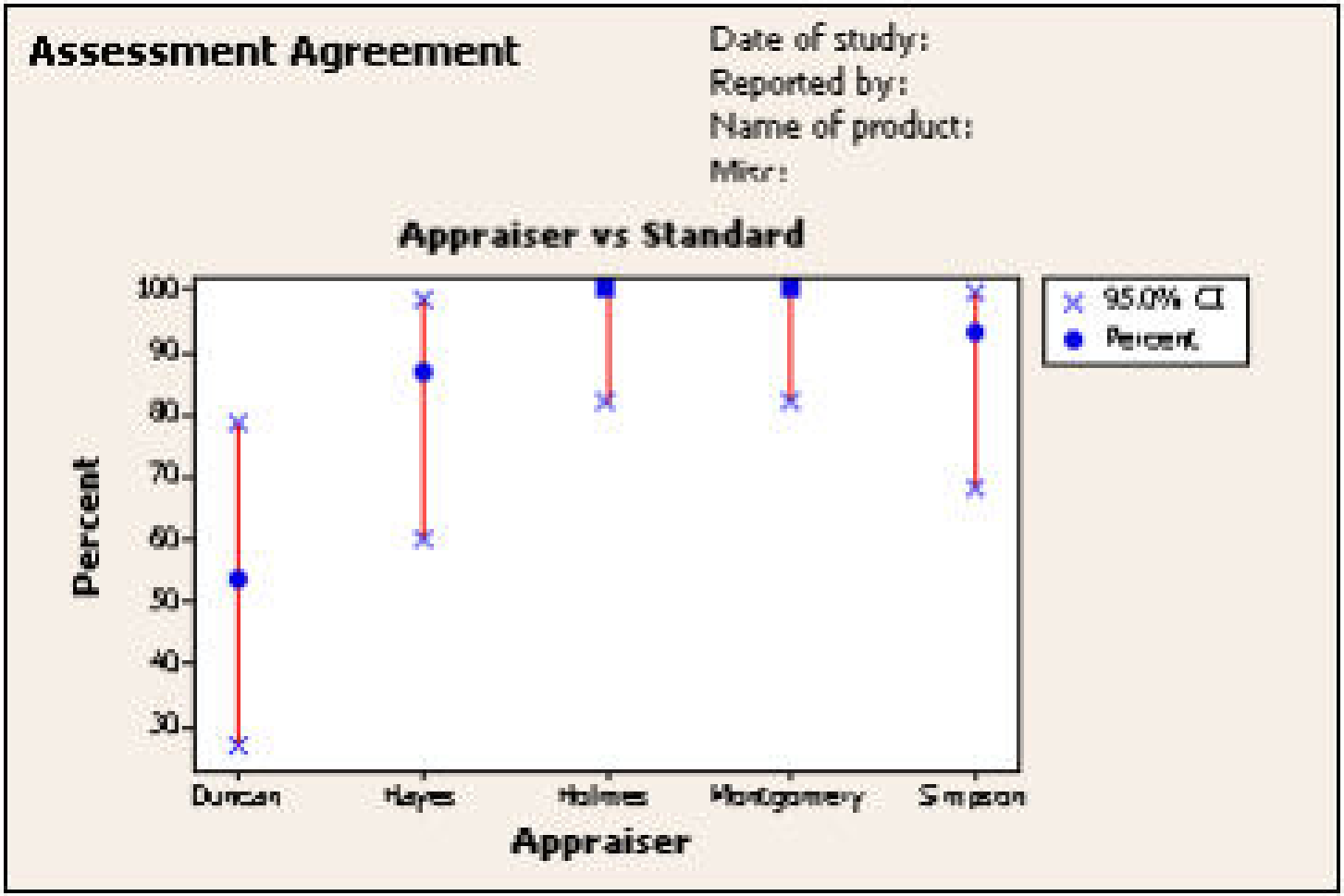
Sigma hat = 1.21 population sigma 5.14



One-way ANOVA: systolic versus Reading

Type	Rbar/d2	Sigma hat	Pooled sigma
Systolic	8.36/1.693	4.94	4.86
Diastolic	5.31/1.693	3.14	3.10
Pulse	2.077	1.23	1.21

Attribute Agreement



Attribute Agreement

- Attributes has much less information than variables.
- Samples need to be good, bad, mostly in “grey zone”
- Need more sample than using variables. Small sample sizes can look good but have poor confidence intervals.

ASQ

Attribute Agreement

- Samples should be blind.
  Time separation can be used for blinding but may add variability.
- Expert can be used as reference but must be internally consistent.
- Watch those confidence intervals !!!!!
- Kappa shows agreement among appraisers.

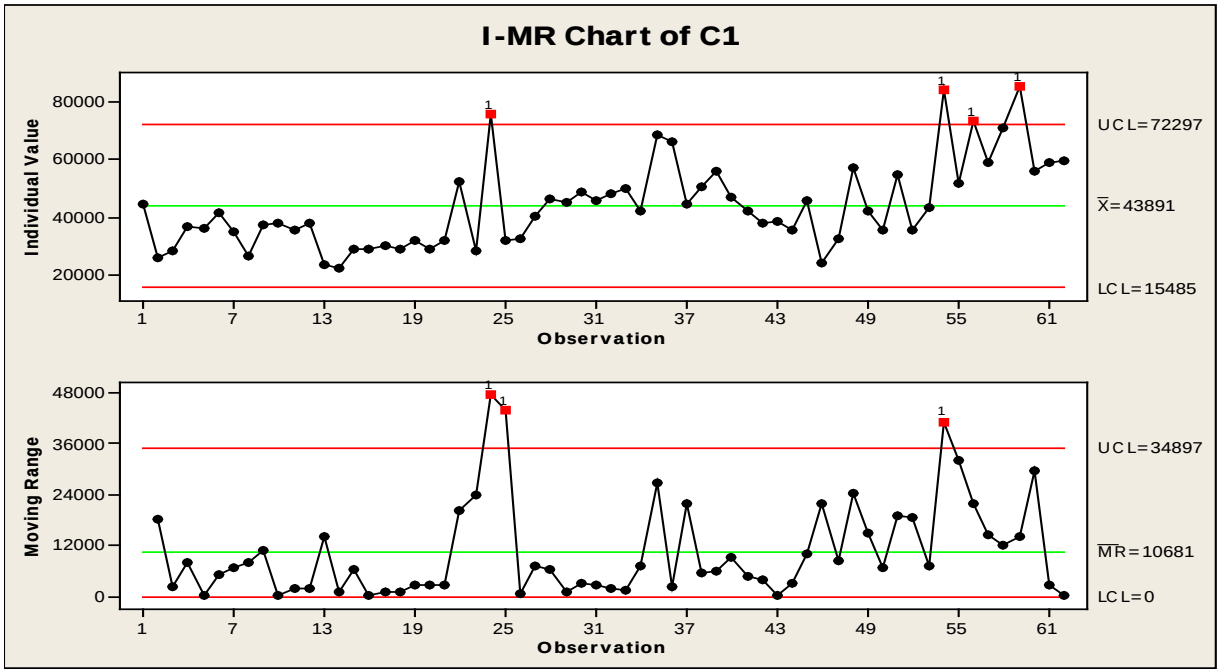
ASQ

Sales Example

```
1-Sample t Test

Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05  Assumed standard deviation = 9469


            Sample  Target
Difference    Size   Power  Actual Power
     26191       5    0.95      0.993548
```



Final Thoughts

- MSA is more than Gage R&R
- Some measure of measurement variation is needed for any improvement.
- Full gage R&R may be a later step.
- Existing data often can give a good picture of the measurement system.
- Any study is a “snapshot” in time.
- All studies have CI’s around the data. Seldom shown but they are there.