Measurement System Analysis

Greg Larsen

G. A. Larsen Consulting

Ops A La Carte

Reasons for Making Measurements

- → Discriminate among products
- Monitor performance of production process
- Manufacturing process improvement
- Specification setting

Proverb

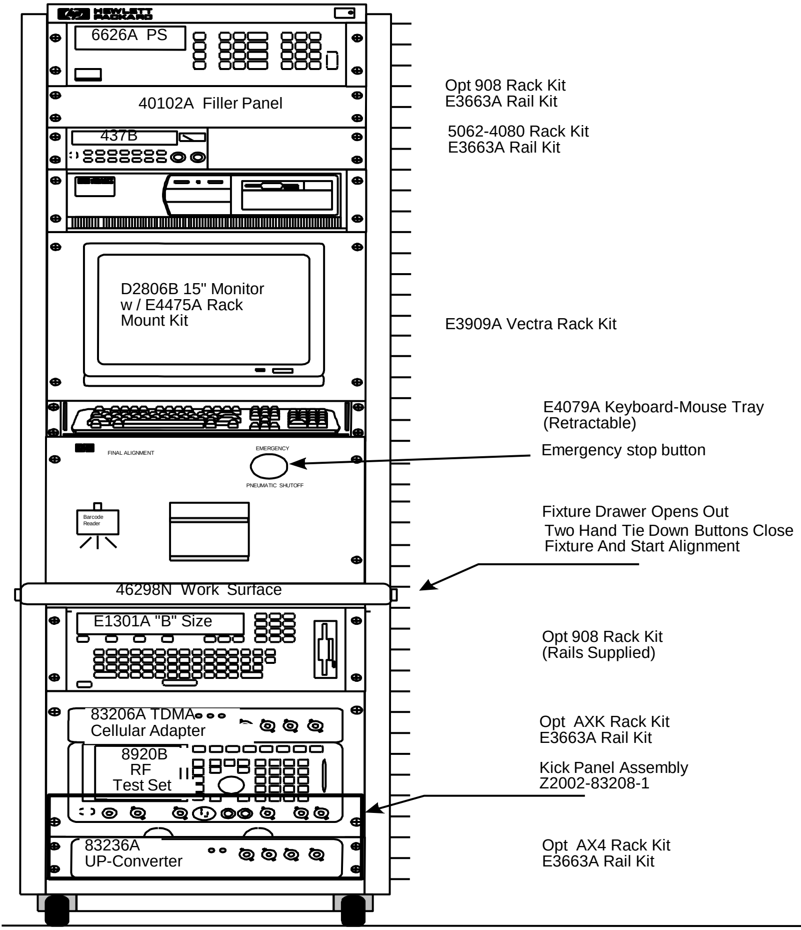
> Acquitting the guilty and condemning the innocent – the Lord detests them both.
>
> Proverbs 17:15

Example - Cell Phone Test

- A cell phone manufacturer uses Agilent equipment to test whether each phone is functioning properly just prior to shipment.
- The test system consists of a fixture which secures the phone and a rack of measurement equipment which tests over 40 functions.

HEWLETT PACKARD

System Configuration Layout



Opt 908 Rack Kit
E3663A Rail Kit

5062-4080 Rack Kit
E3663A Rail Kit

E3909A Vectra Rack Kit

E4079A Keyboard-Mouse Tray
(Retractable)

Emergency stop button

Fixture Drawer Opens Out
Two Hand Tie Down Buttons Close
Fixture And Start Alignment

Opt 908 Rack Kit
(Rails Supplied)

Opt AXK Rack Kit
E3663A Rail Kit

Kick Panel Assembly
Z2002-83208-1

Opt AX4 Rack Kit
E3663A Rail Kit

Example - Cell Phone Test

- Questions arose on an installed base of 20 test systems as to whether the measurement process had excessive noise.
- In particular, the phone manufacturer felt the false failure rate was too high.
- A measurement system analysis was conducted.

Example - Cell Phone Test

- 24 phones were randomly selected
- 6 test systems were selected
- 2 repeated measurements were taken on each phone and test system
- 42 phone parameters were measured on each of the 288 design points
- The data were used to assess the adequacy of the test process.

The General Test Scenario

- Classic truth table
- Augmented with measurement error distributions
- Warranted product specs versus production test limits

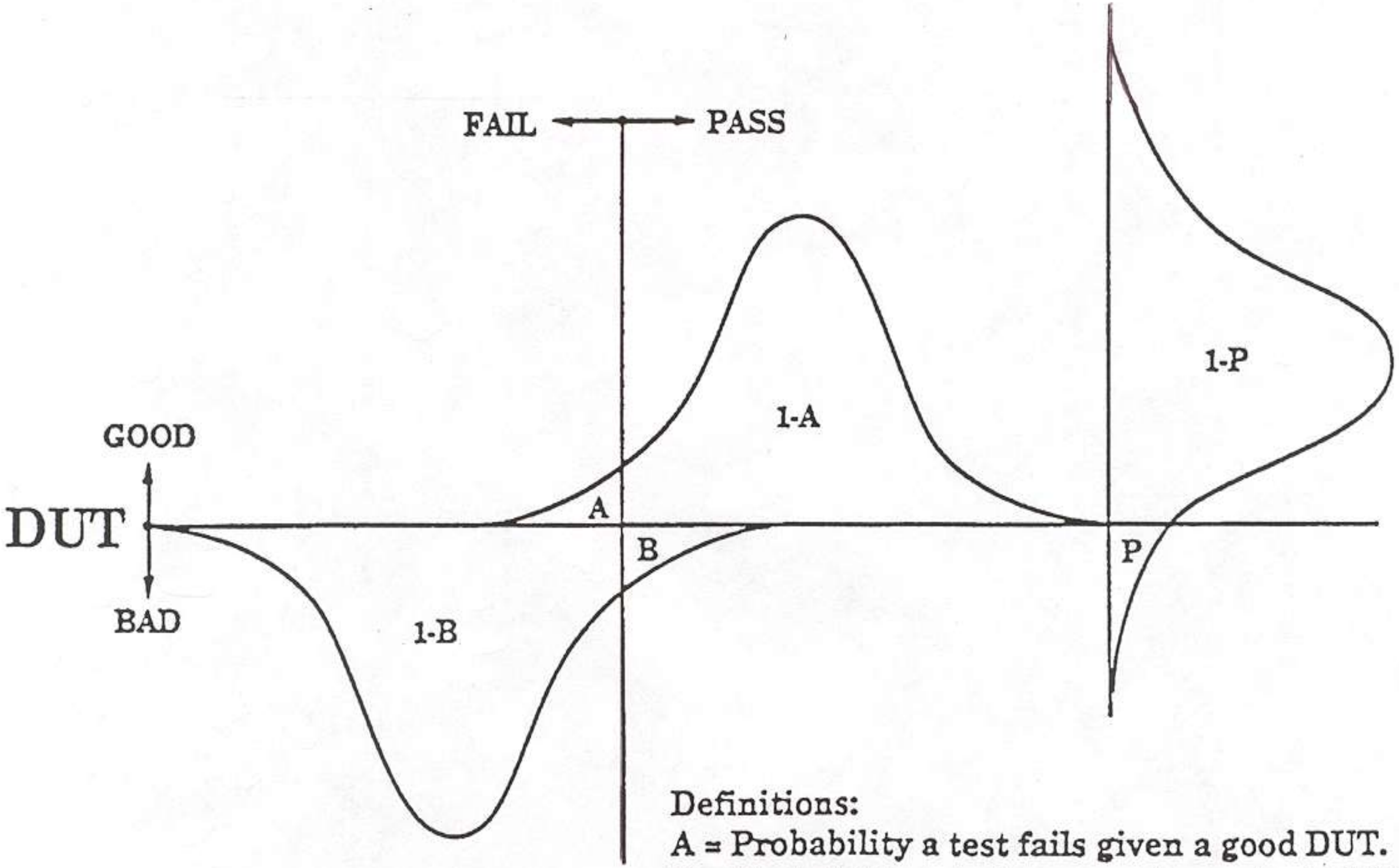
Classic Truth Table

TEST RESULT

		Fail	Pass	
DUT	Good	False Failure	Valid	Yield (1-P)
	Bad	Valid	Missed Fault	Defect Rate (P)

The two types of test errors

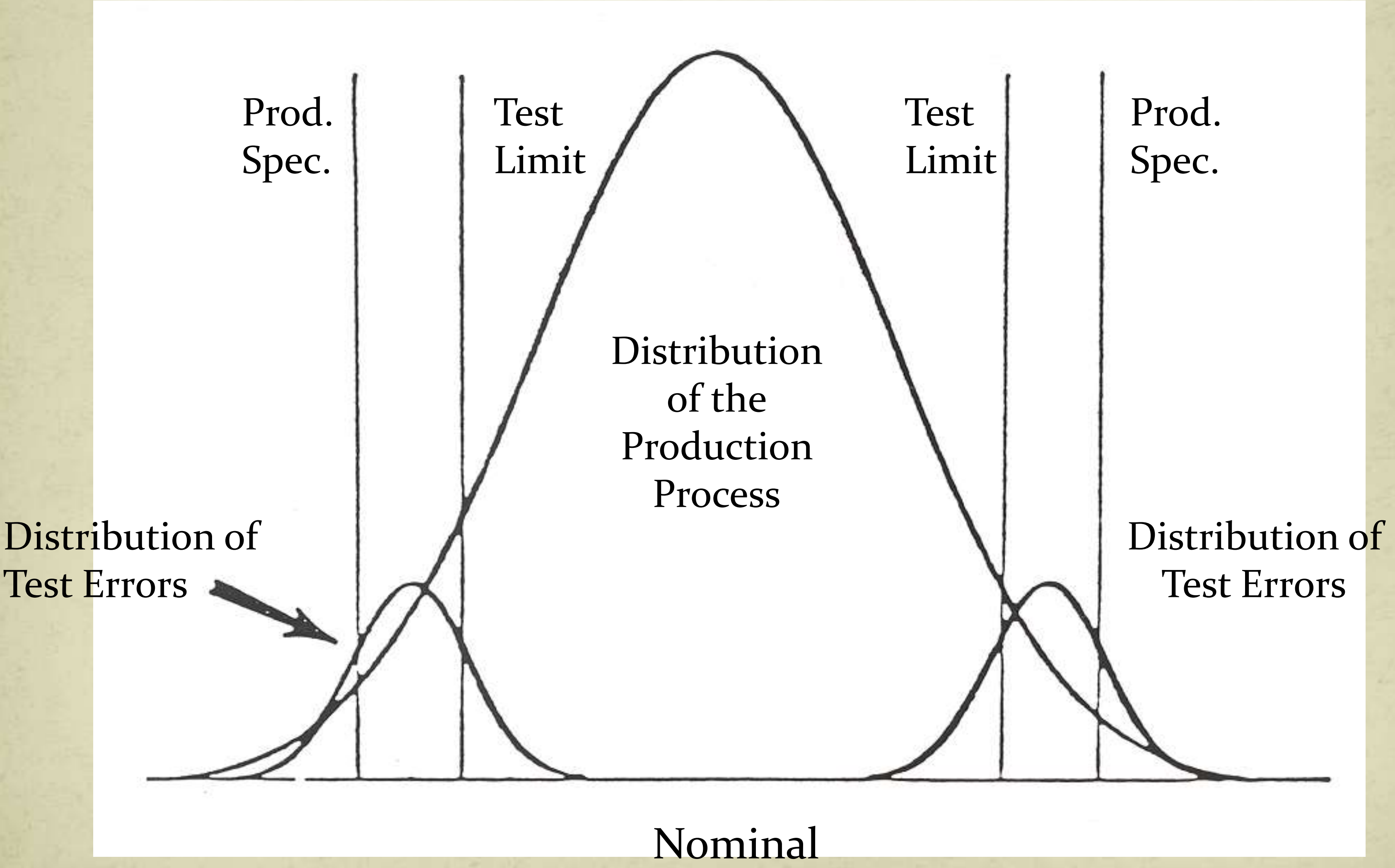
Measurement Error Distributions Added



Definitions:

A = Probability a test fails given a good DUT.
B = Probability a test passes given a bad DUT.
P = Probability the DUT is truly bad.

Product Specifications and Production Test Limits



Estimating Measurement Error

- Different ways to assess the performance of a measurement process
  - Classical “Gauge R&R”
  - ISO (International Standards Organization)
  - DOE/ANOVA

Classical Method

- Typically 2-factor random effects model without interaction
  - "Parts" and "Operators" plus replication
- Uses ranges to estimate sources of variation
  - Not as precise
  - Normality required since the factor (d_2) to convert from a range to a standard deviation depends on it
- Typically no interval estimates
- Sample sizes tend to be small (e.g. 10 parts, 3 operators, 3 reps)

ISO Method

- International Standards Organization
- Evaluates measurement uncertainty using a mathematical model
- Law of propagation of uncertainty (1st order Taylor series)
- Usually no measurements are taken directly on the “measureand” (i.e. the response of interest)
- To avoid understating the estimated uncertainty, all major sources of variation must be included

DOE/ANOVA Method

- Measurement error estimated by direct observation on the response
- ANOVA model is general in that it can accommodate multiple factors in different designs
- Lends itself to both point and interval estimates of the measurement variation
- Fixed or random effects

Definitions

- Measurement:

  The value obtained from measuring the device under test (DUT) a single time.

  Measurement = True Value + Bias + Measurement Error

- True Value:

  The accepted correct value of the parameter being measured.

- Bias:

  The difference between the average of repeated measurements of a parameter on a single DUT, and the true value.

Definitions

- Measurement Error:

  The deviation between an individual measurement of a parameter on a DUT, and the average of many repeated measurements of the same DUT. The standard deviation of measurement error is often called *precision*.

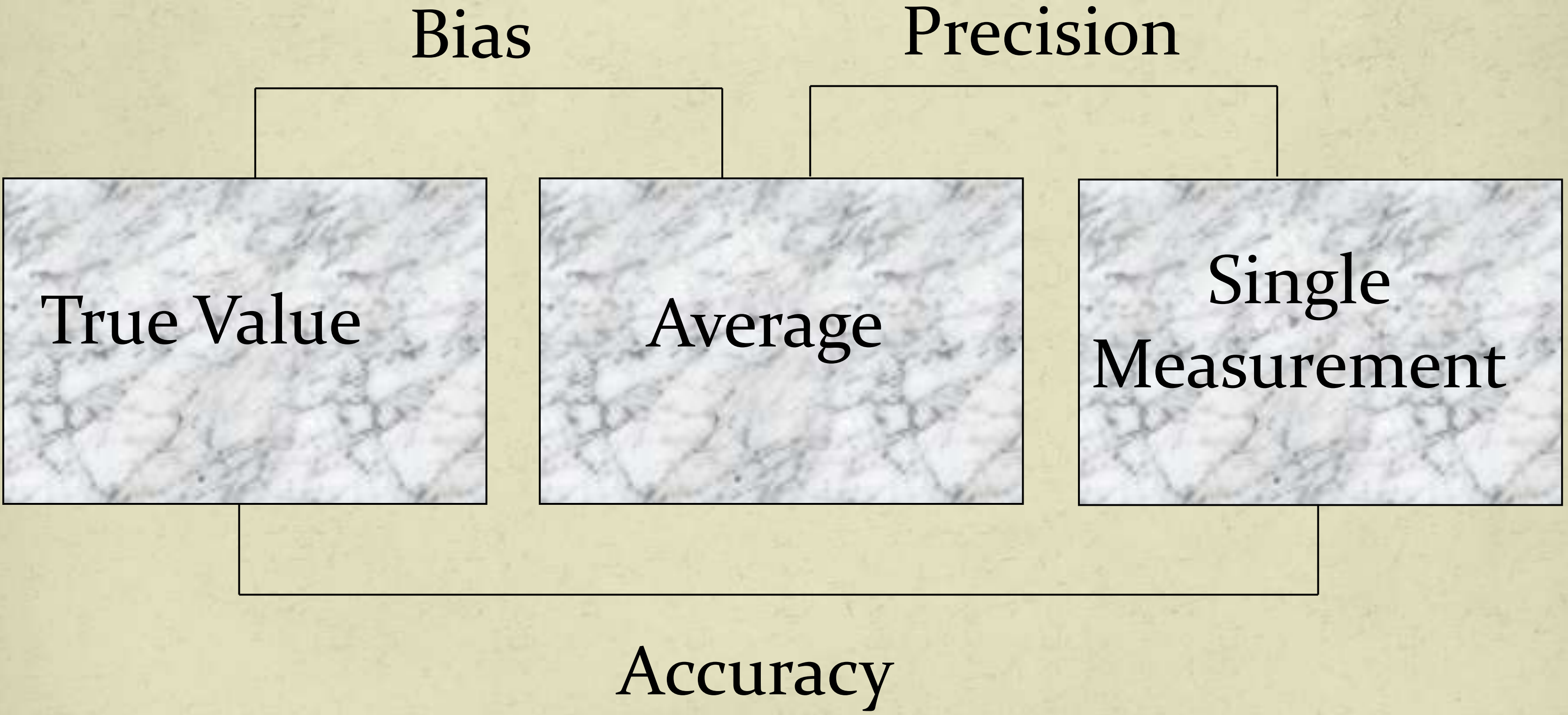
- Standard:

  A measurement standard for which the true value is known.

- Accuracy:

  The difference between an individual measurement and the true value. For a measurement to be accurate, both the bias and the measurement error must be small.

Precision vs. Accuracy



Bias

Precision

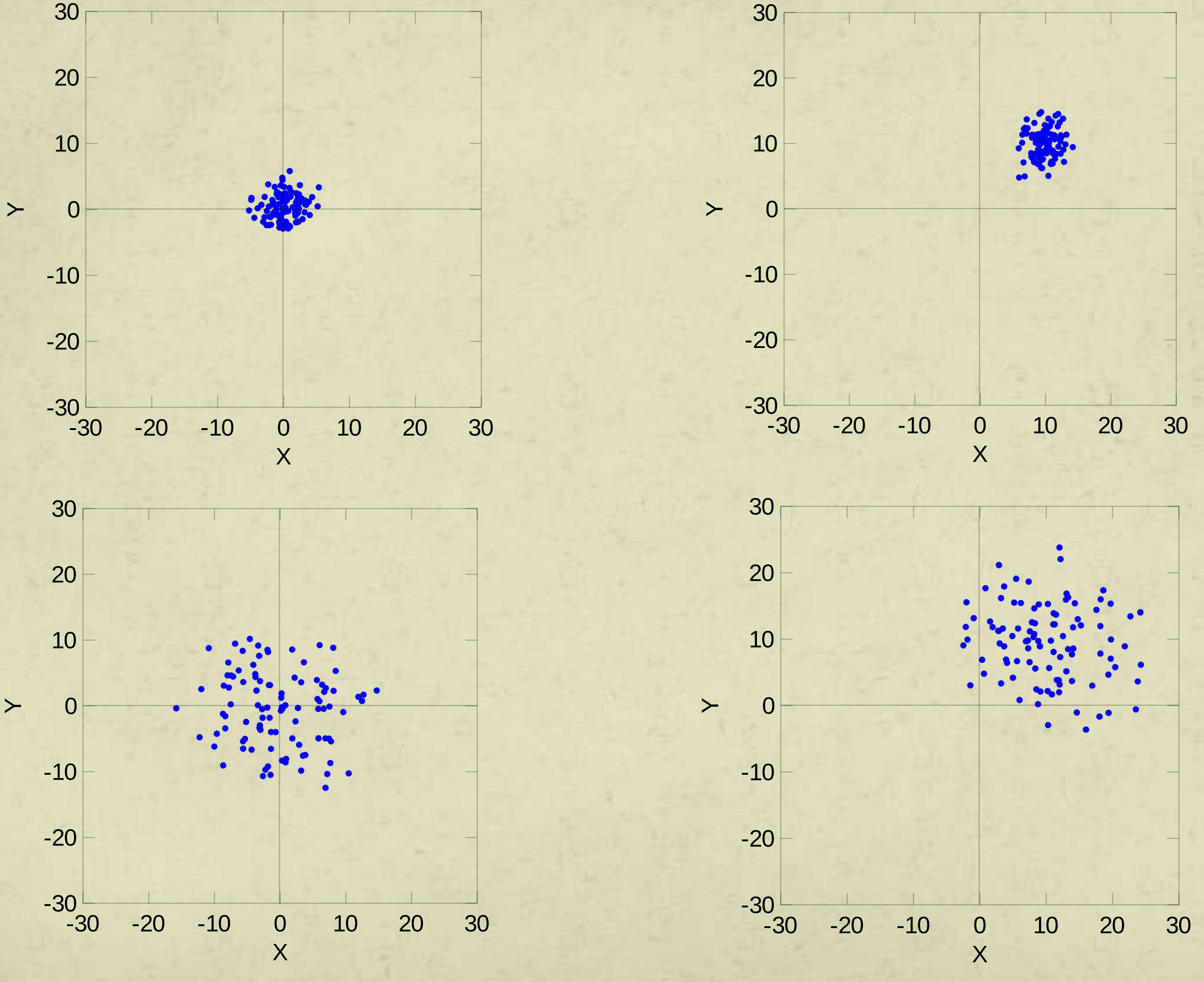
True Value

Average

Single Measurement

Accuracy

Precision and Accuracy



More Definitions

- Reproducibility:

  The variation in measurements due to all sources of measurement variation *except* repeatability.

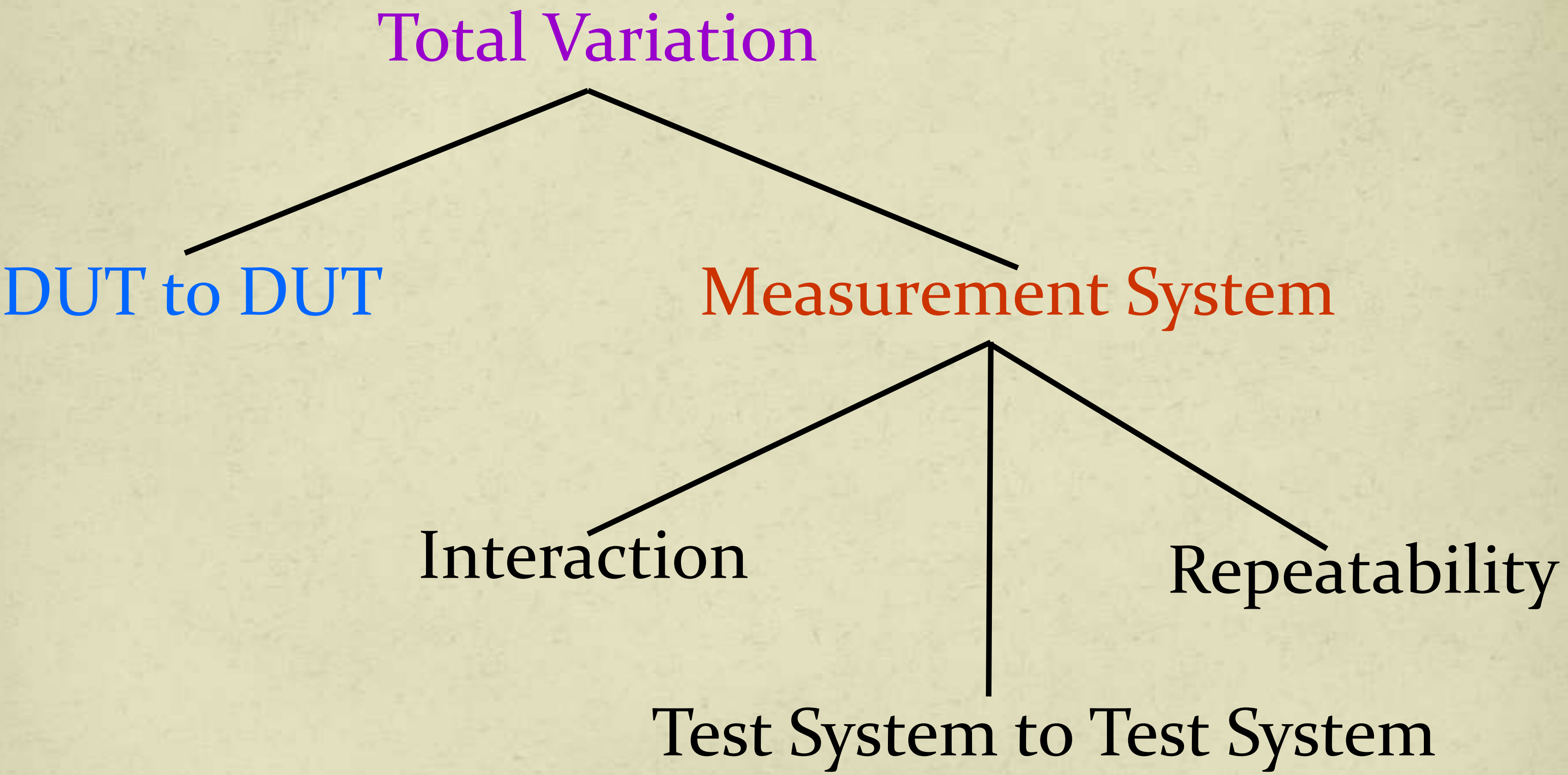
- Repeatability:

  The variation of repeated measurements using a single test system and an individual DUT.

- Measurement System Variation:

  Variation arising from *all* sources in the measurement system. For example, variation among test systems, among fixtures, among repeated measurements.

Partitioning of Variation (Variance Components)



Total Variation

DUT to DUT

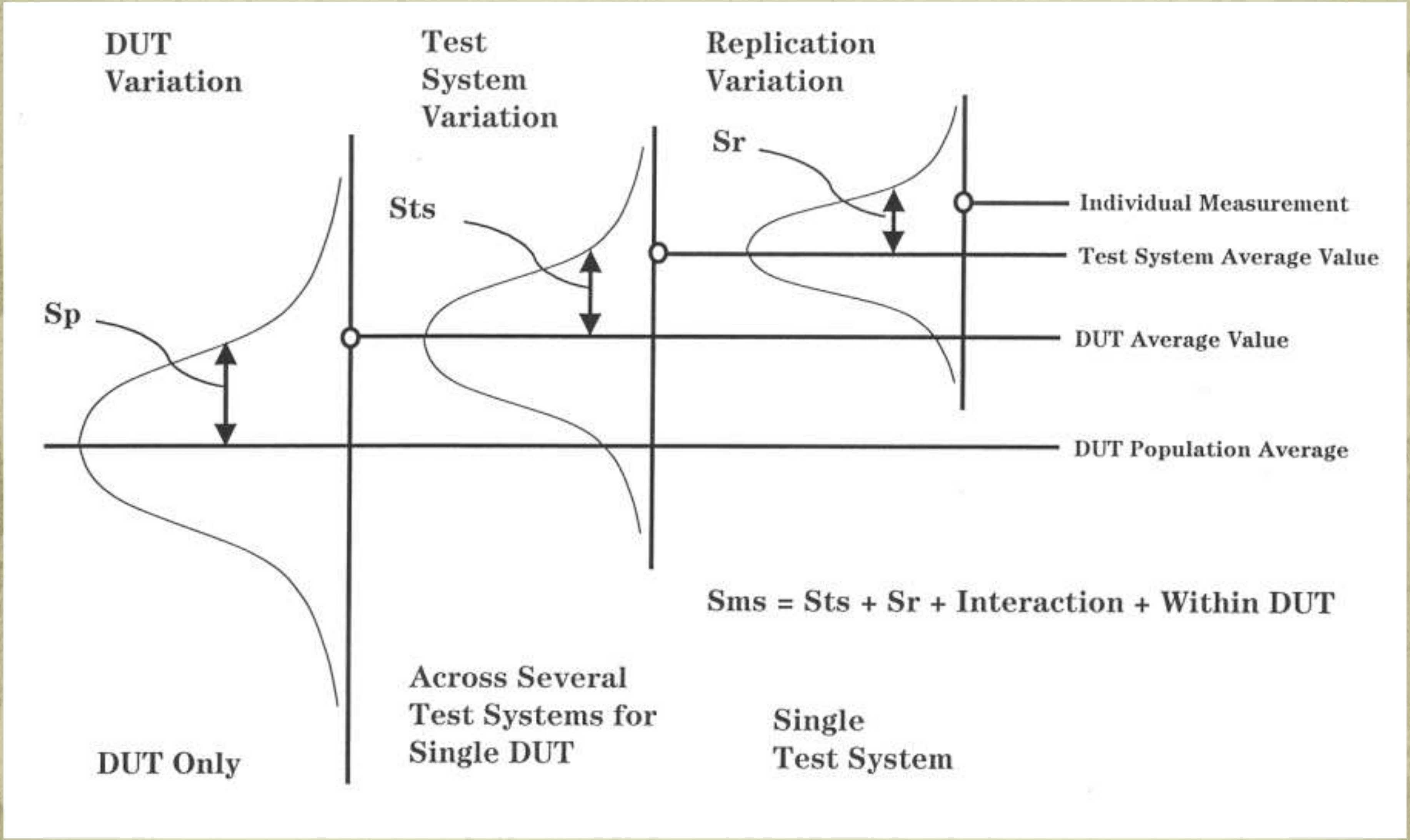
Measurement System

Interaction

Repeatability

Test System to Test System

Variance Decomposition



Sms = Sts + Sr + Interaction + Within DUT

More Definitions

- Repeatability
  - Static:
    Measurements taken in succession without changing the setup.
  - Dynamic:
    The DUT is removed from the measurement device and repeated measurements are separated in time.

- Within DUT variability
  - If present, included in repeatability
  - To resolve, substitute a standard for the DUT and isolate the pure repeatability of the measurement system

Example - Head Test

- A division of Hewlett-Packard that made tape drives for computer backup designed a drive with higher storage capacity that used a new head design.
- The head is the part of the drive that makes contact with the tape and magnetically records information on the tape.
- The heads are manufactured by an outside supplier and must be tested to make sure they meet performance specifications.
- To do this, a head tester was developed by HP and installed at the supplier’s production facility to test the heads prior to shipment.

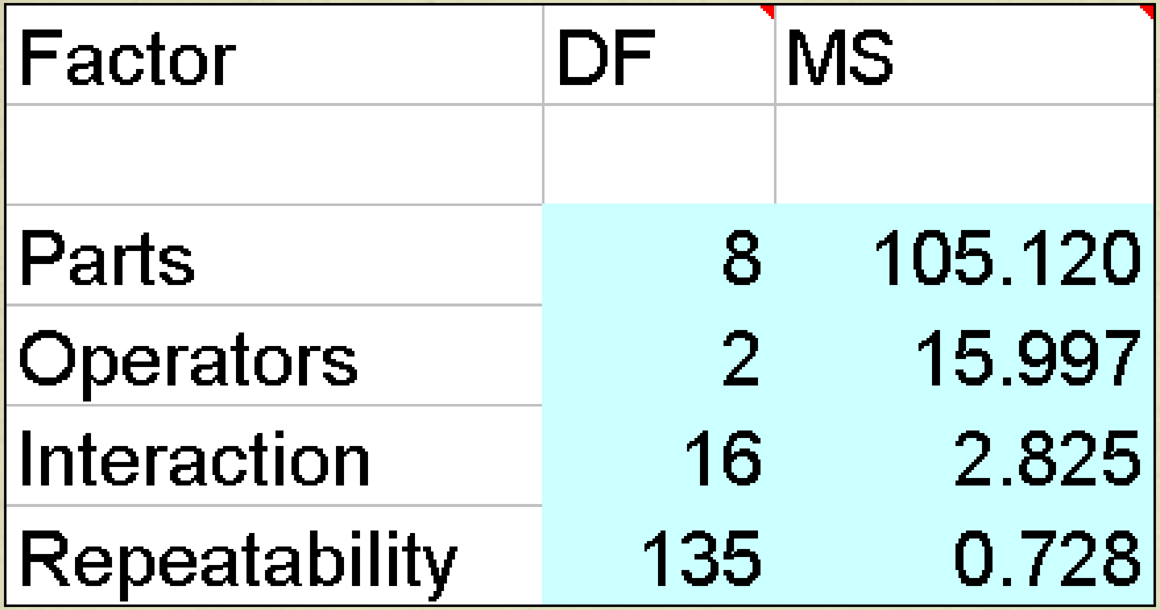
Example - Head Test

- Nine heads were selected at random from the manufacturing process that produces the heads.
- Three different tape cartridges were selected at random to record measurements in the head tester.
- Thus, heads are “parts”, and the tapes are “operators”.
- Six repeated measurements were made on each head/tape combination for a total of 162 points in the design.
- The data were used to determine whether the tester had sufficient performance to discriminate among heads.

Example - Head Test

- 16 response variables were measured, one of which is “reverse resolution”
- In the process of writing information to magnetic tape, voltage is measured at two different frequencies.
- The reverse resolution is a ratio of these two voltages when the drive writes in the reverse direction.
- Ideally, this ratio should be 100%. The specification limits are from 90% to 110%.

ANOVA Table



Factor	DF	MS
Parts	8	105.120
Operators	2	15.997
Interaction	16	2.825
Repeatability	135	0.728

ANOVA Two-factor Model

$$Y_{ijk} = \mu + P_i + O_j + (PO)_{ij} + E_{ijk}$$

i=1,….,I; j=1,….,J; k=1,….K

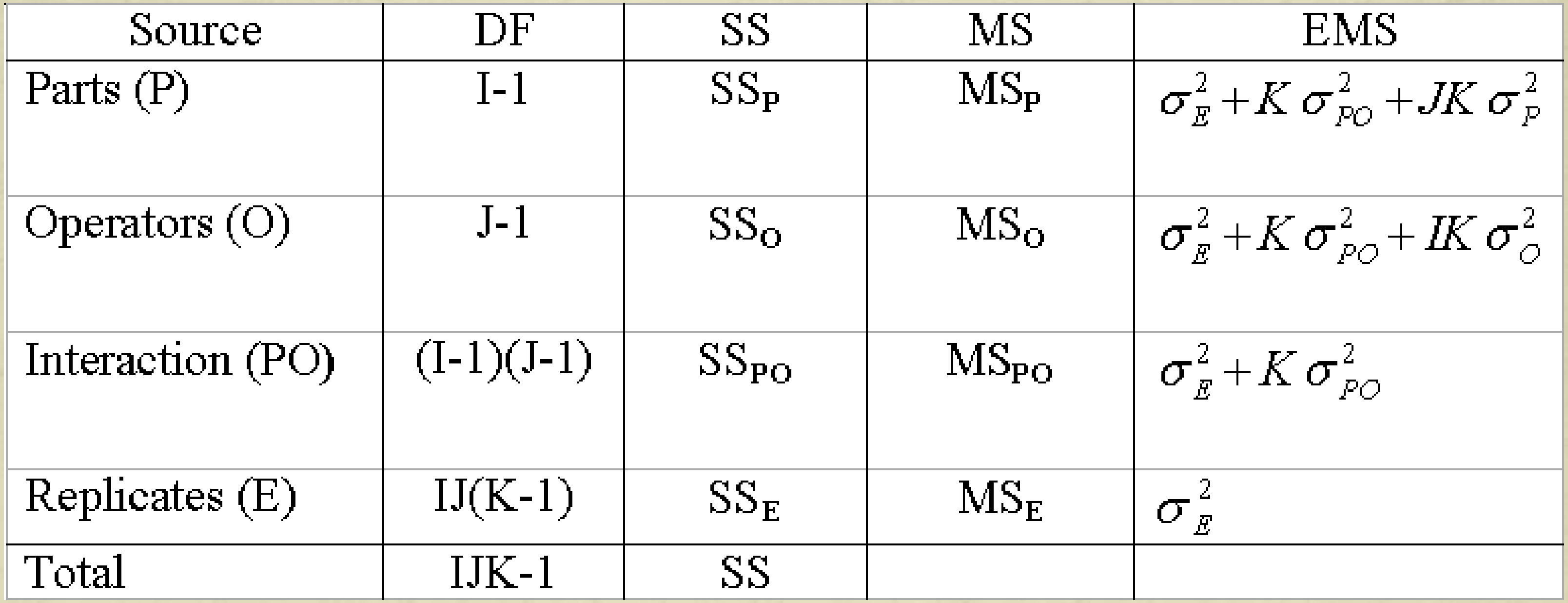
μ is a constant

P_i, O_j, $(PO)_{ij}$, E_{ijk} are iid normal with means of zero and

variances $\sigma_P^2,\ \sigma_O^2,\ \sigma_{PO}^2,\ \sigma_E^2$

These are called *variance components*

Expected Mean Squares



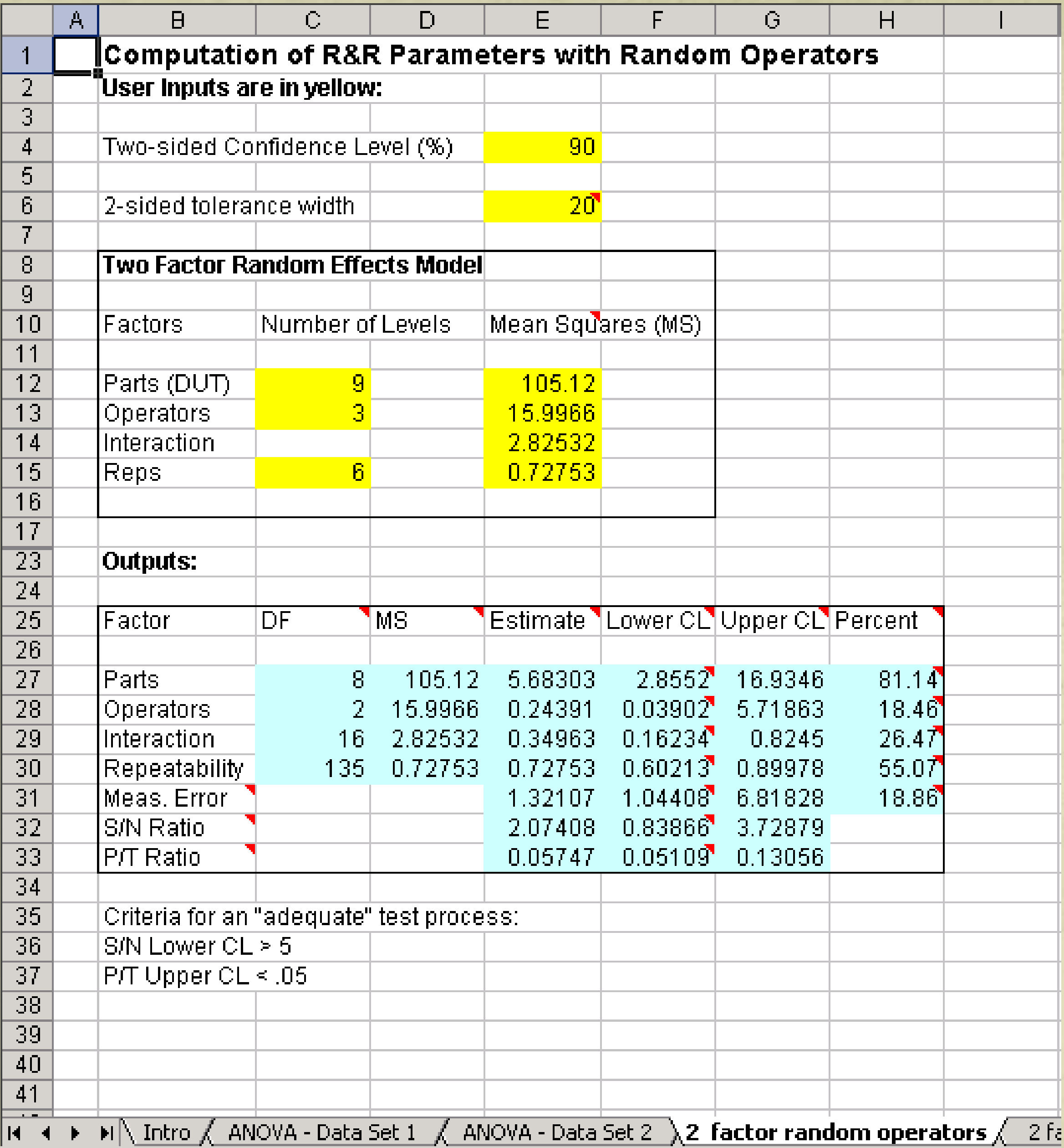
Source	DF	SS	MS	EMS
Parts (P)	I-1	SS_P	MS_P	$\sigma_E^2 + K\sigma_{PO}^2 + JK\sigma_P^2$
Operators (O)	J-1	SS_O	MS_O	$\sigma_E^2 + K\sigma_{PO}^2 + IK\sigma_O^2$
Interaction (PO)	(I-1)(J-1)	SS_{PO}	MS_{PO}	$\sigma_E^2 + K\sigma_{PO}^2$
Replicates (E)	IJ(K-1)	SS_E	MS_E	σ_E^2
Total	IJK-1	SS		

Measures of Adequacy

Measure	Symbol
Repeatability	σ_E^2
Measurement Error	$\gamma = \sigma_O^2 + \sigma_{PO}^2 + \sigma_E^2$
S/N Ratio	$\omega = \sqrt{\sigma_P^2 / \gamma}$
P/T Ratio	$\eta = \sqrt{\gamma}\,/(USL - LSL)$

Criteria

- S/N Ratio
  - Criterion: Lower bound of a 90% confidence interval for $\omega > 5$
- P/T Ratio
  - Criterion: Upper bound of a 90% confidence for $\eta < .05$



Computation of R&R Parameters with Random Operators

User Inputs are in yellow:

Two-sided Confidence Level (%)	90
2-sided tolerance width	20

Two Factor Random Effects Model

Factors	Number of Levels	Mean Squares (MS)
Parts (DUT)	9	105.12
Operators	3	15.9966
Interaction		2.82532
Reps	6	0.72753

Outputs:

Factor	DF	MS	Estimate	Lower CL	Upper CL	Percent
Parts	8	105.12	5.68303	2.8552	16.9346	81.14
Operators	2	15.9966	0.24391	0.03902	5.71863	18.46
Interaction	16	2.82532	0.34963	0.16234	0.8245	26.47
Repeatability	135	0.72753	0.72753	0.60213	0.89978	55.07
Meas. Error			1.32107	1.04408	6.81828	18.86
S/N Ratio			2.07408	0.83866	3.72879	
P/T Ratio			0.05747	0.05109	0.13056	

Criteria for an "adequate" test process:
S/N Lower CL > 5
P/T Upper CL < .05

Extensions

- Randomization
- Alternative designs
- Fixed vs. random operators
- Sample size
- Comparison across time/location

Randomization

- Completely randomized design provides protection from the effects of uncontrolled variables which may be changing
- However, a random sequence is likely to require the same DUT to be measured multiple times in a short time period. If heat affects the measurement, a wait time has to be introduced which lengthens the total time to run the experiment
- Regardless of the degree of randomization, probably best to spread the repeated measurements out in time (dynamic)

Alternative Designs

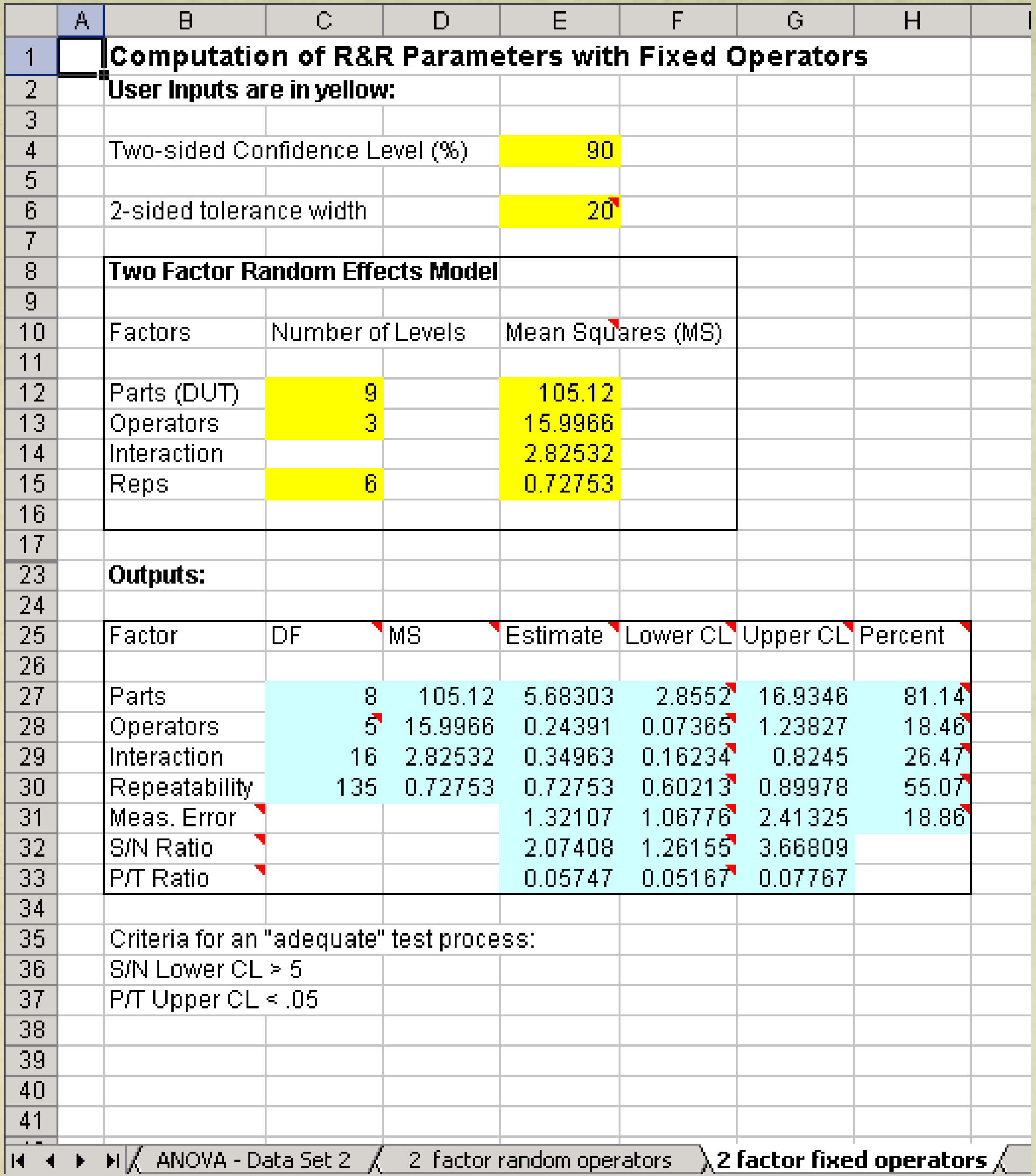
- Single factor
  - Multiple DUTs and repeated measurements on each DUT
- Two factor crossed (parts, operators)
- Three factor crossed (parts, operators, fixtures)
- Nested
  - DUTs are circuit boards. Measurements vary by location and so location is nested within board.
  - Different test systems and fixtures, but fixtures cannot be moved among systems. Thus, fixture is nested within system.
- Nested factorial

Fixed vs. Random Operators

- A factor is considered random if its levels are a random sample from the population of all levels
- A factor is fixed if the levels are a non-random sample or if the levels consist of the entire population
- In this context DUTs should be representative of the whole population which is likely large, so DUT is a random factor

Fixed vs. Random Operators

- The other factors (e.g., operators) may be considered random or fixed depending on the situation
- Methods for computing CI's depend on whether effects are fixed or random
- Requires a modification for the degrees of freedom for the fixed effect



Computation of R&R Parameters with Fixed Operators

User Inputs are in yellow:

Two-sided Confidence Level (%)	90
2-sided tolerance width	20

Two Factor Random Effects Model

Factors	Number of Levels	Mean Squares (MS)
Parts (DUT)	9	105.12
Operators	3	15.9966
Interaction		2.82532
Reps	6	0.72753

Outputs:

Factor	DF	MS	Estimate	Lower CL	Upper CL	Percent
Parts	8	105.12	5.68303	2.8552	16.9346	81.14
Operators	5	15.9966	0.24391	0.07365	1.23827	18.46
Interaction	16	2.82532	0.34963	0.16234	0.8245	26.47
Repeatability	135	0.72753	0.72753	0.60213	0.89978	55.07
Meas. Error			1.32107	1.06776	2.41325	18.86
S/N Ratio			2.07408	1.26155	3.66809	
P/T Ratio			0.05747	0.05167	0.07767	

Criteria for an "adequate" test process:

S/N Lower CL > 5

P/T Upper CL < .05

Sample Size

- Usually impacted by practical considerations
- The “classic” 10 parts, 3 operators and 3 reps probably gives poor precision on the variance component estimates

Sample Size Recommendations

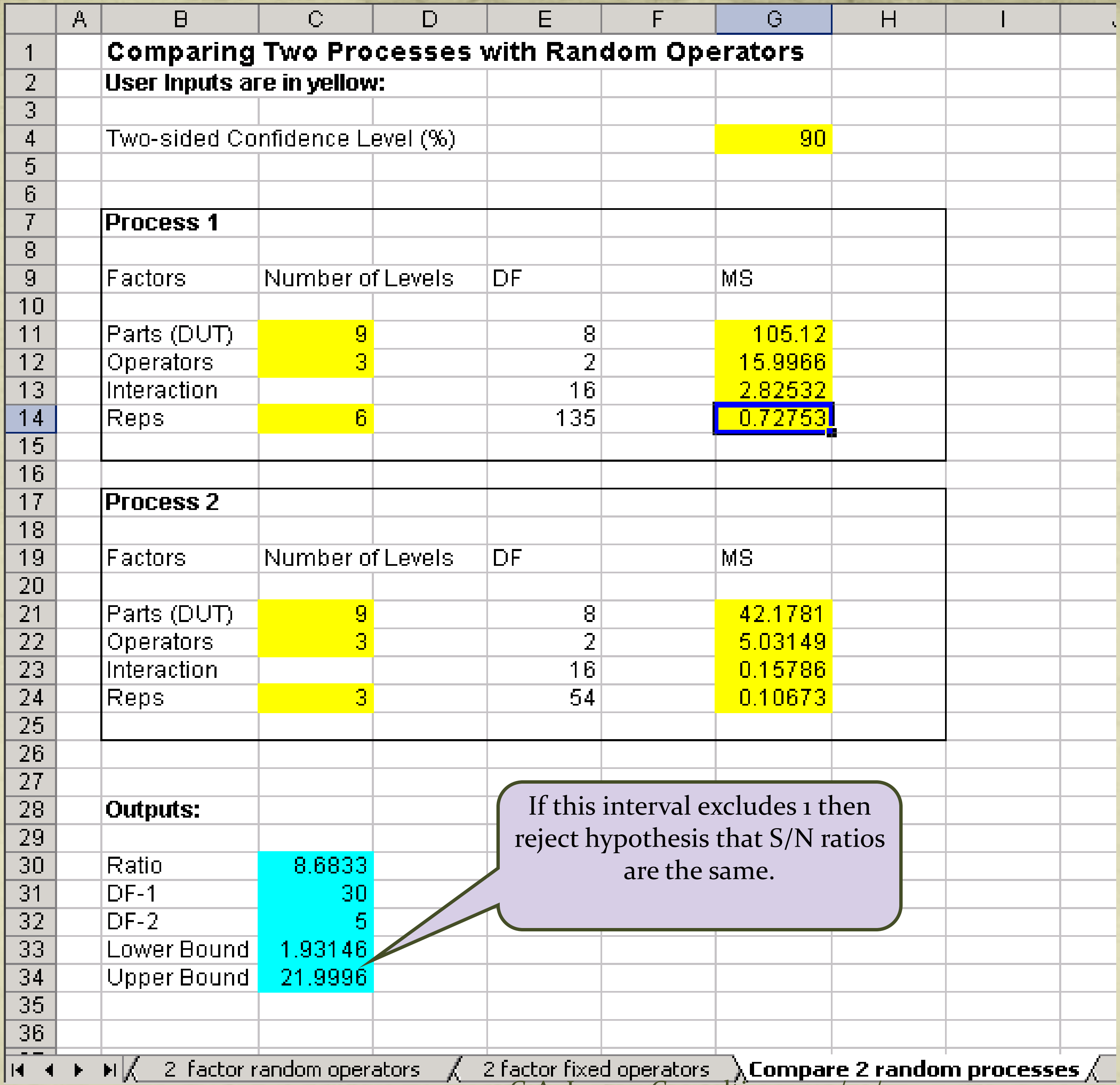
- Single factor design
  - Use a lot more DUTs than reps and get the total design size > 100
- Two factor design
  - Use 2 reps
  - Use at least 6 levels for the other factors (if random)
  - Use at least 30 DUTs if objective is to estimate the manufacturing process capability in the absence of measurement error
  - Depending on the application, a 12x6x2 factorial design might be a good starting point

Comparisons Across Time/Location

- Compare “before and after” to verify an improvement of the measurement process
- Compare a competitor’s test process to an Agilent solution
- Comparison of same measurement process used at two different locations
- Compare using the ratio of S/N ratios
- Confidence interval can be constructed using Cochran’s interval (Satterthwaite)

Example - Before vs. After Head Test

- In our head test example, the initial MSA revealed poor head tester performance on 2 of 16 test variables
- After some changes, a 2nd MSA was done to determine whether the tester had actually improved
- The 2nd MSA had the same 9 heads, 3 tapes, and 3 reps
- Results showed significant improvement and the performance on all 16 test variables was now adequate
- Several additional testers were produced, qualified, and installed in the head supplier's production process



Comparing Two Processes with Random Operators

User Inputs are in yellow:

Two-sided Confidence Level (%) 90

Process 1

Factors	Number of Levels	DF	MS
Parts (DUT)	9	8	105.12
Operators	3	2	15.9966
Interaction		16	2.82532
Reps	6	135	0.72753

Process 2

Factors	Number of Levels	DF	MS
Parts (DUT)	9	8	42.1781
Operators	3	2	5.03149
Interaction		16	0.15786
Reps	3	54	0.10673

Outputs:

Ratio	8.6833
DF-1	30
DF-2	5
Lower Bound	1.93146
Upper Bound	21.9996

If this interval excludes 1 then reject hypothesis that S/N ratios are the same.

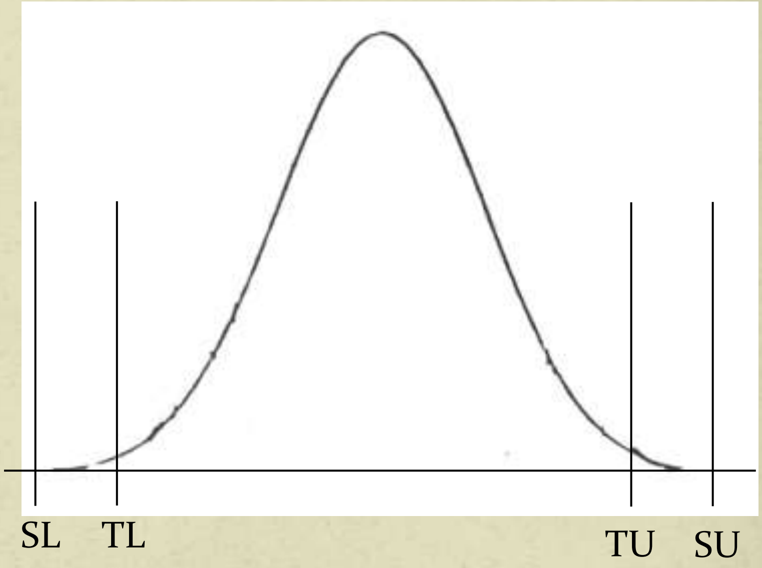
Estimating the Cost of Measurement Error

- False failures and missed faults
- Bi-variate normal integration
- Setting test limits
- Simulation of more complex test processes

Estimating the Cost of Measurement Error

Estimating the two test errors
- False failures
- Missed faults

	FAIL	PASS	
GOOD	FF P(FAIL|GOOD) α	P(PASS|GOOD) 1- α	P(GOOD)
BAD	P(FAIL|BAD) 1- β	MF P(PASS|BAD) β	P(BAD)
	P(FAIL)	P(PASS)	



Estimating the Test Errors

Y = True Value ~ N(μ_P, σ_P^2)

Product Specs (SL, SU)

X = Measured Value

X = Y + E

E ~ N(0, σ_m^2)

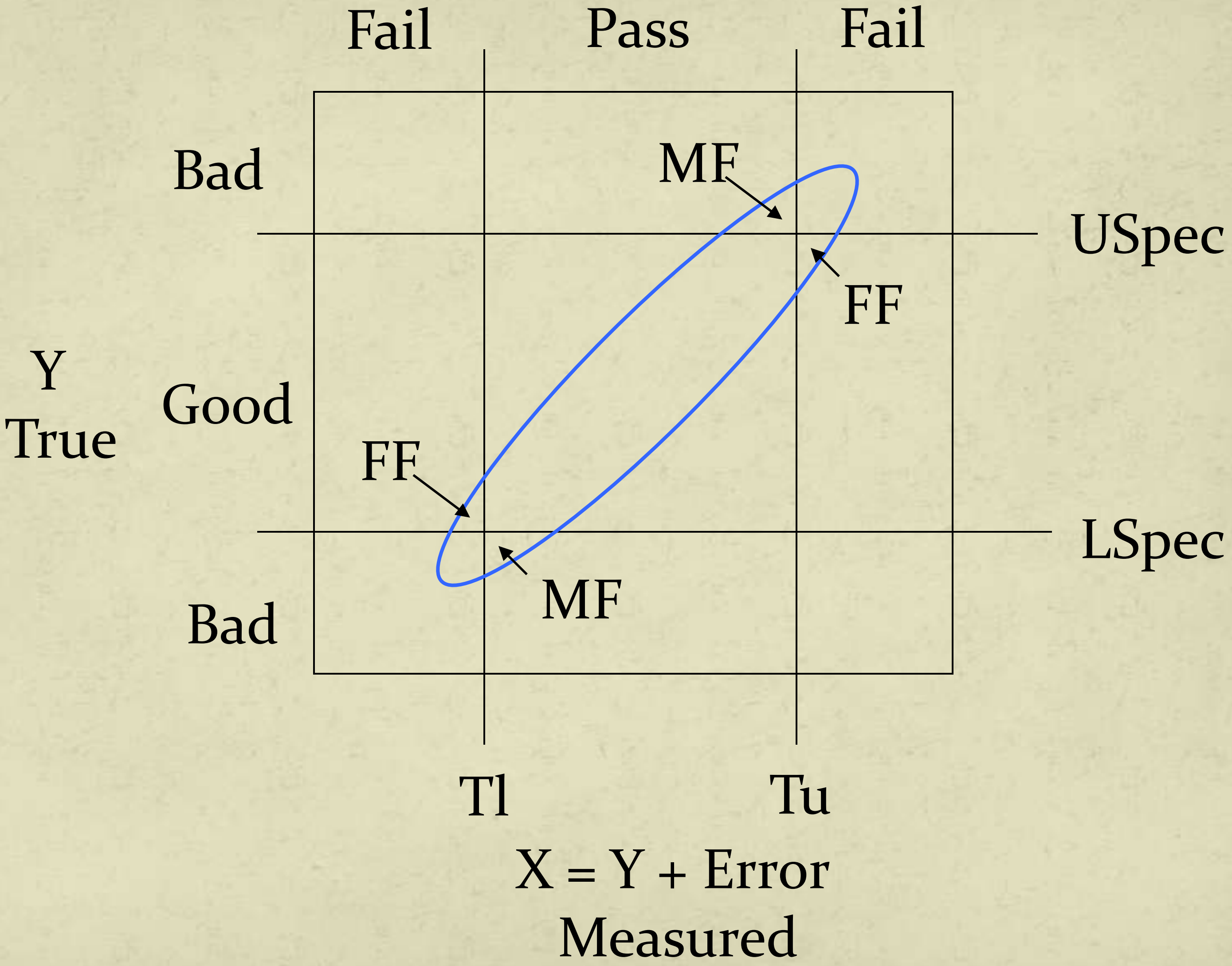
Test Limits (TL, TU)

f(x,y) = joint pdf

f(x) = marginal pdf for x

f(y) = marginal pdf for y

Two-Sided Limits



Fail Pass Fail

Bad

MF

USpec

FF

Y

True

Good

FF

LSpec

MF

Bad

Tl Tu

X = Y + Error

Measured

Estimating the Test Errors

P(BAD) = $\int_{-\infty}^{SL} f(y)dy + \int_{SU}^{+\infty} f(y)dy$

P(FAIL) = $\int_{-\infty}^{TL} f(x)dx + \int_{TU}^{+\infty} f(x)dx$

P(PASS, BAD) = Joint $\beta = \int_{-\infty}^{SL}\int_{TL}^{TU} f(x,y)dxdy + \int_{SU}^{+\infty}\int_{TL}^{TU} f(x,y)dxdy$

P(FAIL|GOOD) = $\alpha = \dfrac{\text{P(FAIL) - P(BAD) + P(PASS, BAD)}}{\text{1 - P(BAD)}}$

P(PASS|BAD) = $\beta = \dfrac{\text{P(PASS, BAD)}}{\text{P(BAD)}}$

P(FAIL, GOOD) = Joint $\alpha = \alpha$ (1 - P(BAD))

Test Error Criteria

- Earlier a pair of metrics were suggested to quantify the adequacy of the measurement process: S/N and P/T.
- Criteria have also been suggested for the test error rates. The idea is to compare the computed error rates with those that would be expected by random chance. If the computed error rates are lower, there is evidence that the measurement process is a better discriminator than pure chance.
- Two indices are defined:

$$FF_{index} = \frac{\text{Joint } \alpha}{P(BAD)*(1 - P(BAD))}$$

$$MF_{index} = \frac{\text{Joint } \beta}{P(BAD)*(1 - P(BAD))}$$

- Values < 1 suggest the measurement process is capable.

Setting Production Test Limits

Criteria is to minimize the sum of the test error costs

$$\text{Cost Ratio} = \text{CR} = \frac{\text{Cost of MF}}{\text{Cost of FF}}$$

$$\text{SN} = S_P/S_m$$

$$K = \frac{(SU - SL)}{2*S_P}$$

$$Z = \text{PROBIT}(1 - 1/(1 + \text{CR}))$$

$$B = \frac{Z\sqrt{1 + SN^2} - K}{SN}$$

$TL = SL + B*S_m$ and $TU = SU - B*S_m$

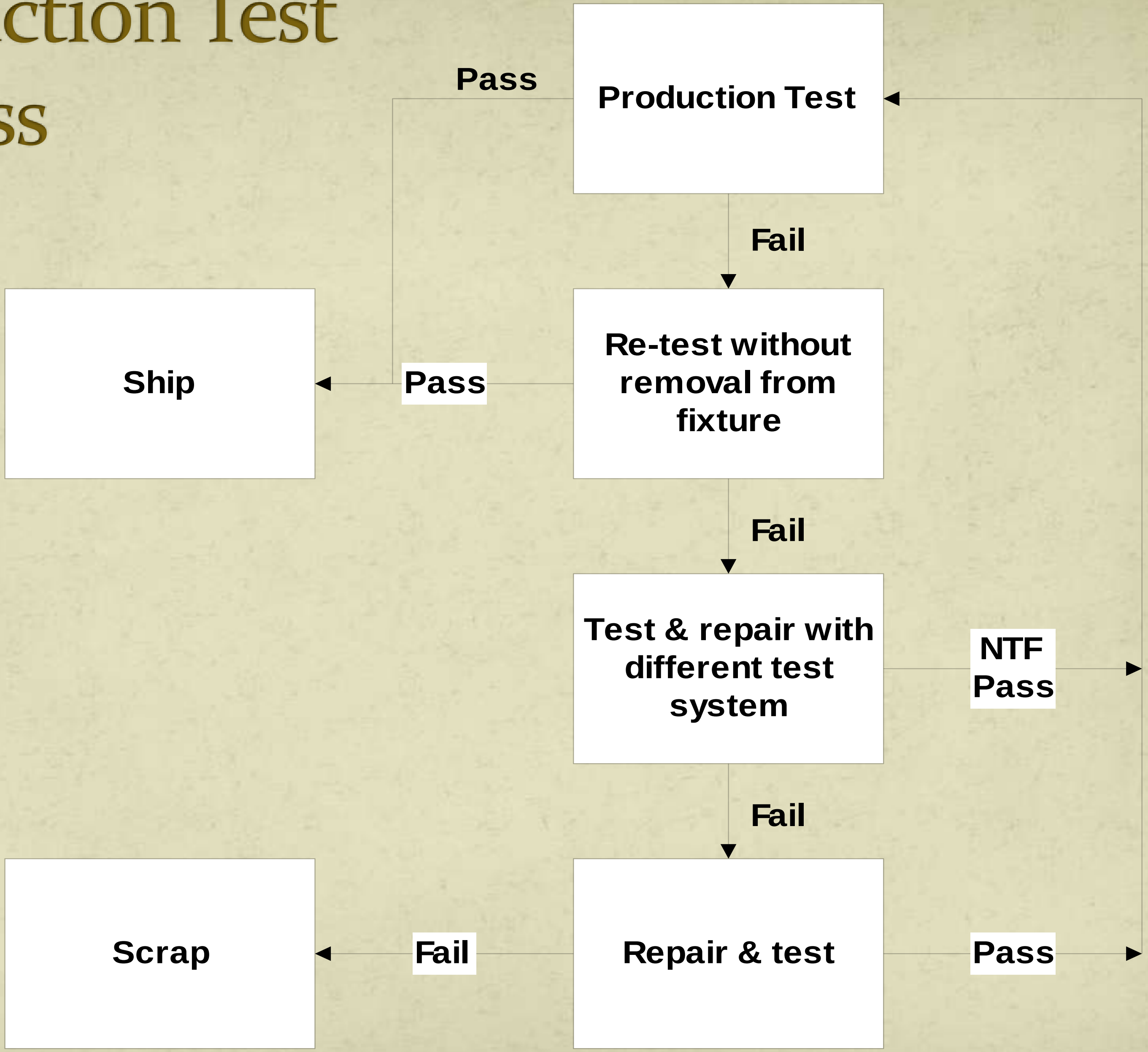
Setting Production Test Limits

- Production test limits can also be set so that a target test error rate is achieved
- For example, the limits and the false failure rate can be determined such that the missed fault rate is equal to 1%
- This is done with an iterative application of the numerical integration described earlier

Simulation of More Complex Test Processes

- So far we have been discussing a single test instance on a single test parameter
- The test process often involves re-test, and/or test and repair loops.
  - Multiple measurements may be averaged.
- There may be many parameters tested on the same DUT
- For these situations, FF, MF and reject rates are simulated.

Production Test Process



Production Test

Fail

Re-test without removal from fixture

Pass

Ship

Fail

Test & repair with different test system

NTF Pass

Fail

Repair & test

Pass

Fail

Scrap

Usage in Manufacturing Environment

- Software application to aid sales
- Qualification of test solution prior to implementation
- Characterize/improve existing test process
- Monitor stability of the measurement process

References

- “Measurement Systems Analysis”, Automotive industry Action Group (AIAG), 2nd Edition, 1995.
- “Confidence Intervals on Measures of Variability in R&R Studies”, Burdick and Larsen, *Journal of Quality Technology*, 29(3), 1997.
- “Analysis of a Two-Factor R&R Study With Fixed Operators”, Dolezal, Burdick and Birch, *Journal of Quality Technology, 30(2), 1998.*
- “Two-Way Random-Effects Analyses and Gauge R&R Studies”, Vardeman and Van Valkenburg, *Technometrics*, 41(3), 1999.
- “Comparing Variability of Two Measurement Processes using R&R Studies”, Burdick, Allen and Larsen, *Journal of Quality Technology*, 34(1), 2002.
- “The Economic Impact of Measurement Error”, Mader, Prins and Lampe, *Quality Engineering*, 11(4), 1999.
- “On Setting Test Limits Relative to Specification Limits”, Grubbs and Coon, *Industrial Control*, 10(5), 1954.

References

- “Measurement System Analysis - the Usual Metrics can be Non-Informative”, Larsen, *Quality Engineering*, 15(2), 2002.
- “Measurement System Analysis in a Production Environment with Multiple Test Parameters”, Larsen, *Quality Engineering*, 16(2), 2003.
- “A Review of Methods for Measurement System Capability Analysis”, Burdick, Borror and Montgomery, *Journal of Quality Technology*, 35(4), 2003.
- “Confidence Intervals for Misclassification Rates in a Gauge R&R Study”, Burdick, Park, Montgomery, and Borror, *Journal of Quality Technology*, 37(4), 2005.
- Design and Analysis of Gauge R&R Studies, Burdick, Borror and Montgomery, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA, 2005.
- “Capability Measures for Measurement Systems Analysis”, Larsen, in Encyclopedia of Statistics in Quality and Reliability, Ruggeri, F., Kenett, R. and Faltin, F.W. (eds). John Wiley & Sons Ltd, Chichester, UK, pp 1070-1074. 2007.

References on Related Topics

MSA on Pass/Fail Data

- "Gauge Capability for Pass-Fail Inspection", Boyles, *Technometrics*, 43(2), 2001.

MSA when Measurements are Destructive

- "Improving and Applying Destructive Gauge Capability", Bergeret, Maubert, Sourd, and Puel, *Quality Engineering*, 14(1), 2001.
- "Gauge R&R Studies for Destructive Measurements", De Mast and Trip, *Journal of Quality Technology*, 37(1), 2005.

MSA when Metric of Interest is Derived from Other Variables

- "Approving Measurement Systems when Using Derived Values", Majeske and Gearhart, *Quality Engineering,* 18(4), 2006.

Effect of Measurement Error on Control Charts

- "Effect of Measurement Error on Shewhart Control Charts", Linna and Woodall, *Journal of Quality Technology*, 33(2), 2001.