Test Engineering Principles

Introduction

This article is an attempt to "condense" some things that I've learned about Test Engineering from 36 years of experience in the industry and after reading many books and spending a lot of time researching and working in the testing of manufactured electronic assemblies. It is not possible to go into all details but hopefully this article can be used as a guideline for further research and consideration.

Classes or degrees in Electronics Test Engineering and Management are not easy to find. It is not a subject that is really taught in schools. It is easy to find information on how to use a sensor or how to program an oscilloscope. Not so easy to find out *why* we test or why some of the statistics are used a certain way or how to apply them. What it all comes down to is how to balance the cost of test with the effects of poor quality.

There are 4 books that have had the most influence on my view of Test Engineering:

1. The Economics of Automatic Testing, 2nd Edition by Brendan Davis
2. Test Engineering by Patrick D.T. O'Connor
3. Improving Product Reliability by Mark Levin and Ted Kalal
4. World Class Quality, Using Design of Experiments, 2nd Edition by Keki and Adi Bhote

Most projects follow a basic outline for test development:

Requirements Analysis
Test Plan Development
Writing Test Cases
Verification Testing
Debugging
Data Collection and Analysis
Test Reporting
Refinement and continuous improvement (kaizen)

Test Development – Measurement Systems Analysis

When developing a testing system, it is important to understand where sources of variation come from. This starts with Measurement Systems Analysis.

A measurement system is the collection of instruments or gages, standards, operations, methods, fixtures, software, personnel, environment and assumptions used to quantify a unit of measure or assessment of the feature characteristic being measured; It is the complete process used to obtain measurements. *[1]

1. The accuracy and precision necessary in the measurement system starts by comparing the criteria for product acceptance against a manufacturer's specifications of the instrumentation being used. See my article at https://testview.wordpress.com/2014/03/10/how-accurate-does-your-measurement-need-to-be-revised/.
2. Once a testing system is ready, a design review of the system may be in order with the Engineering Team.
3. I will refer to the measurement system as the "testing system".
4. We use the instrument manufacturer's specifications to understand how the equipment will perform in these areas: *[2]
    i. Discrimination as the smallest readable unit.
    ii. Resolution as the smallest input that results in a usable output of measure.
    iii. Sensitivity of the instrument to variation.
    iv. Accuracy as the closeness of the reading to an accepted reference value.
    v. Bias as the difference between the observed average of measurements and a reference value.
    vi. Stability as the change in bias over time.
    vii. Linearity as the change in bias over the normal operating range.

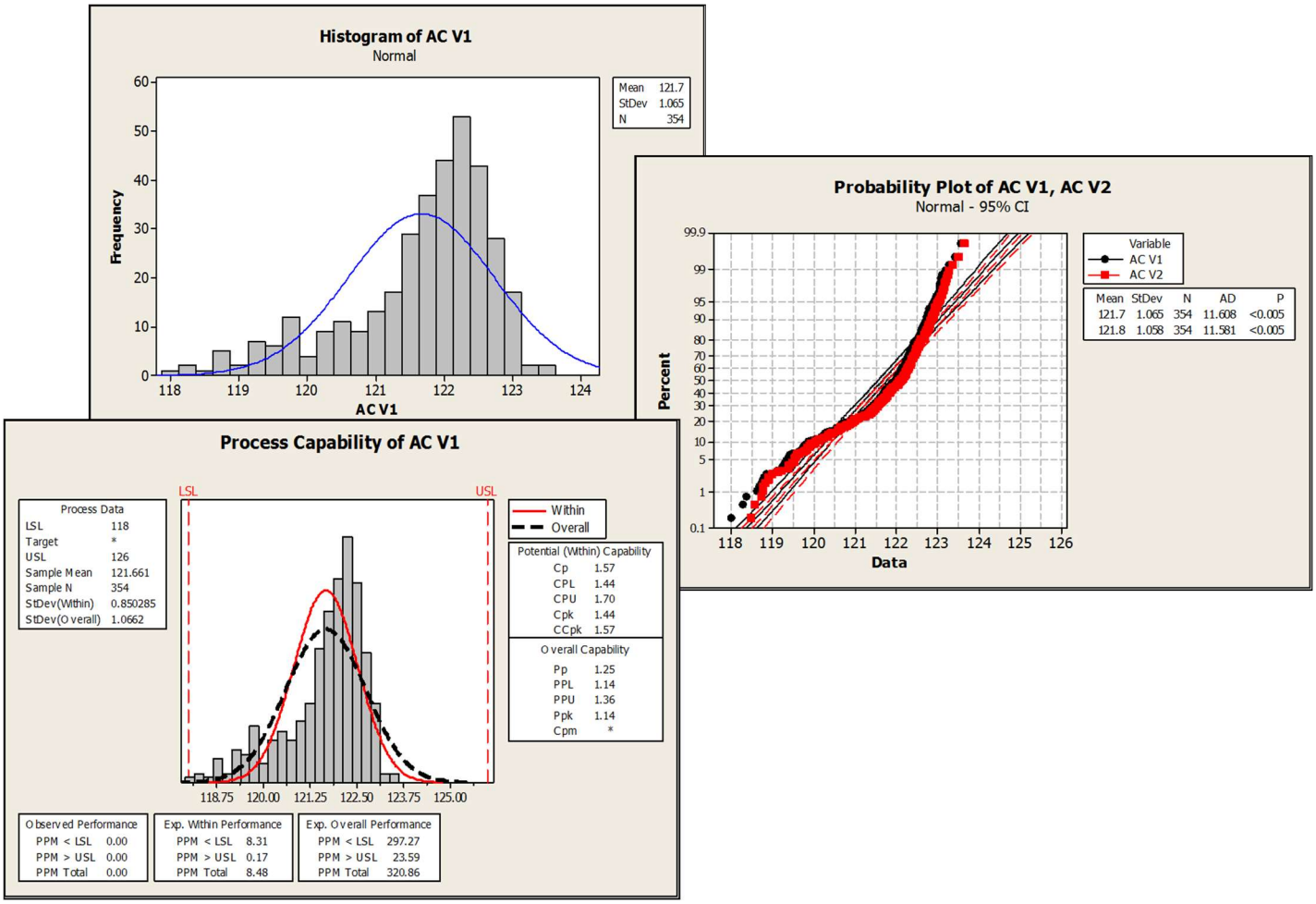
There are 3 main areas when undertaking a Measurement Systems Analysis. They are 1) Precision, 2) Repeatability and 3) Reproducibility.

- Precision – relies on the published accuracy specifications stated by the manufacturer. They may also include guidelines for calibration intervals or expected drift over time. The accuracy tolerances should account for most of the areas listed in No. 4 above. Precision is not a single number, it is the "closeness" of repeated readings to each other.
- Repeatability - by comparing data from testing a single unit, multiple times, at least 10.
- Reproducibility - by comparing data from testing many units several times; at least 10 units, 3 times. This might also be done between systems if there is more than one.

When a new product is being developed, this is another source of variation to consider along with the testing system. To start 30 or more units should be tested and the data between them is reviewed using descriptive statistics such as: Minimum, Maximum, Standard Deviation, Average Mean, Cp, Cpk. In this regard we use Cp and Cpk to understand the *spread and centering* of the data set, not as a process capability index.

Test Engineering Principles

Data should be plotted with histograms to see how the data is dispersed and grouped. A Probability Plot helps to see how well the data correlates to a normal, bell curve distribution. For example, the data below was taken from the output voltage of a DC-AC converter and shows how the data is skewed towards the lower specification limit.



Outliers in the data should be brought to the attention of Engineering. They may wish to perform a closer analysis of why those units are under or over performing.

During each of these stages of development there is an opportunity to improve the test and check that all features of the product are being tested as required to determine fit-for-use against established criteria. This is also the time to identify and resolve any special causes of variability in the testing system or in the product and to review testing limits.

Here are some of the things to be looking at:

a. Do the measurements agree with each other within an acceptable variance?
b. If not, what explains the difference? Is it in the test program, such as timing or delays? Is it in the measurement system, such as the meter ranging? Are all the commands being sent to the instruments correctly? Is the variation coming from within the product itself, such as noise?
c. It's not always clear to where variation is coming from and can require further investigation and consultation with engineering to review product/circuit design, procedures and methodology.

I've found that many of the examples provided in SPC and Quality training seem to be aimed at operators taking *dimensional* measurements with *gages*. These examples often study part-to-part variation and do not consider the special nature of electronic components and electrical design which should be studied part-to-tolerance. The application of these examples toward electronics and electronic components can be a cause of some confusion and can lead to wrong conclusions.

It should be noted that "special causes" are very much an accepted part of electronics by design. The effects of filters, attenuation, noise, part tolerances, binning, vendors and date codes are all testament to this. It is not always possible to see a normal standard distribution in electronic measurements due to these factors. This doesn't mean that variation does not exist, just that electronic or electrical test data can be very poor candidates for process capability indices and GRR studies. Unless you are careful about what the data is you may be studying the output of another supplier's processes instead.

The issue of GR&R studies comes up from time to time. This stands for Gauge Repeatability and Reproducibility and is typically done to study variation between appraisers. See the section on GRR below.

Here is an interesting table from a Quality Spreadsheet I found. You can see all the different ways that a quality metric can be calculated.

	# Opp	Calculated Values										1.5 shift
	1000	dpo	dpu	p	FTY	dpmo	dpm	ppm	Z	Zσ*	Ppk	Cpk
G	dpo	**0.0000034**	0.0034	0.0034	99.66%	3.4	3,400	3,394	2.707	4.207	0.902	1.402
I	dpu	0.000045	**0.045**	0.044002518	95.60%	45.000	45,000	44,003	1.706	3.206	0.569	1.069
V	p	2.53178E-05	0.025317808	**0.025**	97.50%	25.32	25,318	25,000	1.960	3.460	0.653	1.153
E	FTY	5.12933E-05	0.051293294	0.05	**95.00%**	51.29	51,293	50,000	1.645	3.145	0.548	1.048
N	dpmo	0.00012	0.12	0.113079563	88.69%	**120**	120,000	113,080	1.210	2.710	0.403	0.903
	dpm	0.0000012	0.0012	0.00119928	99.88%	1.20	**1,200**	1,199	3.036	4.536	1.012	1.512
	ppm	4.00001E-09	4.00001E-06	0.000004	100.00%	0.00	4	**4**	4.465	5.965	1.488	1.988
	Z	6.91435E-05	0.069143456	0.066807201	93.32%	69.14	69,143	66,807	**1.500**	3.000	0.500	1.000
	Zσ	0.002705944	2.705944401	0.933192799	6.68%	2,705.94	2,705,944	933,193	-1.500		-0.500	0.000
	Ppk	0.000693147	0.693147181	0.5	50.00%	693.15	693,147	500,000	0.000	1.500		0.500

Hypothesis Testing

The formal declaration for designing a test case might be described as "a test of significance as a Bernoulli trial where we perform an experiment designed to provide enough evidence that a unit *has a defect*", since we cannot prove a unit *has no defect*. Tests are based on 4 Bayesian outcomes in a Bernoulli trial which has either a success or failure: *an alternative hypothesis with 4 possible outcomes: True Pass, False Pass, True Fail, False Fail*.

$P_{(A\|B)} = [P_{(B\|A)} \times P_{(A)}] / P_{(B)}$	True Fail
$P_{(\sim A\|B)} = [P_{(B\|\sim A)} \times P_{(\sim A)}] / P_{(B)}$	Producer Risk / False Fail
$P_{(\sim A\|\sim B)} = [P_{(\sim B\|\sim A)} \times P_{(\sim A)}] / P_{(\sim B)}$	True Pass
$P_{(A\|\sim B)} = [P_{(\sim B\|A)} \times P_{(A)}] / P_{(\sim B)}$	Customer Risk / False Pass

When designing a test, we first form a hypothesis, a supposition about what it is we wish to know. We start with the Null Hypothesis H_0 which is what is normal and accepted and most likely to occur (what is a Pass). The Null Hypothesis is assumed to be true unless there is strong enough evidence to the contrary. As for manufacturing, we are normally trying to build a defect-free product.

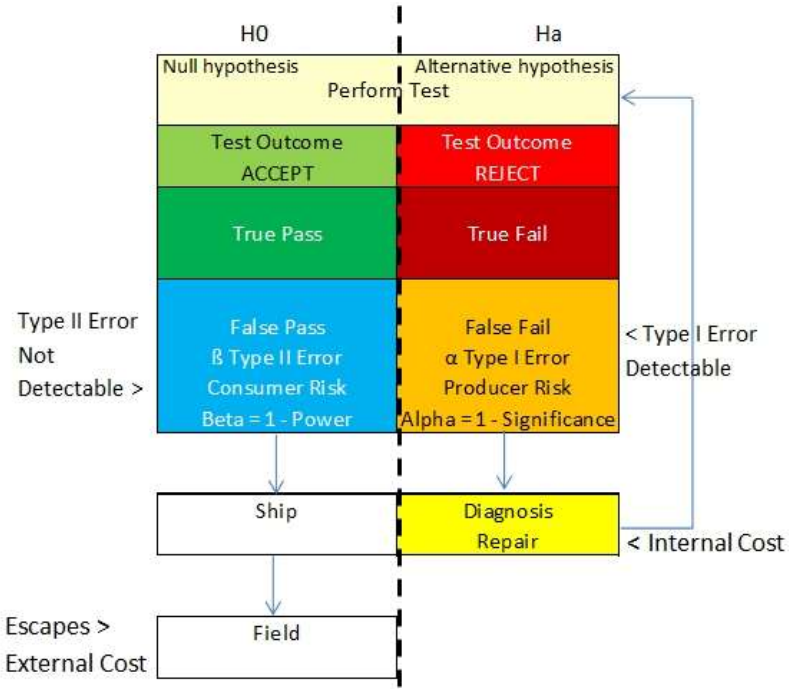
From this we form a basis for an Alternative Hypothesis Ha of what is not normal (what is a Fail) to provide evidence for or against the Null Hypothesis. Just because a unit *passes* a test is not *conclusive* evidence that there was in fact no defect. Evidence comes from the Alternative Hypothesis for which we have devised a series of tests with significance. Some examples can be found in the table below.

Hypothesis	Type	H_o (Null hypothesis) Accept, Pass, Normal	H_a (Alternative hypothesis) Reject, Fail, Abnormal
1	Fault	X <= .03	X > .03
2	Fault	X within a specification	X outside a specification
3	Fault	Efficiency > 90%	Efficiency < 90%
5	Defect	No Missing Parts	Missing Part
6	Defect	Proper Alignment of Parts	Improper Alignment
7	Fault	Functional Parts	Non-functional Part
8	Defect	Acceptable Solder	Too Much Solder
9	Defect	Acceptable Solder	Too Little Solder

In statistical terms, a True Failure is said to be determined by "Rejecting the null hypothesis when the alternative is true". The test engineering interpretation is to "Assign an outcome that rejects the result when a defect exists". The 4 possible outcomes we saw earlier can be explained in the table below for both interpretations.

Test Outcome	Statistical Interpretation[3]	Test Engineering Interpretation
True Pass	Fail to reject H_0 when the null is true	ACCEPT the test result when no defect exists
True Fail	Reject H_0 when the alternative is true	REJECT the test result when a defect exists
False Pass	Reject H_a when the alternative is true	ACCEPT the test result when a defect exists
False Fail	Fail to Reject H_a when the null is true	REJECT the test result when no defect exists

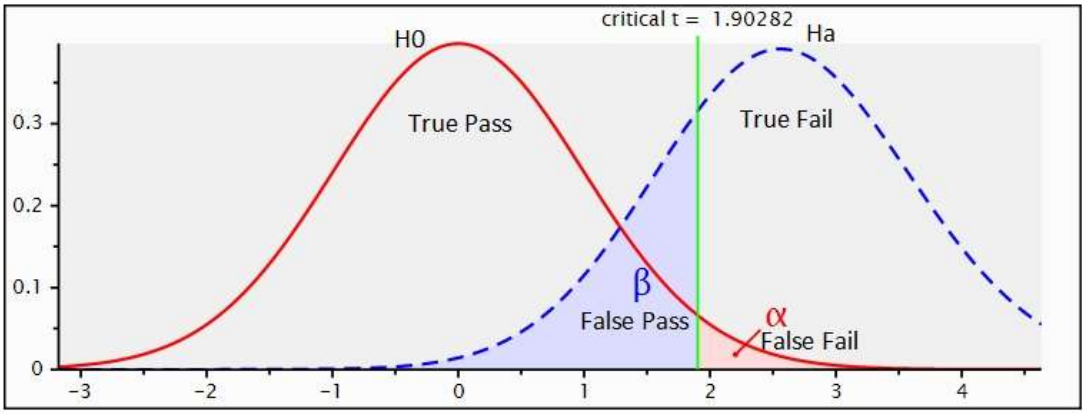
Below is a depiction of what happens during a test process for each of the 4 outcomes above.



False Failures (Alpha Type 1 Errors) are immediately detectable (known as Producer Risk). *False Passes (Beta Type II Errors)* are not detected (known as Consumer Risk) which are called Escapes or Skips.

What is not well understood is that both sides of the "true pass" and "true fail" outcomes have their own distribution curves (see image below).

The shaded areas between the 2 distribution curves below are where we find the sampling errors Alpha False Fail and Beta False Pass. The Effect Size is the distance between the 2 distributions measured in standard deviations. The further apart the 2 distributions are the greater the evidence that a FAILED outcome is true.



Please refer to this article I wrote: https://testview.wordpress.com/2015/07/10/the-effect-of-power-and-confidence-in-electronics-testing/

A fully developed testing program then consists of many of these singular measurement outcomes that result in a success or failure. Each of these results has its own correlation to the Power of that test case to determine if the result is true, within a level of confidence. So then, Power is the statistical term we use for Fault Coverage, Defect Coverage or Test Coverage nomenclature.

In electronics testing, the Confidence Level can be determined by the rate of rejection from the False Fail statistic. This means that if a test is seeing a false fail rate of 5% then the Confidence level would be 95%. As can be imagined, if a test was seeing 50% failures diagnosed as No Trouble Found this would result in low confidence that the test could correctly "Reject the test result when a defect exists".

The confidence level is the inverse of Alpha false failures. Power is the inverse of Beta false passes.

Below is a list of some of the causes of Type I and Type II errors in electronics testing.

False Pass (Beta Type II Error – Consumer Risk)

1. Low fault coverage
2. **Measurement uncertainty (tolerance error outside of specification)**
3. **Incorrect specification limits (too loose or too wide)**
4. Unverified test methods

False Failures (Alpha Type I Error – Producer Risk)

False Failures can also include what is known as the "No Trouble Found" or "No Problem Found" phenomena. Some causes are:

1. Poor preventive maintenance (e.g., intermittent connections and switches, broken wires, wear-out, etc.)
2. **Measurement uncertainty (tolerance error inside of specification)**
3. **Incorrect specification limits (too tight or too narrow)**
4. Poor training of operators and technicians involved in the testing.
5. Poor training of technicians and engineers involved with development of the test.
6. Unverified test methods
7. Undefined test strategy
8. Using the wrong testing method
9. Unclear or confusing test plans and procedures
10. Miss-application of product or feature in its intended function
11. Out of calibration

Yield

We often see Yield reported as (# of units tested - # of defective units / # of units tested). I'll call this Standard Yield. However, Test Engineering and Quality prefers to use Defects per Unit (DPU) rather than the Standard Yield when compiling statistics related to Bayesian outcomes for pass/fail tests that represent a discrete probability distribution. DPU reporting also supports how tests of significance are developed as the alternative in a hypothesis test.

DPU is defined as:

number of defects / number of trials or units tested

DPU is also called the p-value in statistics. Another similar metric in common use is DPMO or Defects per Million opportunities.

Many studies done over the years in semiconductor and electronics manufacturing have found that defects tend to model the Poisson distribution. There are some exceptions but generally this model has been taken as a standard in manufactured electronics for how defects tend to be distributed.

Poisson is a discrete probability distribution of rare events in a large population (>=30)* with each occurrence being independent of other occurrences.

FTY (First Time Yield) can be calculated to model the Poisson as:

$FTY = \text{Log } e^{-dpu}$ Where Log e is the natural log with base e = 2.71828.

As an example, take 1000 trials where 50 defective units were detected. In this case 5% are defective resulting in the 95% Yield commonly reported.

However, when using DPU and the natural log formula we get 95.12% Yield which more closely models the Poisson distribution. We also know that it is possible to have more than 1 defect at a time on the same unit or on the same board.

We can use Binomial Distribution to get the distribution of the number of 0, 1, 2 or more defects per unit. Here is an example for 1000 units at .05 Defects Per Unit:

X	Probability	Number of units on Average
Units with 0 defects	0.9512	951.2 per 1000
Units with 1 defect	0.0476	47.6 per 1000
Units with 2 defects	0.0012	1.2 per 1000

*Central Limit Theroum

The number 30 is defined by the Central Limit Theorem, which is the minimum number needed to establish a normal distribution.

From Wikipedia - It contains a partial solution to a general problem for what is the limiting behavior of the sum of independent randomly distributed variables.

Apparent Yield

We should be aware that the Standard Yield model has an assumption built into it. It assumes 100% fault coverage which does not account for any Beta Type II Error. *Standard Yield assumes there are no escapes.*

To *include* fault coverage in the yield equation, multiply the defects per unit with fault coverage as:

$$Y_a = \text{Log } e^{(-dpu * fc)}$$

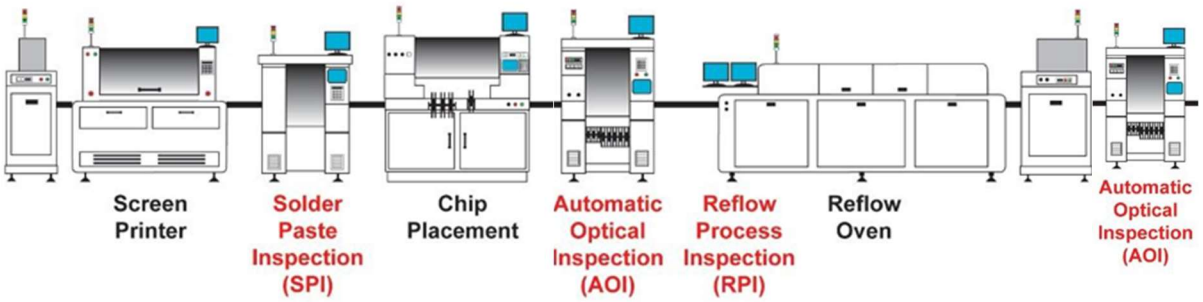
This becomes what is called Apparent Yield or Ya. For example, if we have 200 trials and get a DPU of 0.045 with 90% fault coverage we would get Ya = 95.6%. This is about 1 out of every 22 boards on average with a defect. The Standard Yield would have a result of 96.03%.

The difference between these 2 yields, the one with fault coverage and the one without coverage, is the probability of a Beta Type II Error. In this example the difference is .0043 Escapes Per Unit. If we multiply .0043 x 200 boards, we might see .862 escapes on average, almost 1 in every 200 boards on average.

Now, this is not the most accurate method, which is about 9-10% lower than the Williams & Brown method that will be discussed in the section on Estimating Escapes.

However, the Apparent Yield method may be a bit easier to visualize the importance of controlling the Beta Type II error by doing a Testability and Coverage Analysis.

DPMO – Defect per Million Opportunities



IPC9261 defines how to calculate DPMO or Defects per Million Opportunities in the manufacture of printed circuit assemblies. One of the benefits of using DPMO is that it normalizes defects in products across different boards with varying characteristics and densities so that you can focus on how best to deploy resources for making improvements and in making comparisons.

DPMO is calculated as

(Number of defects / (Number of Boards x Number of Opportunities Per Board)) x 10^6

Looking at yield alone can be misleading as it doesn't consider the level of difficulty in building different types of boards. Let's compare the two methods and see how their priorities might line up.
The table below lists 8 products with varying sample sizes, defect opportunities and yields, in no particular order.

Product	Sample Size	Opportunities	Defects	FTY	DPU
1	90	100	22	78.3%	0.244
2	90	150	3	96.7%	0.033
3	120	1200	25	81.2%	0.208
4	100	1400	5	95.1%	0.050
5	750	100	20	97.4%	0.027
6	1200	120	4	99.7%	0.003
7	1500	1500	22	98.5%	0.015
8	1500	1000	6	99.6%	0.004

FTY would order this priority as 1, 3, 4, 2, 5, 7, 8, 6.
DPMO below would order priority as 1, 5, 2, 3, 4, 6, 7, 8.

Sample Size	Opportunities	Defects	FTY	DPU	DPMO
90	100	22	78.3%	0.244	2444
750	100	20	97.4%	0.027	267
90	150	3	96.7%	0.033	222
120	1200	25	81.2%	0.208	174
100	1400	5	95.1%	0.050	36
1200	120	4	99.7%	0.003	28
1500	1500	22	98.5%	0.015	10
1500	1000	6	99.6%	0.004	4

Since printed circuit assemblies use many hundreds or thousands of parts on each board assembly even a 6-sigma process could see many defects. We can predict the yield from an SMT process that is in statistical control using the following method for a board with 2000 defect opportunities running at 6-sigma (3.4 ppm = 0.9999966):

Ysmt = .9999966 ^2000 = 0.9932 = 99.32%

DPUsmt = Absolute value of the natural Log x 0.9932 = 0.0068

For 100 thousand boards produced we could expect (100,000 x .0068) 680 defects on average.

In the formula above we used Yield to get DPU. In the earlier formula we used DPU to get Yield.

https://testview.wordpress.com/2011/02/28/use-dpmo-to-normalize-defects-across-printed-circuit-board-assemblies/

Rolled Throughput Yield RTY refers to the cumulative effects of the entire process as a performance metric. It multiplies the process yields. DPUs can be summed.

	Y1	Y2	Y3	Y4	RTY
Yield	0.98	0.97	0.99	0.99	0.92
DPU	0.025	0.030	0.015	0.010	0.08

Fault Coverage

Fault coverage is defined as the proportion of detectable defects or faults over the total number of defects or faults. Defects are seen to be physical (such as missing or incorrect parts) and faults are virtual (functional).

To be used as a *defect category* for Pass/Fail tests the defect must be measurable as either a success as or a failure (0 or 1) or on a variable scale between 0 and 1. For example, a category for Missing Parts can be detected (1) or not (0) while solder quality could be measured on a scale between 0 and 1. It should be noted that a category such as Missing Parts is a single-sided distribution while Solder Quality falls on a 2-sided distribution as either too little on one side or too much on the other.

The equation for fault coverage is:

FC = Number of detectable defects / total number of defects

The resulting coverage value is a number between 0 and 1.0 with 1 being 100% coverage.

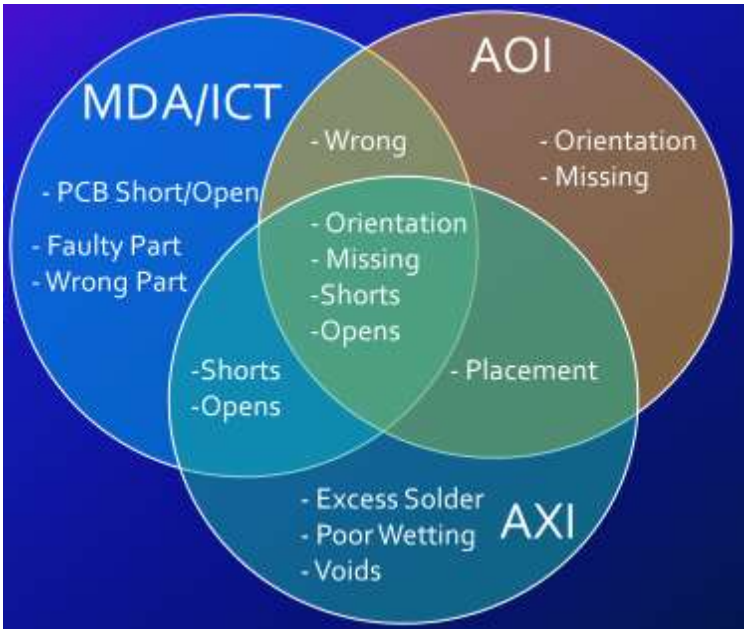
What is not intuitive is that as test engineering works to *improve* fault coverage, well, more units with failures would be detected so the resulting yield might go down, thereby causing some disappointment amongst some Management teams. It is the process that must be improved in order to *raise* yields. The focus on better testing, by itself, does not improve yield.

Since we cannot expect one type of test to have full coverage, the overall test process should be reviewed to identify gaps in coverage and design tests to fill in these gaps where it is most economical so that the total reaches 100% through Testability Analysis.

We know that defects not covered by a test could be in a region that is highly controlled by the manufacturing process or may not be randomly distributed after all. Some defects tend to occur more frequently in some places due to certain design aspects. But these are special causes which should be dealt with using continuous improvement techniques. Fault coverage models can apply weighting averages to reflect some of these situations. However, in general we tend to regard each defect opportunity as independent from the others and with an equal probability of occurring.

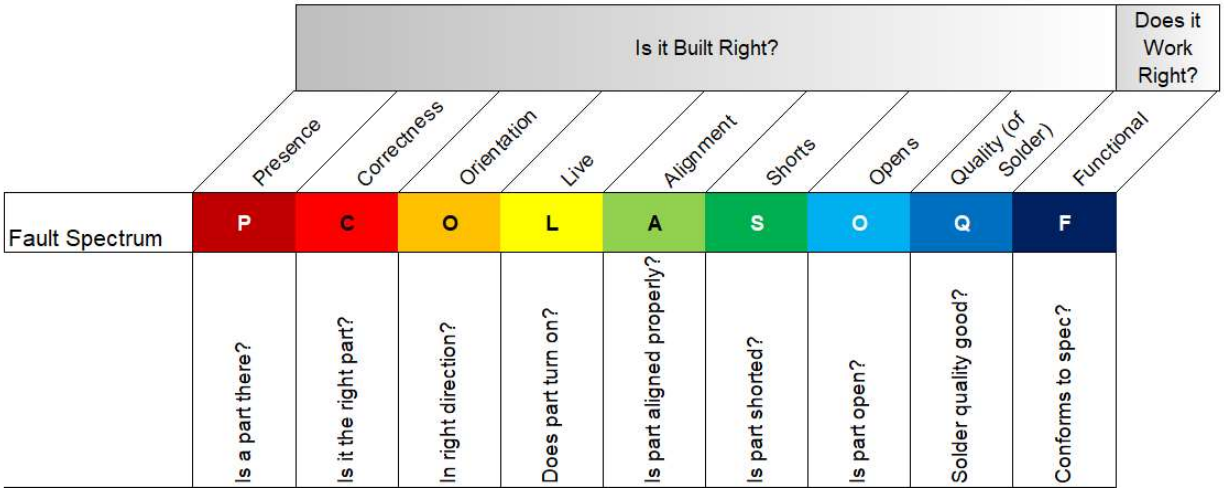
Fault Coverage Analysis

How is Fault Coverage determined? Fault coverage analysis is performed by first defining the defect and fault categories. Below is a chart of the defect universe as a Venn diagram for the appropriate inspection method.



Defects are found to fall within certain categories for printed circuit assemblies:

- Presence – is the part present?
- Correctness - is the part the correct value?
- Orientation – is the part in the correct direction? (For polarized parts)
- Live – does the part turn on?
- Alignment – is the part aligned properly? (Variable scale)
- Shorts – is part shorted?
- Opens – is part open?
- Solder Quality – too much or too little solder (variable scale)
- Functional – is the part functional? (Conforms to specification)



Once the defect or fault category is defined then the inspection or test process is reviewed for detectability in those categories. These fall into discrete or continuous variables. For example, Presence or Correctness for parts on circuit board assemblies may each get a value of 0 or 1 for whether the test can detect that defect or not. This process repeats for every component in the design.

Take an Automated Optical Inspection for example:

- 9 out 10 parts can be detected for Presence = 0.9
- 7 out of 10 parts can be detected for Correctness = 0.7
- 2 out of 2 polarized parts can be detected for Orientation = 1.0
- 10 out of 10 parts can be detected for Alignment = 1.0

The results are then averaged across the total number of parts, so we get (0.9 + 0.7 + 1.0 + 1.0 = 3.6) / 4 categories = 0.90 fault coverage. Once the other 10% of uncovered parts are identified those could be checked by some other means.

Functional tests can be seen as testing a collection of parts outlined by circuit blocks that are designed to perform certain functions. A circuit consists of many parts so a functional failure can have many causes that would need further diagnosis or troubleshooting.

The coverage process can be repeated for faults related to components within the circuitry which the functional test is attempting to measure. Faults in functional tests are defined by limits and tests are meant to determine if a circuit is functioning correctly. For example, is the measured output less than, greater than or within certain limits? Therefore, what is a fault can be redefined by changing limits.

Functional Tests are not necessarily designed to target individual components that are in-circuit. So, this may require simulation software to inject faults into a virtual model of the circuit to get coverage. The faults that are the most detectable tend to be related to the parts that are most critical to the operation of the circuit.

Fault Coverage Analysis as PFMEA

A fault coverage report could be used to feed information into a PFMEA (Process Failure Mode and Effects Analysis). FMEA contains a rating system for considering Severity, Occurrence and Detection and often relies on a cross-functional team effort that can be time-consuming and somewhat subjective. This is just my opinion, however, since I have not seen this used in practice.

Severity

The Severity rating can help to shorten the list of priorities in building the FMEA, to make a big job a little easier to digest. It is true that some components play a more critical role than others, however the goal is defect-free products regardless of the criticality of any specific part. Therefore, fault coverage analysis simply considers all parts as having the same importance regardless of their effect, so Severity need not be considered.

Occurrence

Occurrence can be handled by the assumption that each defect has the same probability of occurring. But if the historical defect rate is known from process data a weighting system could be applied. In this regard reliability data could also be used for how many parts per million or billion could be expected to fail. Historic defect rates together with Reliability data could help produce a more accurate model.

Detection

Then that leaves Detection which is based on how well the test process, system, or equipment can detect a defect if it exists.

BOM and PCBA CAD data can be imported so there is no subjectivity in this regard. It is all based on whether a part can be detected within a defined defect category by the inspection systems to be employed and that all parts have the same importance and have a known probability or at least the same probability of occurring if not known.

Estimating Escapes

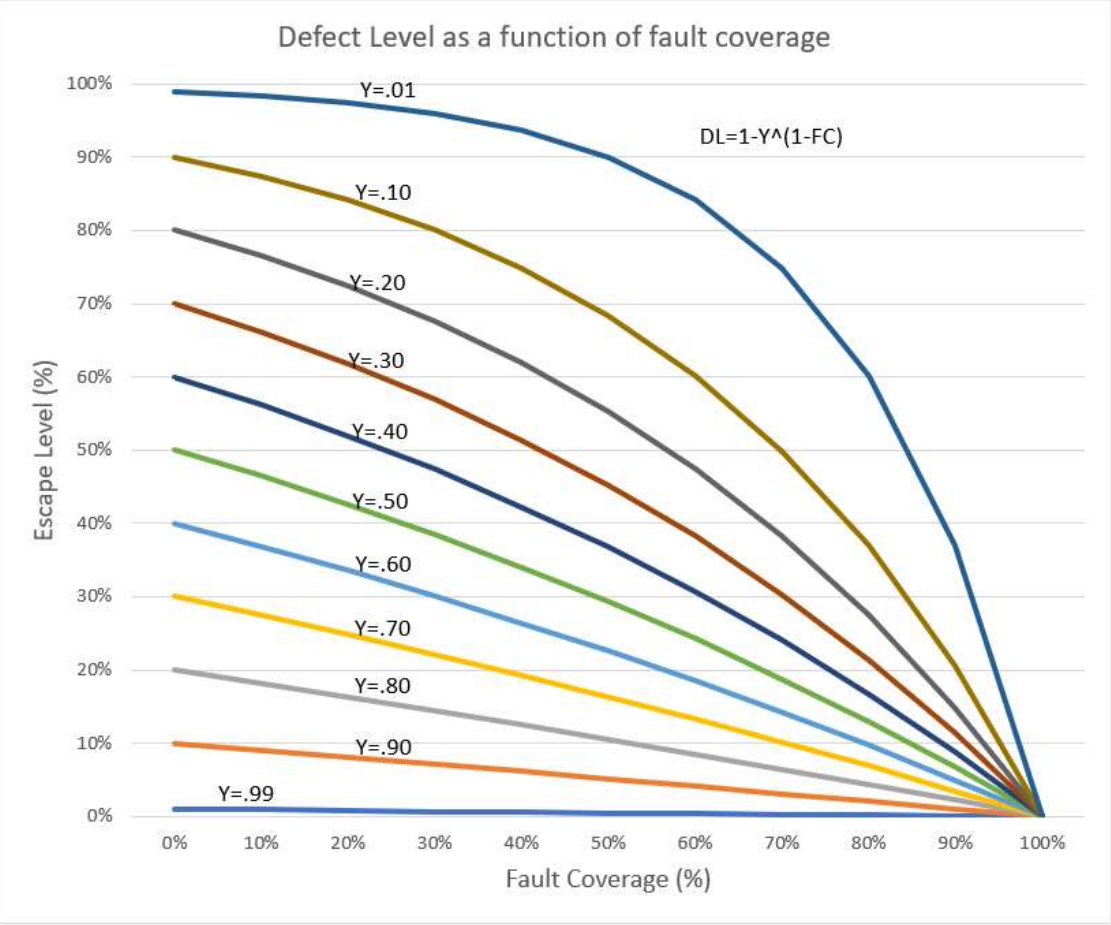
As part of an intensive study into the subject of Escapes I found an IEEE paper written in 1981 by T.W. Williams and N.C. Brown. There are very few papers written on the subject. Their paper describes how the DL (defect level) changes as a function of fault coverage that follows a Poisson distribution.

DL = 1-Y^(1-FC)

where Y is the Yield and FC is Fault Coverage

In this paper, "The defect level (DL) is equal to the probability that a bad chip is accepted, divided by the probability of accepting a bad chip plus the probability of a good chip, which under our assumptions will always be accepted."

I created my own chart below from the Defect Level formula above for a family of yield curves:



There is also a similar equation for the Escape Level that follows a power law function:

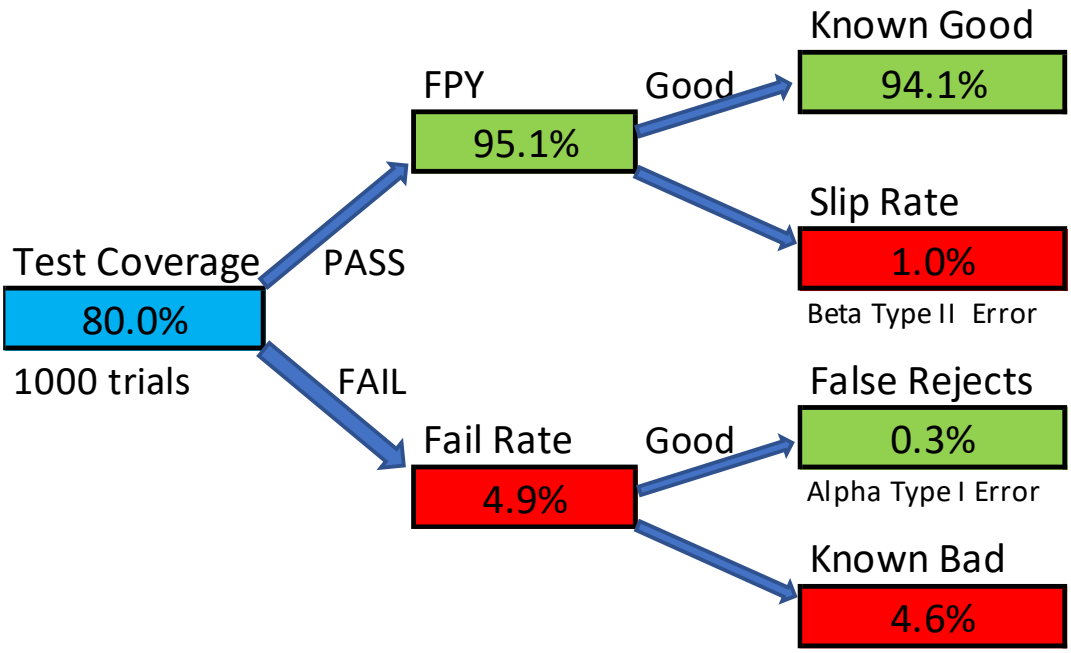
EL = 1-(FC^(1-Y)

For example, In the process below we have 50 failures from 1000 samples at 80% coverage. Using the Defect Level equation, we might get 1 escape on average for every 92 units tested.

Fault Coverage	80.0%
Samples	1000
Fail	50
Defects Per Unit	0.0500

Yield, % 0 defects	0.9512	95.12%	
% with 1 defect	0.0476	4.76%	1 defect for every 21 units on average
% with 2 defects	0.0012	0.12%	2 defects for every 841 units on average
% with 3 defects	0.0000	0.00%	3 defects for every 50461 units on average
DL 1-Y^(1-FC))	0.0100	1.00%	1 escape for every 92.4 units on average

Williams and Brown 1981



Escape Study and Prediction Model

I developed an experimental study of my own where I simulated a process that created random defects in order to see how many Escapes there might be at varying levels of coverage. During this process I discovered another equation on my own for estimating escapes that closely followed the Williams & Brown equation, of which I was unaware at the time.

I wrote a paper on this and attempted to submit it to the IEEE. They did not publish my paper but instead pointed me to some prior papers on the subject including the Williams and Brown paper from 1981.

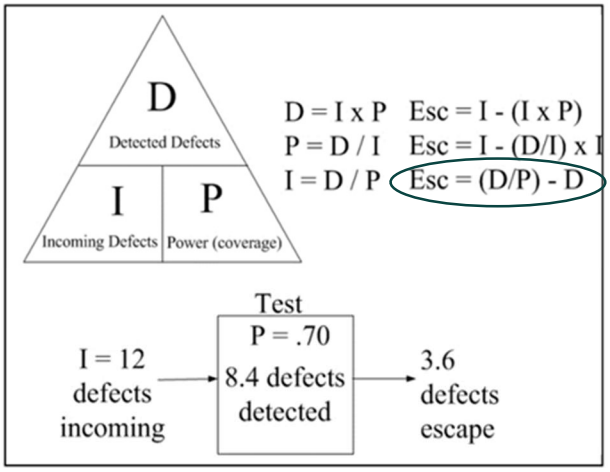


Fig. 3. Incoming, detected defects and escapes seen from test process.

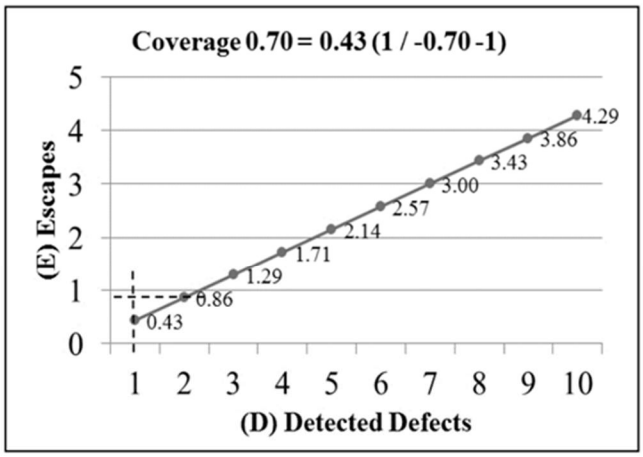


Fig. 4. Linear regression line for detected defects for coverage 0.70 (for any sample size with mean = 0 and standard deviation = 1).

In the paper I wrote I started by looking at the relationship between Incoming Defects, Detected Defects and Power or Fault Coverage. Figure 1 above shows this relationship. Many relationships like this can be found, such as with Ohms Law for example. I then added the various equations for Defects, Power and Incoming defects and how one might derive Escapes from what values were known.

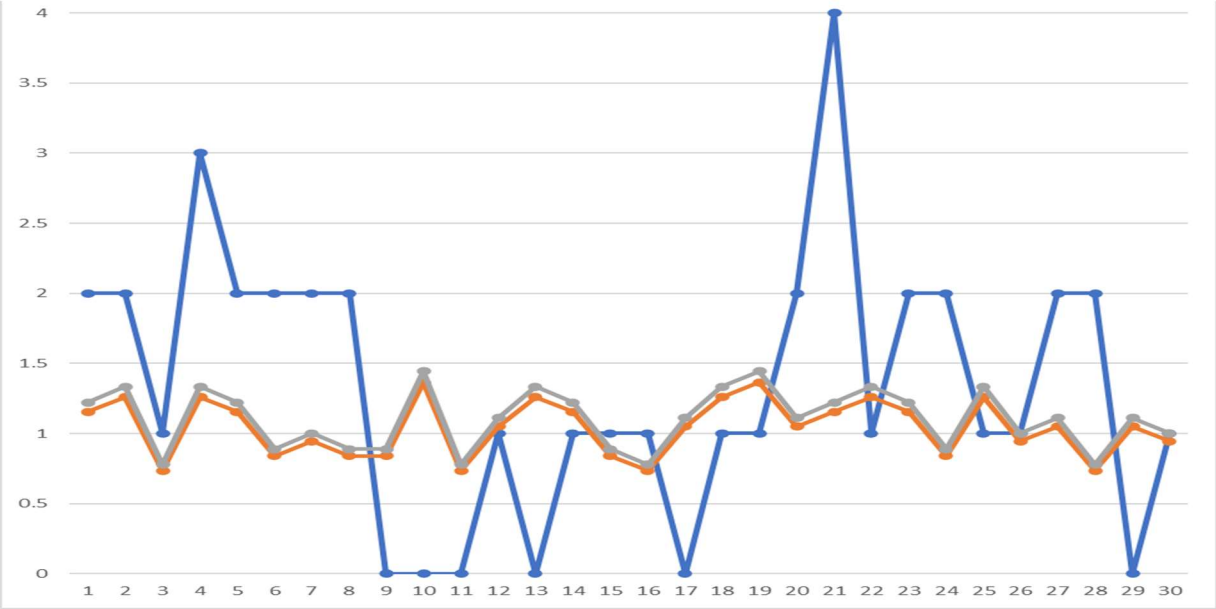
When we run a real test, we can know the defect rate and we might be able to know the fault coverage. What we don't know is how many defects are actually there until they've all been tested (with 100% coverage). From this, I noticed the formula for Escapes (circled above) which is (Number of Detections/Coverage) – Number of Detections. So I started running the simulation comparing this formula to actual escapes, which is not something we would normally know.

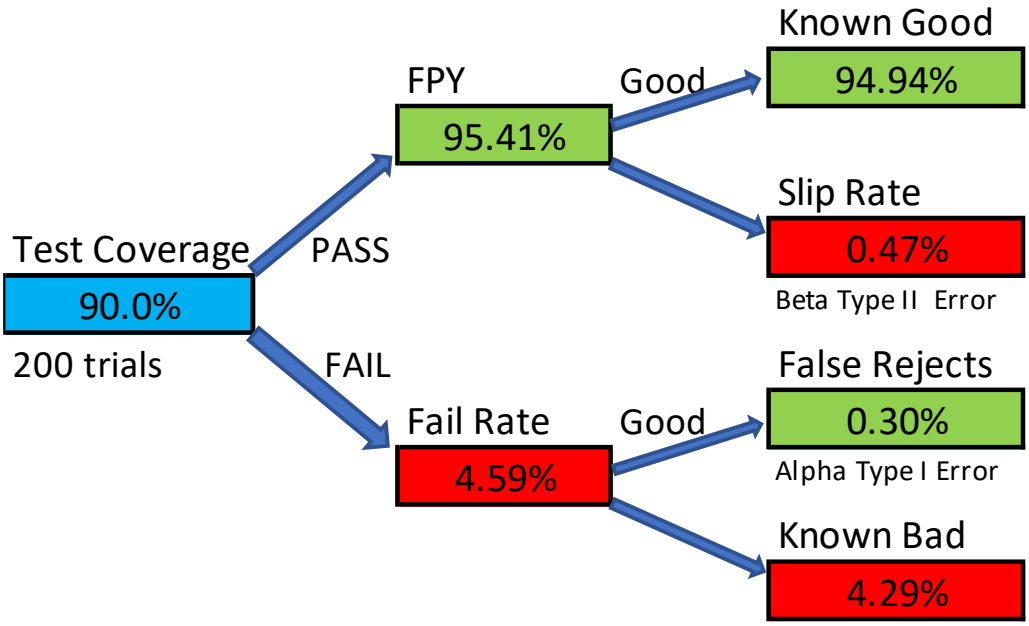
The result of this study was that I found what I call a k covariate which can be used as a kind of "Z-Score" for Escapes which describes the relationship of this value to the mean of a group of defect values as shown in Figure 4 at any level of coverage.

I then later found that by multiplying this *k* covariate with DPU you get a number very close to the result of the Williams & Brown equation. Without going into too much detail on how the simulation was carried out I then attempted to see how the 2 equations tracked with known escapes in the simulation model.

From the chart below you can see how well the 2 models track each other where the blue line is Actual Escapes. This dataset was run using 200 trials 30 times with DPU 0.0473 and 90% coverage.

You can see there can be wide difference from one set of trials to the next which could be anywhere from 0 to 3. What was found was that the average escapes across all 30 trials in this dataset was is 1.33. The average for the *predictions* on this dataset is 1.04 for W&B and 1.10 for Schoen. So, the predictions are quite close to *actual* esapes which is 1 Escape per trial on average across the board.





While no model found yet can be completely accurate when predicting escapes *from one trial to another*, the value comes from the knowledge that you can expect to get 1 escape on average for a process with 90% coverage at a defect rate of 0.0473 DPU.

Cpk - Process Capability Index

As a capability index Cp represents the **spread of the data** (related to precision) and is defined as:

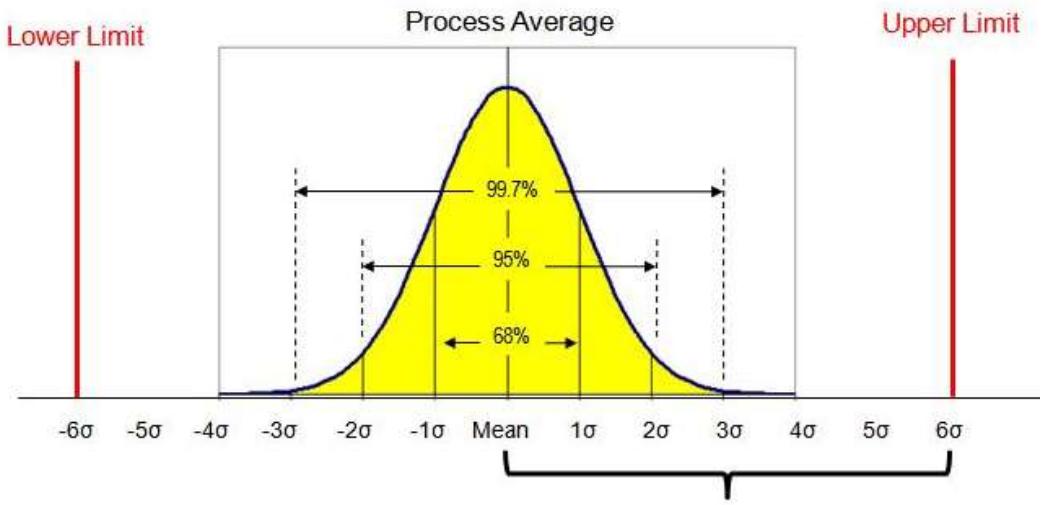
Cp = Tolerance Range as USL - LSL / 6σ (6 standard deviations)

Cpk is a short-term index (Ppk is the longer term) and represents how **well-centered the data is** (related to accuracy), which takes the minimum or the worst case of either:

(Average mean – LSL) / 3σ or (USL- Average Mean) / 3σ (3 standard deviations)

The use of Cpk as a measure of Process Capability has some stipulations and rules regarding its use. All the index does is compare a process that is *normally distributed and in statistical control* to a specification that you or your customer have decided on. If a process is not in control, then Cpk becomes meaningless. It is the control charts that tell us when a process is no longer in control and that corrective action is needed. What the index *can* do is tell you the *probability* of the number of defects that *might* be produced by *a process that is in statistical control.*

https://www.leansixsigmadefinition.com/glossary/six-sigma/



The data for Cpk is usually taken from Control Charts that plot either a measurement that is 1) a variable attribute of a process that is measurable such as the dimension of a part or 2) an average of a sub-group of parts such as the defect rate of a sample population. It is the second one that electronic test data is most often concerned with.

Control charts are used to compare one set of *short-term data* (that may might have special cause) to *longer-term historical data* that represents a process that was in statistical control at the time with common causes. I've seen teams use process limits instead of specification limits for Cpk. Then what happens is that the team may begin to apply countermeasures that end up shifting the process limits to make it appear that the process is back in control.

It is up to the Quality and Engineering Teams to decide on what are the most important indicators for Special Characteristics that are best served by tracking Cp and Cpk. There is a cost to the effort and time

taken to collect, analyze and report process capabilities. Not all measurements need to be tracked or even make sense to be used for Cpk.

PPAP Appendix F.7 and the AIAG define Special Product Characteristic as a product characteristic for which "*reasonably anticipated variation could significantly affect a product's safety or compliance with government standards or regulations or is likely to affect customer satisfaction with the product*".

Though Cpk can be derived mathematically as a function of Yield it is less accurate than control charting and an out-of-control condition cannot be determined from Cpk.

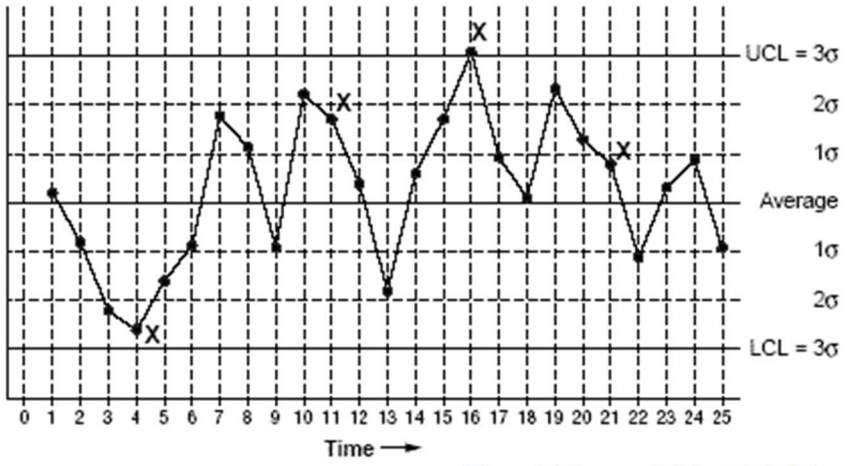
To calculate Cpk using Yield we take what is called the Z-score of an inverse normal distribution + 1.5 (sigma shift for process variation) divided by 3 as the number of standard deviations.

One thing to remember is that when Cpk = 1.0 this means the specification tolerance = the process spread, and the process mean coincides with the tolerance mean.

Special Cause Criteria

Identifying special cause is important in working towards making process improvements. So, what determines Special Cause? If we are looking at data as it is plotted on a c-chart, some criteria are listed below:

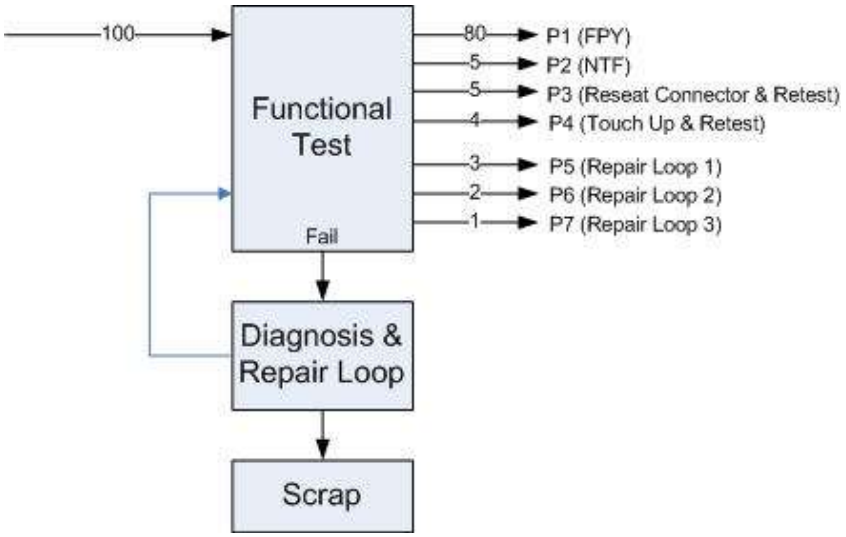
	Summary of Typical Special Cause Criteria for Control Charts
1	1 point more than 3 standard deviation from centerline (describes an outlier)
2	7 points in a row on the same side
3	6 points is a row all increasing or decreasing
4	14 points in a row, alternating up and down
5	2 out of 3 points > 2 standard deviations from centerline on the same side
6	4 out of 5 points > 1 standard deviations from centerline on the same side
7	15 points in a row within 1 standard deviation of centerline (either side)
8	8 points in a row > 1 standard deviation from centerline (either side)



On the Floor - Test Cost Analysis

As each unit in a population is tested there are different ways that they may be processed. We can illustrate this with an example shown in the diagram below.

For example, out of 100 units 80 passed on the first attempt, 5 boards were retested and passed immediately, 5 boards passed after reseating a connector, 4 boards had some solder that was touched up before they passed, 3 failures had 1 repair before passing, 2 failures had 2 repairs before passing, and 1 unit had 3 repairs before passing. Thus, closing the loop on all 100 units. No units were scrapped.



- P1 – the total number of units that passed on the first try.
- P2 – boards that failed the first pass and were then retested and passed immediately without being touched. This may be able to uncover timing or other problems related to the test method – Type 1 errors related to False Fail.
- P3 – are units that failed but passed after reseating the interface connector or reloading the board in the fixture. This can help to reveal connector/terminal pin wear-out or other fixture related maintenance problems. These are also Type 1 errors.
- P4 – boards with solder bridges or some other minor defect. How you decide to track these defects is up to you. I placed the possibility of this happening in this position as some departments may try to correct these faults at Test. Others may send those boards to a Rework location regardless as not all test operators are trained to do this kind of work.
- P5 – are units that were then passed after being through the first diagnosis/repair loop.
- P6 and P7 – are units that passed after 2 or 3 trips through the diagnosis/repair loop. The earlier diagnosis may have been incorrect. However, some units may in fact experience more than 1 or even 2 faults per board.

So how do we analyze this data? What is the cost of poor quality? First, we tally up the units and how many times these units were tested:

Test Process Analysis

Results	True Pass	True Fail	False Pass	False Fail	Number times through Test	Total Times Tested
P1 (FPY)	80				1	80
P2 (Retest)	5				2	10
P3 (Reconnect/Retest)	5				3	15
P4 (TouchUp/Retest)	4				4	16
P4 (RP1)	3	3			4	24
P5 (RP2)	2	1			5	15
P6 (RP3)	1	0			6	6
SCRAP [1]		0				0
False Pass			0			
Total	**100**	**8**	**0**	**10**		**166**

Next, we assign a cost structure. What is the cost per test based on the labor rate and cycle time + handling time, the cost to repair, etc.

Test Process Cost Setup Data

Enter information into these cells:

Number of units in lot	100	
Cost of WIP	$ 50.00	Cost invested in product at this level of the process
Labor Rate	$ 30.00	Labor rate that does not include overhead costs
Load/Unload Time (Seconds)	10	
Cycle Time (Seconds)	30	Average time it takes to test 1 good unit including load/unload times
Avg Visual Inspect Time (Seconds)	30	Average time it takes to inspect a board for bridging, etc at test

Cost of Load/Unload UUT from fixture	$ 0.08
Cost of Test	$ 0.25
Total Test Cost	$ 0.33
Average Repair Cost	$ 7.50
Average Visual Inspection Cost	$ 0.25
Average Cost to Repair an Escape	$ 100.00

Then we take the standard cost per test and compare that to the total cost to test the 100 units that includes retesting, repair, etc. Here we find the total cost was $100.33 or $67.00. 3 times over the standard reported cost.

Test Process Costs & Summary Report

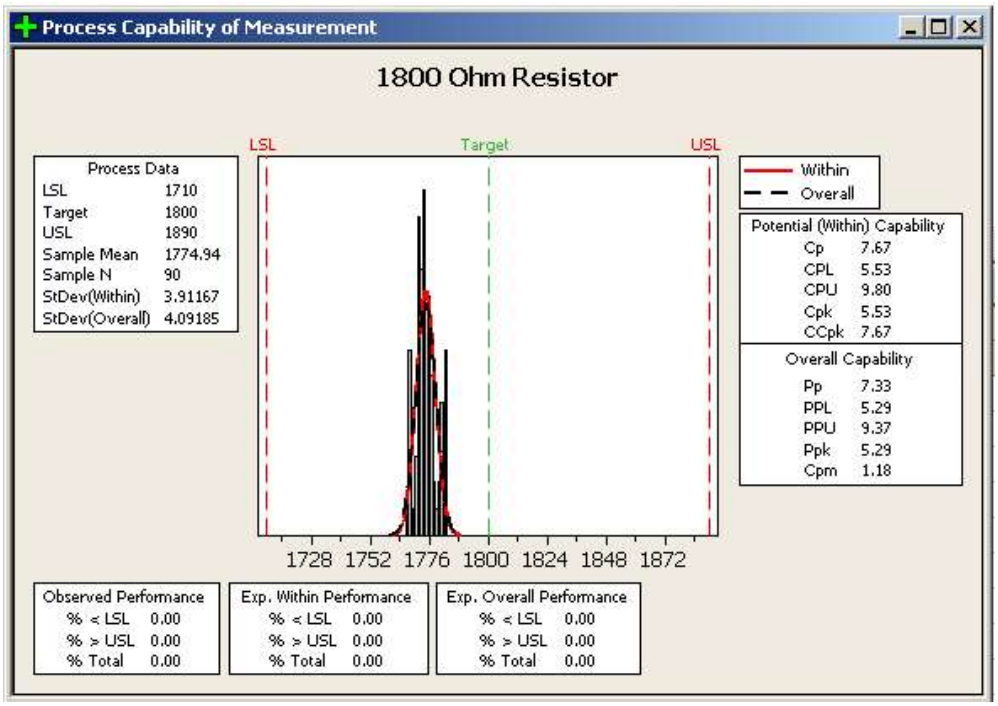
Results	Cost	Description
Standard Cost	$ 33.33	Standard test cost reported to Finance for the quantity tested
Total Test & Repair Cost	$ 100.33	Actual total cost to test and repair all units in lot OR the total cost to detect 8 failures
Type 1 Error Cost	$ 8.33	This is the cost of false failures due to Producer error
Test Cost	$ 55.33	Total test cost
Repair Cost	$ 45.00	Repair cost
SCRAP COST	$ -	Scrap cost
FALSE PASS COST	$ -	Cost to repair escapes
Cost of Poor Quality	$ 67.00	Additional cost over standard cost
Overhead Cost Factor	3.0	Number of times actual cost over standard cost
Capacity Reduction	39.8%	Percent loss in productivity of tester
Standard time for lot	0.833	Standard time it should take in hours to test lot
Actual test time for lot	1.38	Actual time it took to test all units in lot to final resolution

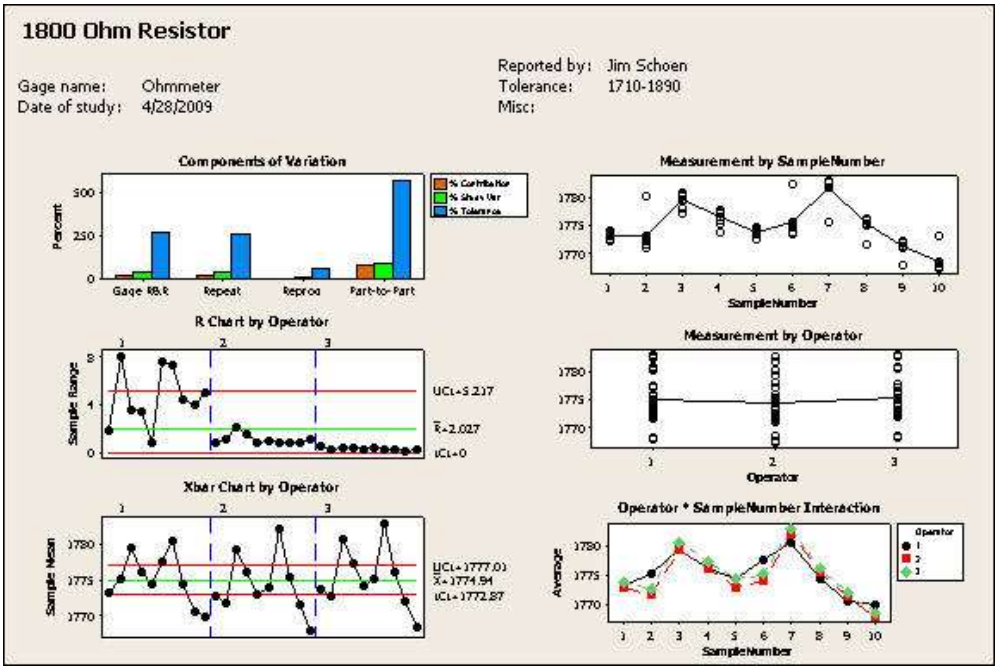
GR&R Studies

GR&R (Gage Repeatability and Reproducibility) is the variance equal to the sums of within-system and between-system variances. There is a difference when studies are done for Product Control (as with electronics testing) or Process Control (as in the dimensional measurements taken after a cutting machine). GRR applied to Electronics testing should use the Part-to-Tolerance method and dimensional measurements the Part-to-Part method. I have seen some confusion on this aspect where the wrong method has been applied.

For example, I took a sample population from a reel of 1800 Ohm 5% axial resistors. I had them measured by 3 different operators with a DMM. A GR&R study using ANOVA was run with Minitab based on part-to-part tolerance. It was found to be 42.51%! A failure is anything over 30%. The same data when ran as part-to-tolerance was 13.5%.

This group of parts represents how electronic components are purchased and used. What we see here is only a sub-set of a much larger population that is used at any one time. The measurements taken from any population sample might shift above or below the center target from those taken on another reel, yet these would all be perfectly acceptable components and meet the specification tolerance. The same thing can apply to integrated circuits and functional tests.





Also, in situations where automated test systems are used there are no human appraisers; computers are doing the work. If all the parts are handled, fixtured and measured by the same equipment then reproducibility is zero, only repeatability needs to be studied. However, if multiple fixtures are used, then reproducibility is the between-fixture variation. [3]

So much of what is considered a population sample in statistics regards taking a random sample of a population that already exists. And yet in electronics we are building a larger population out that is released to the field in smaller batches over time using parts that are highly selective.

Numerous parts used in one build can be shifted compared to the parts used on the next build. We find that the populations of parts (in electronics) are not truly random, they are comprised from a smaller subset of parts that were also binned, sorted or culled. Only when taken overall after many builds you might then start to see a distribution that looks more normalized and even that should be within the expected tolerances of the parts that were used.

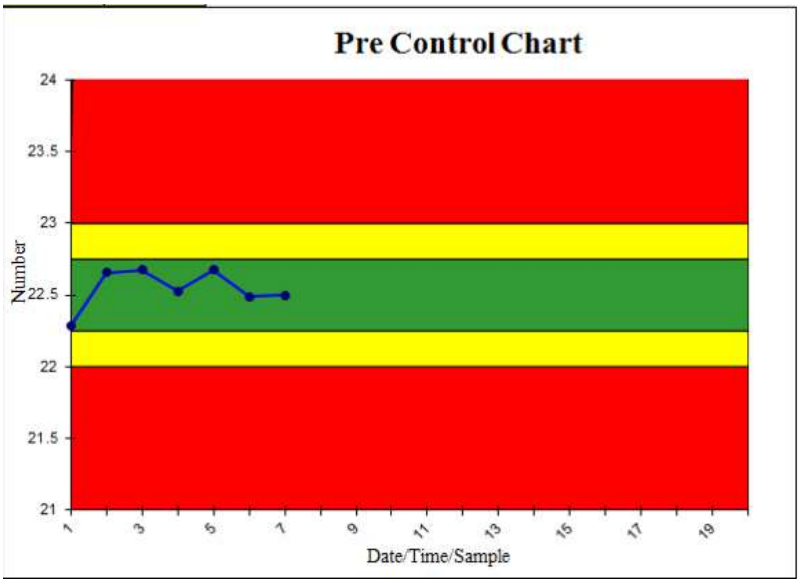
This situation eventually leads to DOE (Design of Experiments, Monte Carlo, etc.) which can be used help to understand the interactive effects of varying tolerances of parts on the output of functional circuits.

Pre-Control

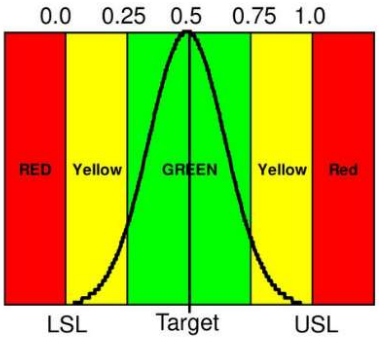
If you can test items as they come off a machine, then Pre-Control may be something to know about. This method was developed by Frank Scatherwaite for the consulting company Rath and Strong in the 1950s.

Pre-Control does not require a normal distribution or any assumption concerning the shape or stability of that distribution. There are no control charts, and you don't even need to know the capability of a process. You can begin production right away using specification limits provided those have been set correctly. You can also verify the control limits from trial production runs.

It is a very simple concept and easy for operators to use. The chart looks something like this:



Pre-Control Charts use limits relative to the specification limits. This is the first and ONLY chart wherein you will see specification limits plotted for Statistical Process Control. This is the most basic type of chart and unsophisticated use of process control.



Red Zones. Zone outside the specification limits. Signals the process is out-of-control and should be stopped

Yellow Zones. Zone between the PC Lines and the specification limits indicating caution and the need to watch the process closely

Green Zone. Zone lies between the PC Lines, signals the process is in control

LSS Green Belt v11.1 MT - Control Phase 43 © Open Source Six Sigma, LLC

Pre-Control can determine if the process is in control within the first 2-5 units. Here are the basic rules:

1. The first 5 consecutive units must all pass. If even 1 unit falls anywhere outside of the green zone, then the process is not in control and production cannot begin. You must stop production and investigate why.
2. If 5 in a row pass, then the next 2 consecutive units must follow the next set of rules:
3. If one unit falls in the green zone and one other in the yellow zone, the process is still in control.
4. If both units fall in the SAME yellow zone, this is an indication of drift. Momentarily stop and adjust.
5. If 2 units fall in OPPOSITE yellow zones, it indicates a major shift. This is not a reject yet, but you must stop production and investigate.
6. If even 1 unit falls in the red zone, this is a clear reject. Stop production and investigate.
7. Whenever production is stopped you must start over again from Step 1.

There are also some methods for setting the sampling frequency. Here is some of the theory behind Pre-Control as spelled out by Bhote and Bhote in *World Class Quality*.

In the worst-case scenario, the process limits = the specification limits (Cpk = 1.0). Assuming a normal distribution the area under the green zone is 86% and the yellow and red zones are 7%. The probability of 2 units landing in the same yellow-red zone is 7% x 7% = 0.49%. The probability of 2 units landing in the same yellow-red zone is roughly 1 in 200.

But there are 4 ways in which 2 units can land in the 2 yellow-red zones. So, the probability is 2% or a 1 in 50 chance of over-correction (falsely passing) which means a 98% chance that this is *not due to chance* and that a correction is needed.

So, for a process with a higher Cpk of 1.33 the probability of 2 units landing outside Pre-control limits by chance shrinks to 0.84%. This is now a 99.16% probability that any unit outside the yellow-red zone is not due to chance. With Cpk's of 1.33 there is virtually no risk of a bad product ever being accepted which is the basis for the 2 consecutive unit rule. With Cpk's of 2.0 hundreds of thousands of products can be produced without defect.

Types of Probability Distributions

A quick description of some of the different types of distributions we might see from data taken from manufacturing processes.

Poisson – the *number of defects* in a given period of time, length, area or volume if they are occurring randomly and are independent. Poisson closely approximates the Binomial if n is large (>30) and p is small (<0.05).

Binomial = what is the distribution of the *number of defects* in n independent Bernoulli trials?

Negative Binomial – the *distribution of trials* needed to get a certain number of defects in repeated independent Bernoulli trials.

Geometric = the *distribution of the number of trials* to get the first defect.

Hypergeometric - the *distribution of the defects* when we are drawing *without replacement* from a source that contains a certain number of defective items and a certain number of non-defective items.

References

[1, 2] Chapter I, Section A, Page 5-6, Measurement Systems Analysis Reference Manual, Third Edition, Second Printing, May 2003 by Daimler Corporation, Ford Motor Company, General Motors Corporation

3 Chapter III, Section B, Page 97, Measurement Systems Analysis Reference Manual, Third Edition, Second Printing, May 2003 by Daimler Corporation, Ford Motor Company, General Motors Corporation

Books

- World Class Quality, Using Design of Experiments to Make It Happen by Keki R. Bhote and Adi K. Bhote
- World Class Reliability, Using Mulitple Environment Overstress Tests to Make It Happen by Keki R. Bhote and Adi K. Bhote
- Accelerated Stress Testing Handbook, A Guide for Achieving Quality Products, Edited by H. Anthony Chan and Paul J. Englert
- Building a Successful Board-Test Stategy by Stephen F. Scheiber
- The Practice of Engineering Management, A New Approach by Patrick D.T. O'Connor
- Six Sigma for Electronics Design and Manufacturing by Sammy G. Shina, Ph.D.
- Statistical Methods for Testing, Development and Manufacturing by Forrest W. Breyfogle III
- Out of the Crisis by W. Edwards Deming

An excellent website for learning statistics can be found at is https://www.jbstatistics.com/