26th annual INCOSE international symposium

Edinburgh, UK
July 18 - 21, 2016

A Practical Guide to Statistical Verification

James Kolozs

syncroness™
complex problems inspired solutions

A Question…

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016

- **Verification** – Why did you test the **number of samples** that you tested?
  A. Because that's how many we have.
  B. We always test 30. It's statistics I think.
  C. 5 is good, right?
  D. I only have one, so I test it a lot.
  E. Never thought about it.

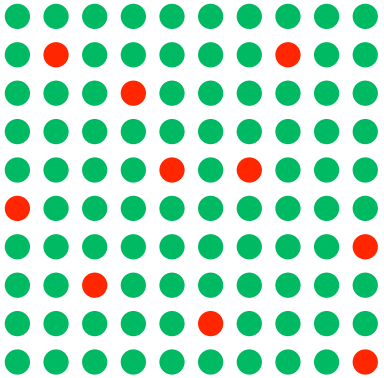
Confidence is Key

- As a stakeholder, which gives you the information to make an informed decision?

"10 samples were tested without failure."

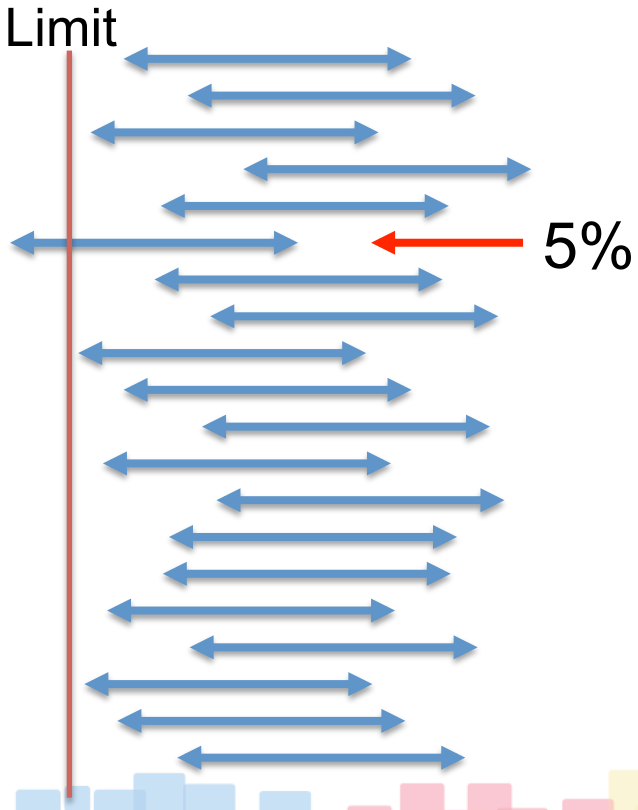
"After testing 10 samples, there is 90% confidence that 90% of the population meets the requirement."

100% Pass, right?



Or a 10% Failure Rate?

What is 95% Statistical Confidence?



- Example:
- We want our data to be above a limit
- Assume we have calculated 95% confidence
- Test many times, and 5% of the time the result will be below the limit

Why You Should Care

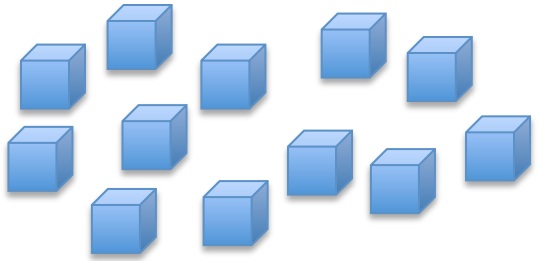
- Do you develop medical devices?
  - FDA and ISO **require** “valid statistical techniques” for:
    - Sampling plans
    - Verifying product characteristics

- In any industry, testing is expensive!
  - Know the **minimum** number of samples required **before** starting
  - **Know** that you are going to **pass**
  - **Account** for **risk**
  - **Predict** your **population**, not just your samples

A Better Way

- Statistical verification enables you to:

Know how many samples to test



Have statistical confidence in your results

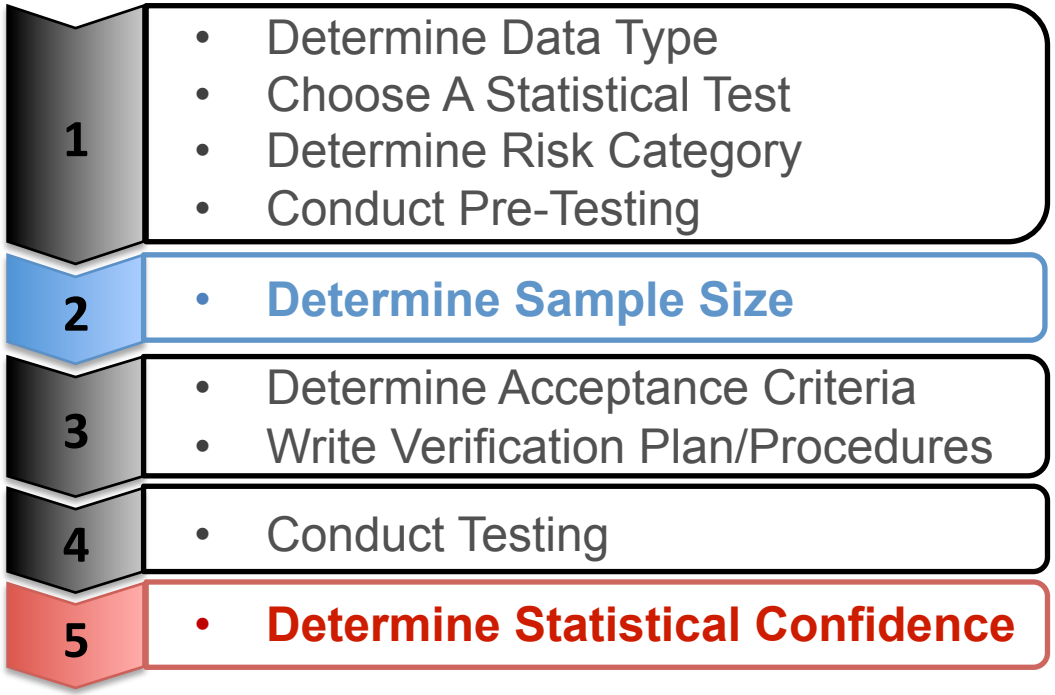
But I'm Not a Statistician!

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016

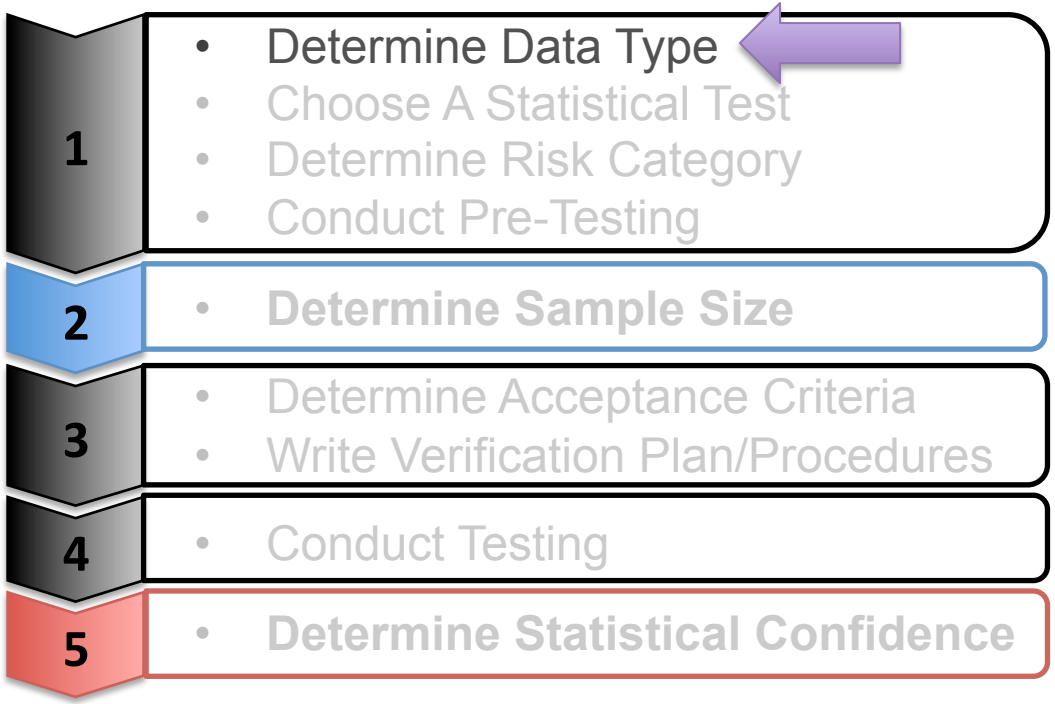
- You don't have to be (neither am I)
  - The hard math has been done for you

- You do need some basic statistics knowledge
  - Mean
  - Standard Deviation
  - Normal Distribution

Basic Process

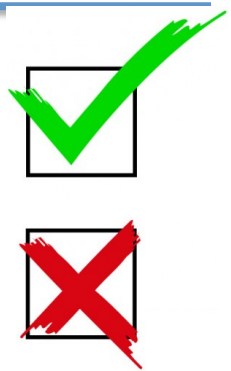


Basic Process

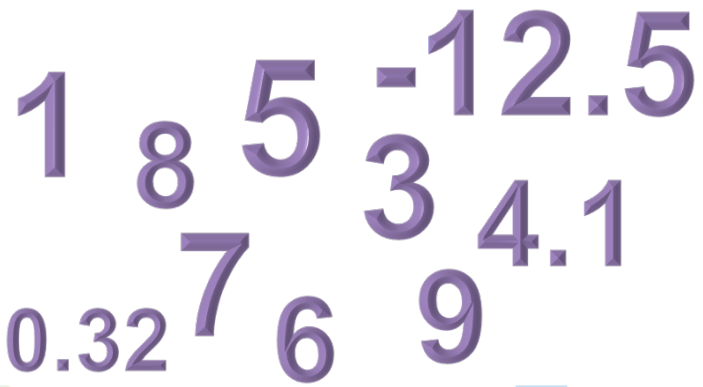


Determine Data Type

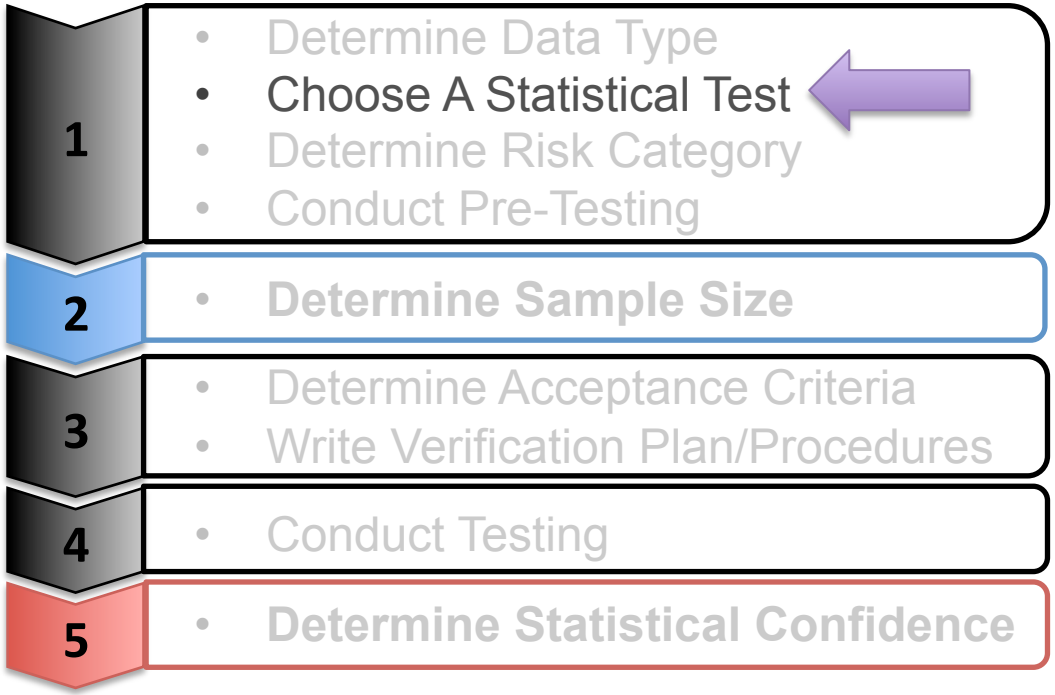
- Attribute Data
  - Have only two outcomes
  - Typically expressed as pass/fail



- Numeric Data
  - Have numbers as values
  - Numeric data are measured
  - Preferred

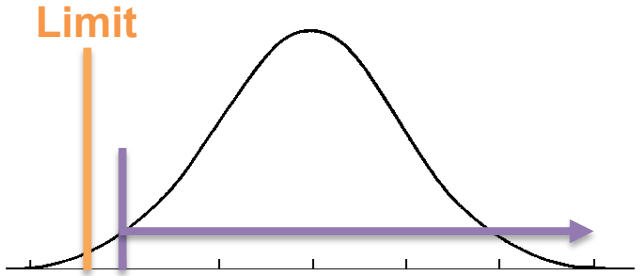


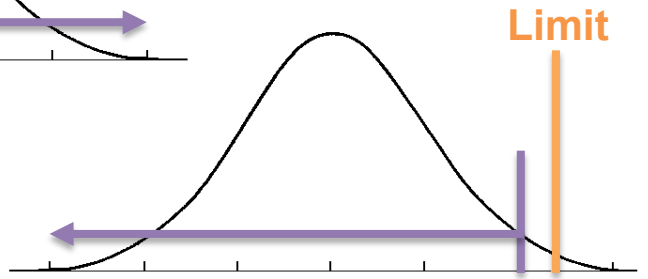
Basic Process

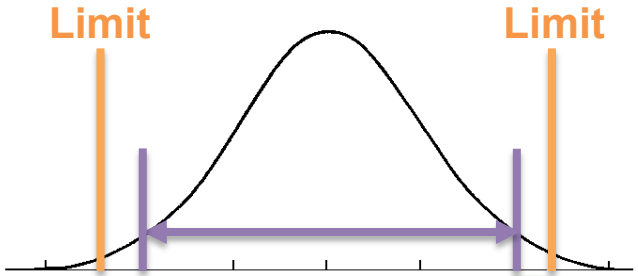


One or Two Sided?

- One-Sided
  - Above limit
  - Below limit
- Two-Sided
  - Between two limits

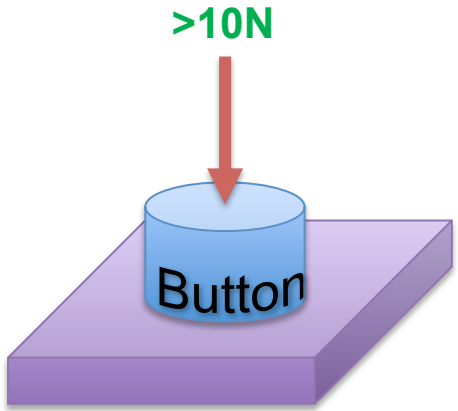






Statistical Tests

- Pass/Fail Tests
  - (Binomial)
- Tests on the Population
  - (Tolerance Interval)
- Tests on the Mean
  - (t-Tests)
  - (see paper)



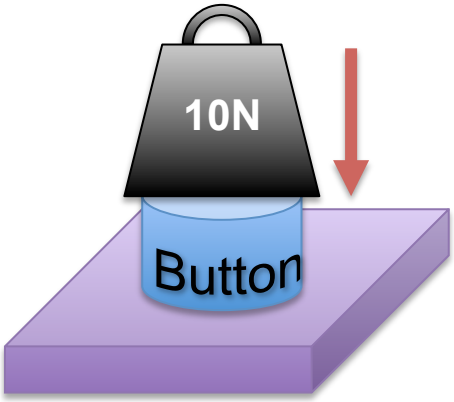
Example Requirement:
The button shall engage with >10N force.
(One-Sided Test)

Pass/Fail Tests (Binomial)

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016

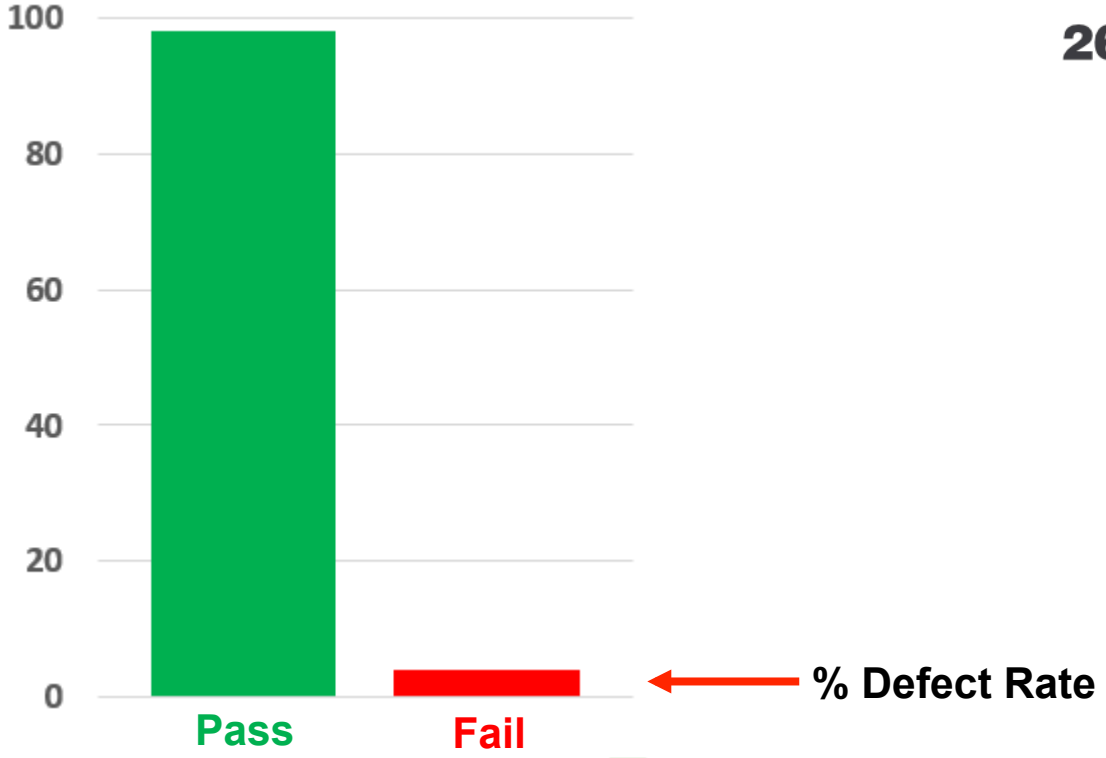
- Attribute (pass/fail) data
- Typically % of population is possibly defective
- OK if data is not normally distributed

- Example: Apply a known load (10N) to a set of buttons, to get a series of pass/fail results
  - (if it depresses, it fails)



Claim: There is 90% confidence that the button defect rate is less than 10%.

Binomial Test Visualized

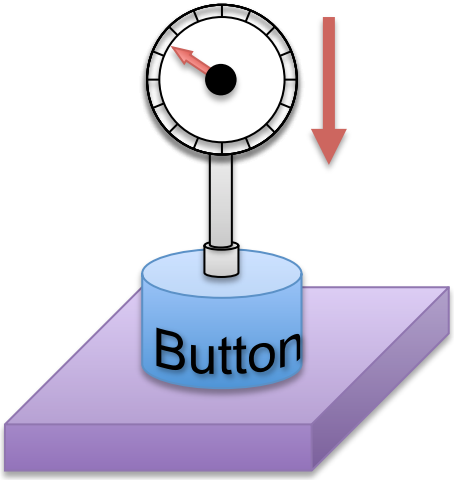


Tests on the Population (Tolerance Interval)

- Numeric data
- % of the population is above/below/within particular values
- Data must have a normal distribution

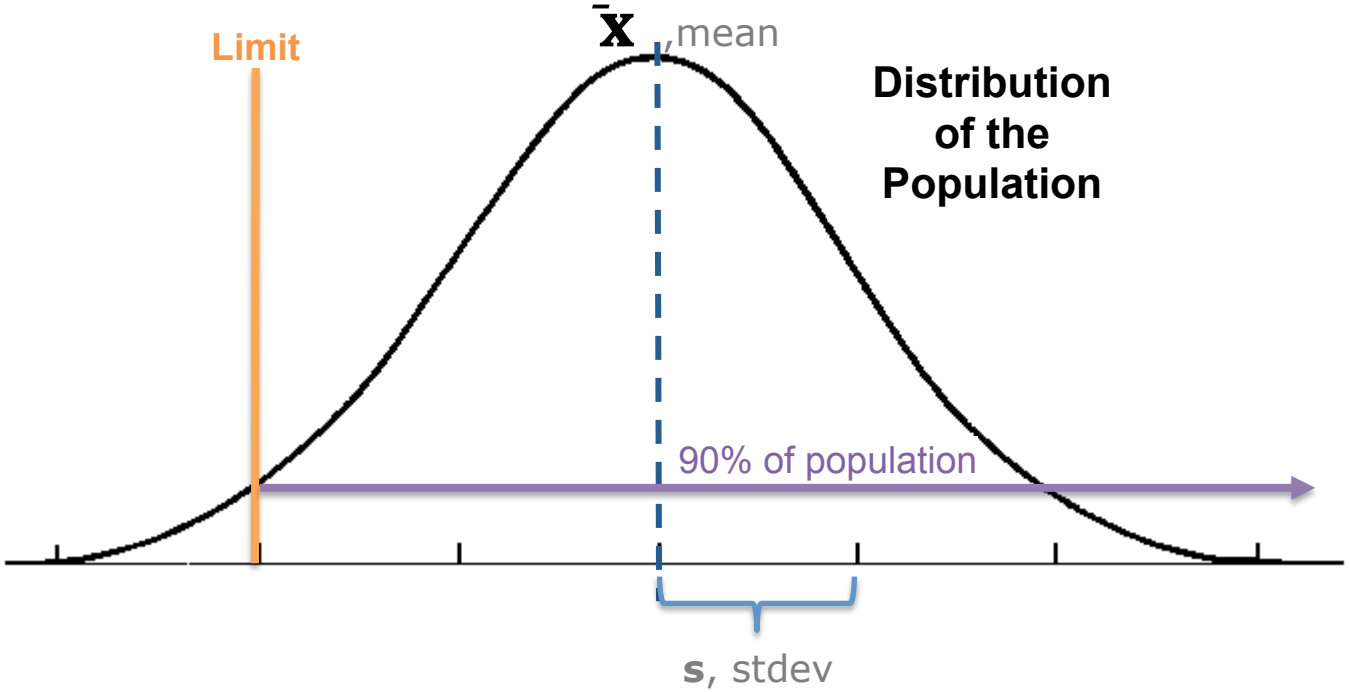
- Example: Measure the push force of a set of buttons with calibrated equipment and get numeric data
  - E.g. 12.1N, 10.5N, 9.7N, etc.

Claim: There is 90% confidence that 90% of the buttons require >10N push force.

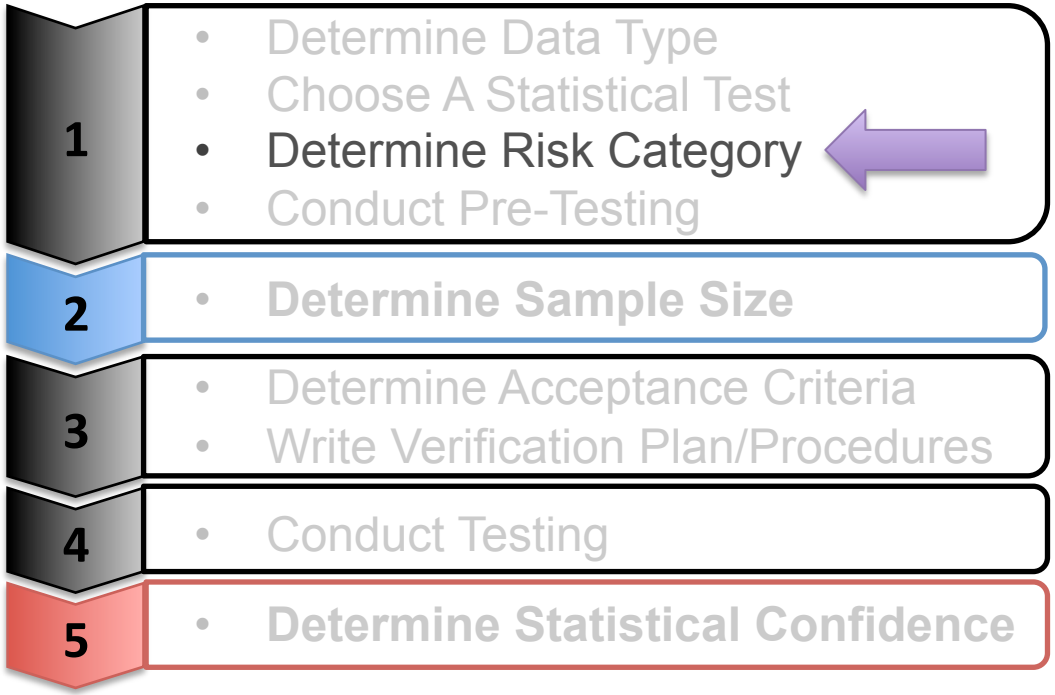


Tolerance Interval Test Visualized

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016



Basic Process



Risk Categories (Typical Med. Device)

Risk categories for requirements that…

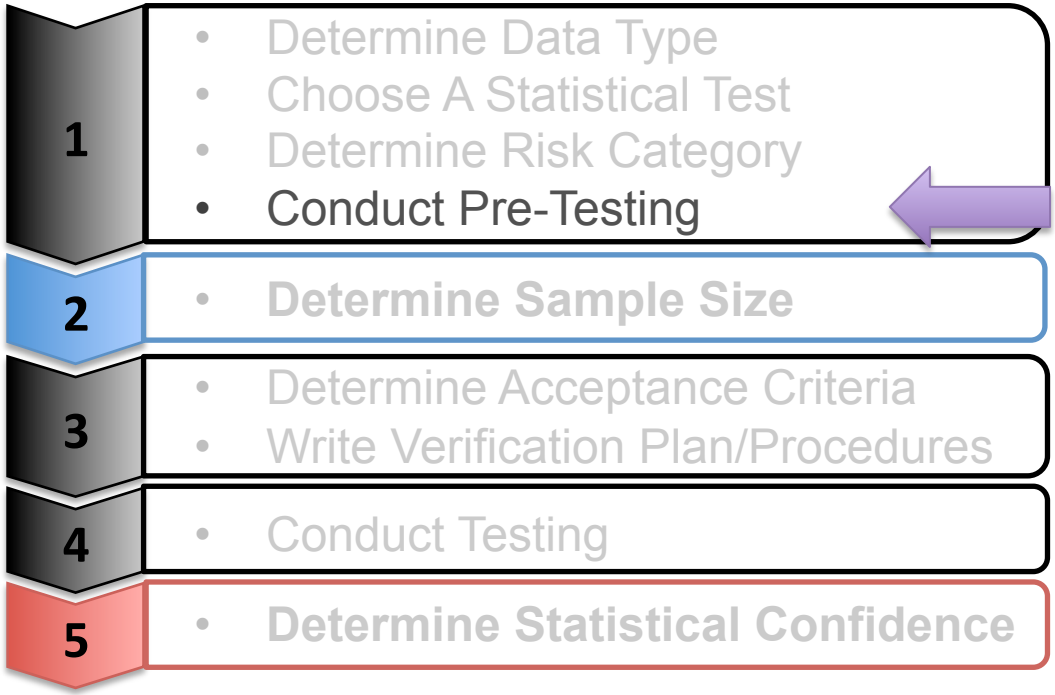
- **Minor Risk**
  - Not used to mitigate risks
  - Not critical to device function
  - Nuisance injuries
- **Moderate Risk**
  - Important to device functionality
  - Moderate damage
  - Moderate injury
- **Critical Risk**
  - Serious damage
  - Serious injury
  - Death

Risk Parameters

- You are really testing a risk level
- Establish the parameters to use in later calculations
- True failure rates must be much lower to have a chance at passing

Test Type	Minor Risk	Moderate Risk	Critical Risk
Pass/Fail Tests (Binomial)	90% Confidence 10% Defect Rate	95% Confidence 5% Defect Rate	95% Confidence 1% Defect Rate
Tests on the Population (Tolerance Interval)	90% Confidence 90% Population	95% Confidence 95% Population	95% Confidence 99% Population

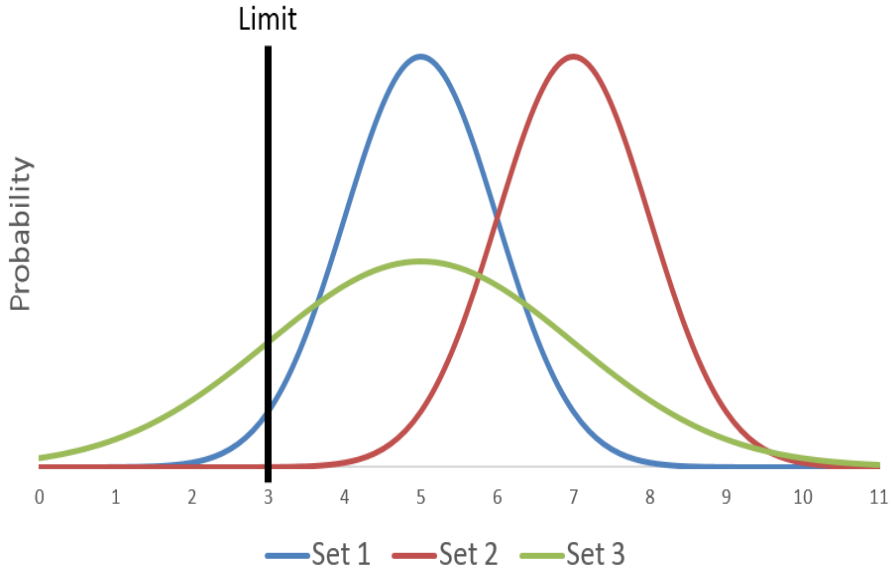
Basic Process



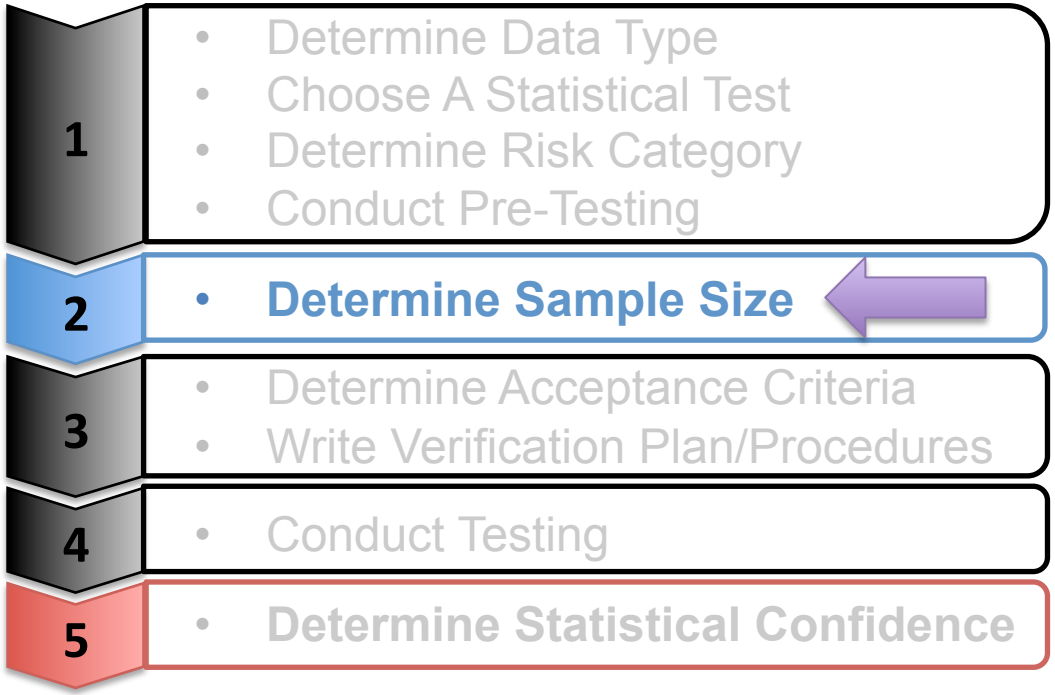
Gather Preliminary Data

- How many samples do I need? It depends…
- Need to **estimate** the **mean** and **standard deviation**
  - Pre-testing
  - Analysis
  - Prior product experience
- Until you know how close you are to your limit, it is impossible to know how many samples you need (numeric data)

Example: Which has >95% population above limit?



Basic Process



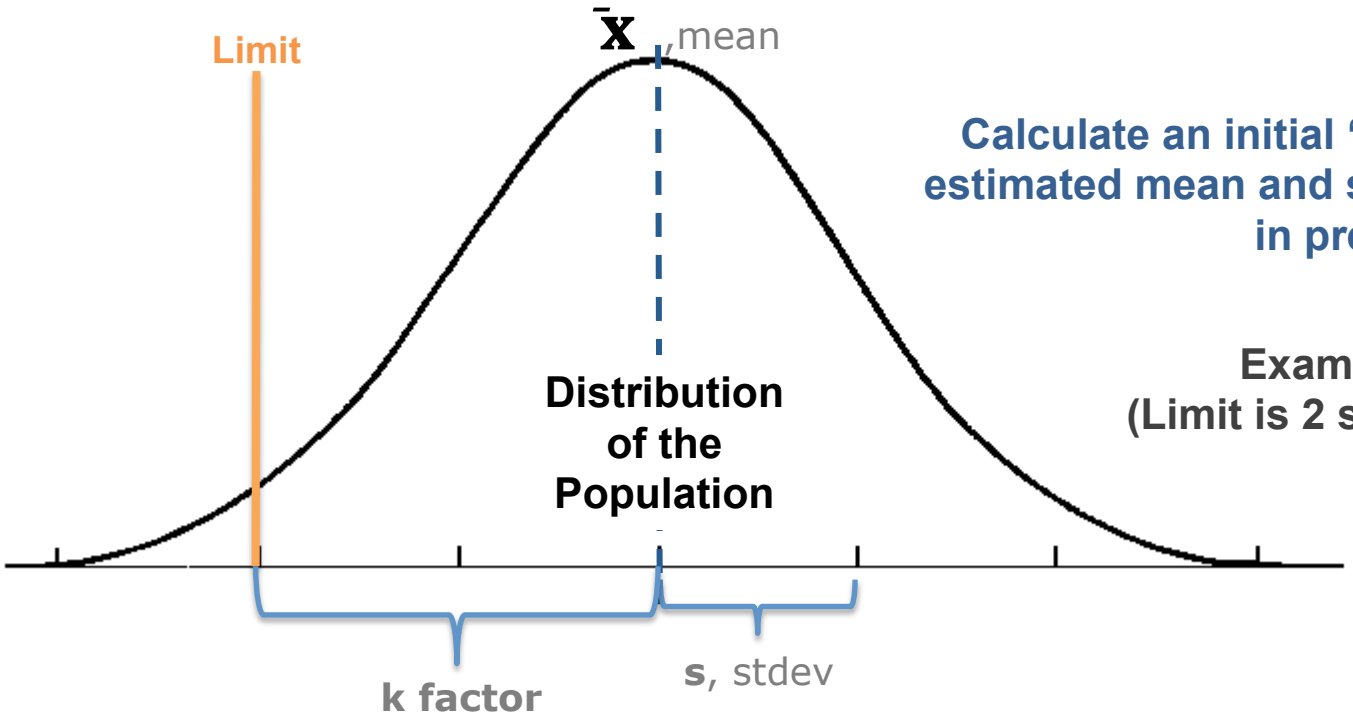
Pass/Fail Test (Binomial) # of Samples

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016

- Can use sequential testing

	90% Confidence				95% Confidence				99% Confidence			
Defect %	0	1	2	3	0	1	2	3	0	1	2	3
<1%	230	388	531	667	299	473	628	773	459	662	838	1001
<5%	45	77	105	132	59	93	124	153	90	130	165	198
<10%	22	38	52	65	29	46	61	76	44	64	81	97

Tests on the Population (Tolerance Interval) # of Samples



Calculate an initial "k factor" based on the estimated mean and standard deviation found in pre-testing

Example: k factor = 2
(Limit is 2 st dev from the mean)

Tests on the Population (Tolerance Interval) # of Samples

Samples	k Factors for 90% Confidence, 90% Population
5	2.743
6	2.494
7	2.333
8	2.219
9	2.133
10	2.066
15	1.867
20	1.766
50	1.560
100	1.471
200	1.412
300	1.387

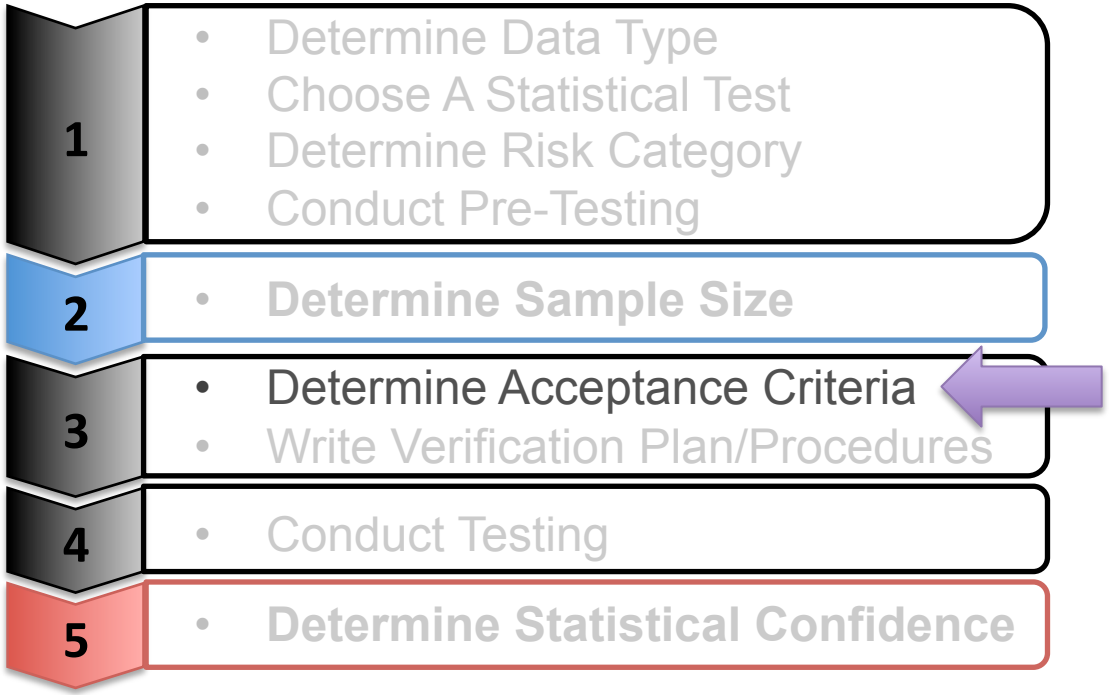
- Look up # of samples in table
- Button Example:
  - Limit = 10
  - Estimated mean = 11
  - Estimated st dev = 0.5
  - k factor = 2
  - For 90%/90% we need 15 samples (round up # samples)

What if I don't have Enough Samples?

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016

- Samples ≠ Devices
- Samples can be multiple measures from one device
  - In this case you are making claims about the future performance of **one particular device**, not the population of devices
- You can calculate confidence intervals for any number of samples
  - See references, not done in paper for arbitrary #'s
- <5 samples and the confidence intervals get very wide

Basic Process



Determine Acceptance Criteria

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016

- **Binomial**

  There is 90% confidence that the defect rate is less than 10% if no more than 0/22 or 1/38 or 2/52 or 3/65 samples fail.

- **Tolerance Interval**
  - $\bar{x}$ = Mean from pre-testing
  - s = Standard deviation from pre-testing
  - Lookup k factor

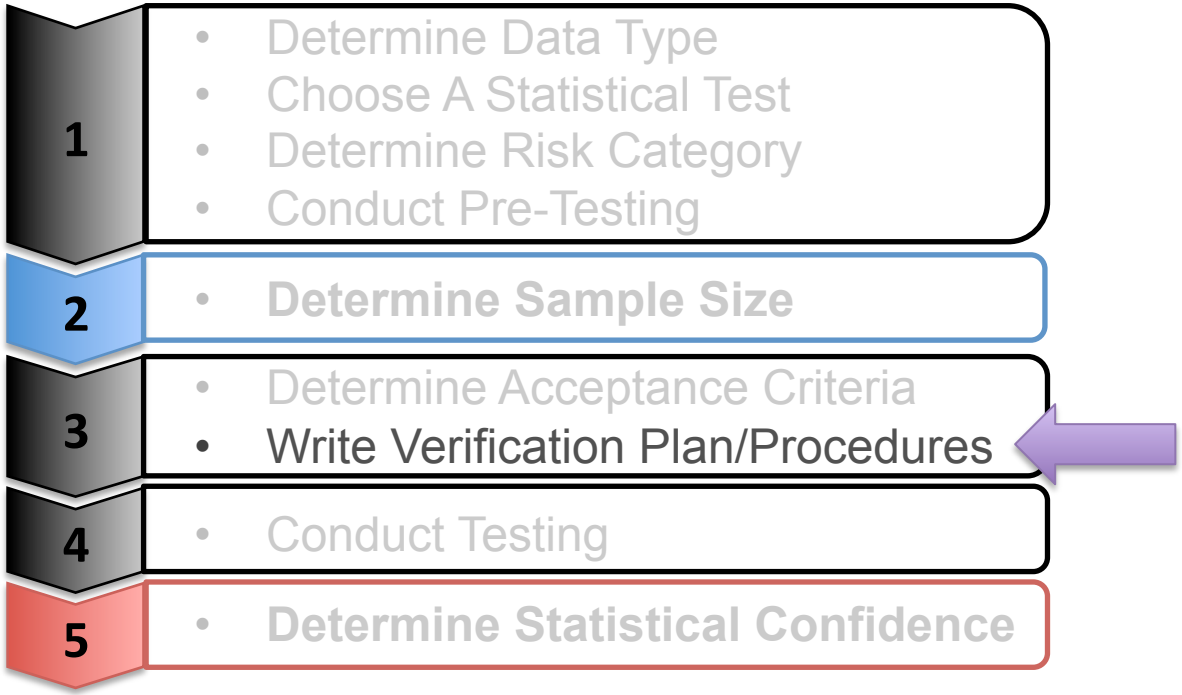
There is 90% confidence that 90% of the population is greater than 10 if:

$$10 < (\bar{x} - 1.867\ s)$$ *with 15 samples*

k factor

 # samples associated with k factor

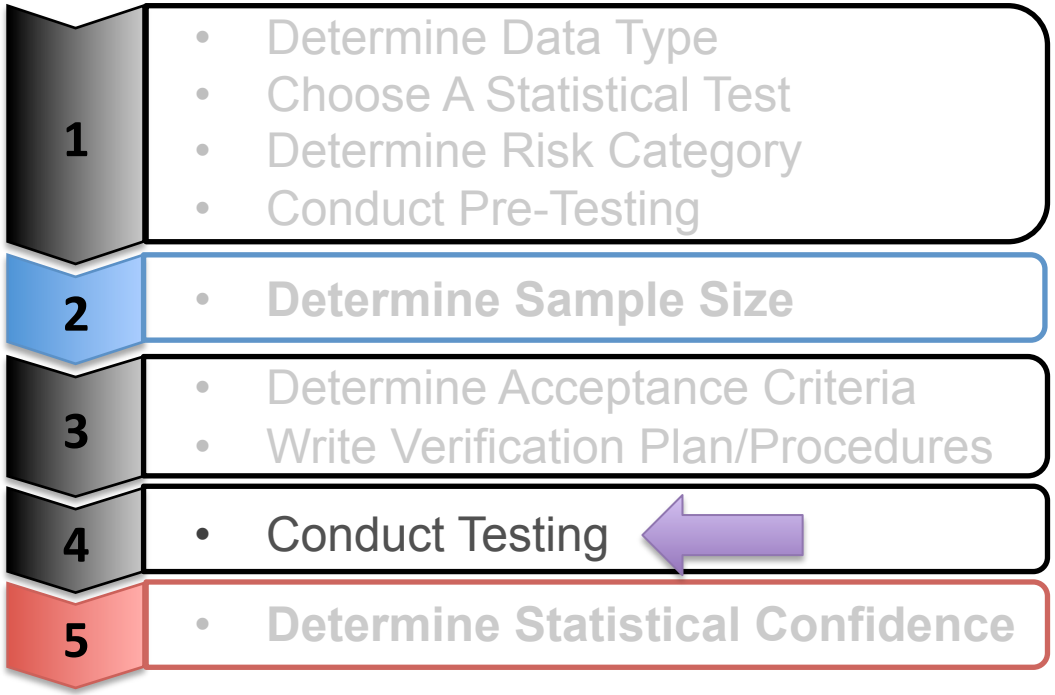
Basic Process



Write Verification Plan/Procedures

- Create the plan and procedure for testing the requirement
- Write the acceptance criteria directly into the verification plan or procedure

Basic Process



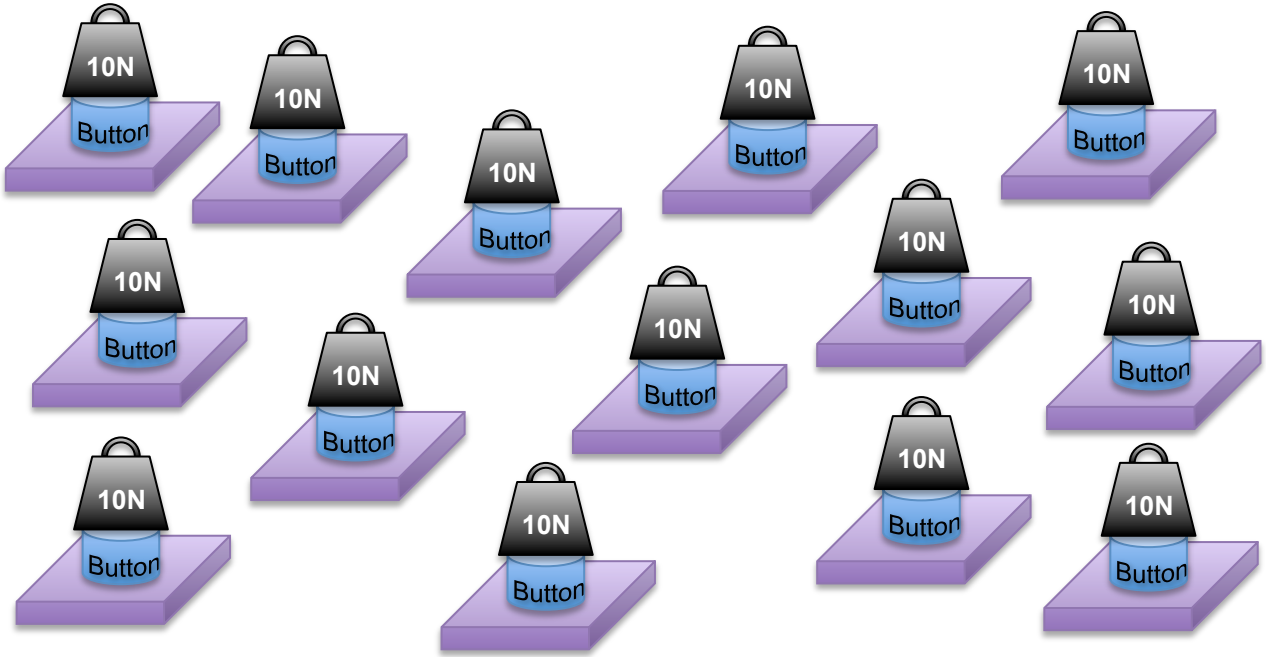
Conduct Testing

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016

- Use the number of samples determined by the process
- Take attribute or numeric data as appropriate
- Calculate number of pass/fails, or
- Sample mean and sample standard deviation

Conduct Testing (Binomial)

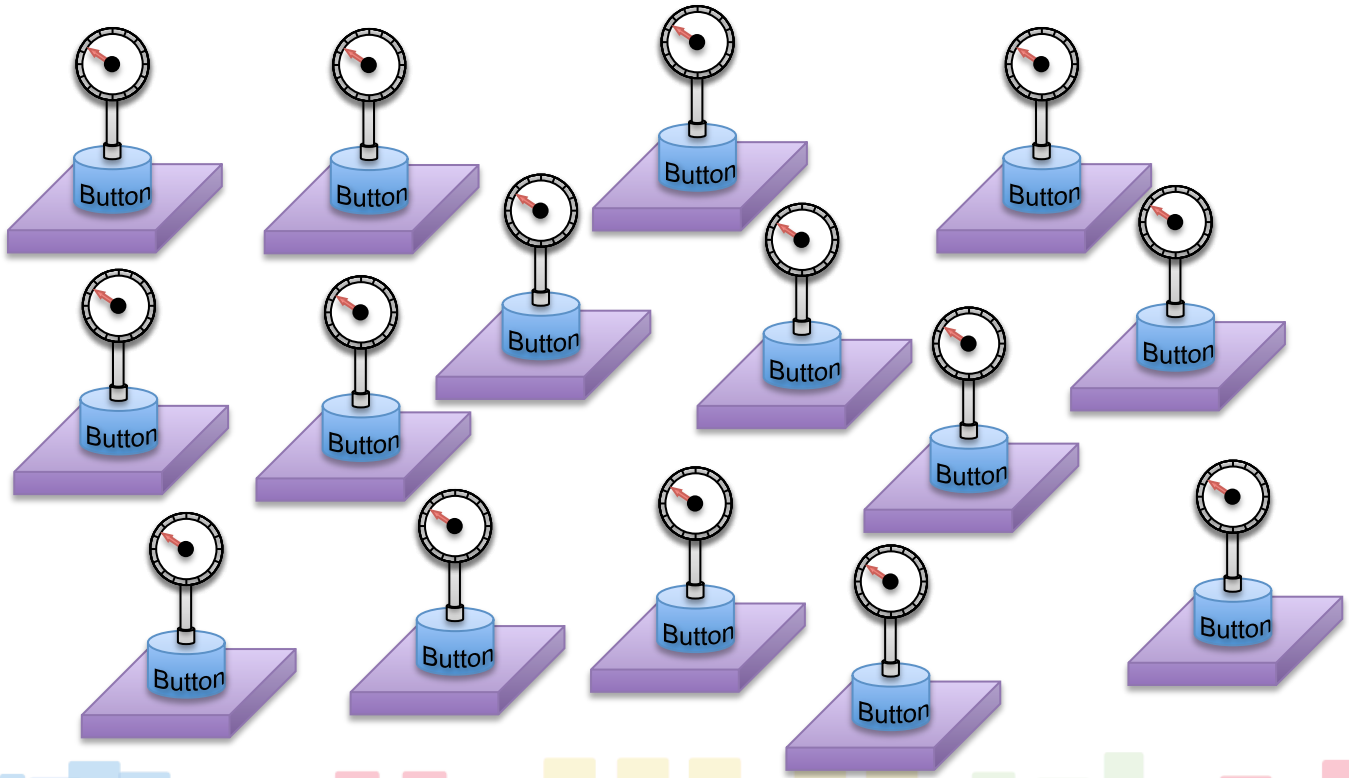
26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016



= # Passes;
Fails

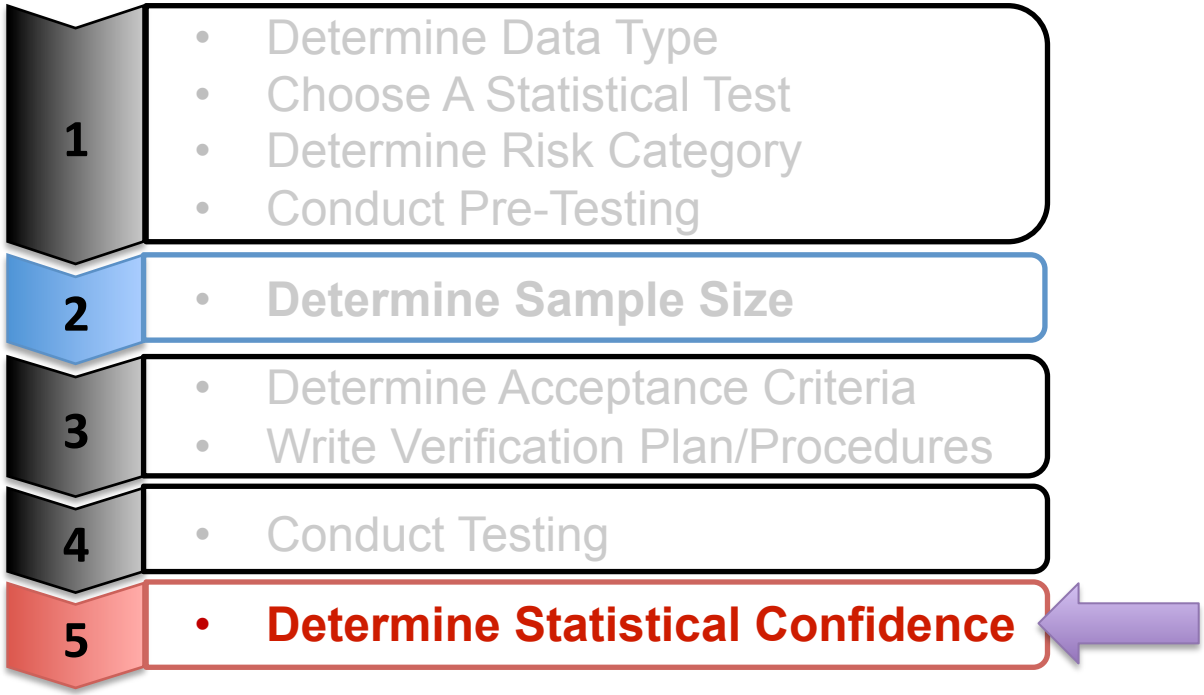
Conduct Testing (Tol Int)

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016



$= \bar{x}, s$

Basic Process



Determine Statistical Confidence

Pass/Fail Test (Binomial)

Acceptance criteria was:

There is 90% confidence that the defect rate is less than 10% if no more than

0/22 or 1/38 or 2/52 or 3/65 samples fail.

Test result:

Tested 38 units with 1 failure. Therefore there is 90% confidence that the defect rate for the button actuation force is less than 10%.

Determine Statistical Confidence

Test on the Population (Tolerance Interval)

Acceptance criteria was:

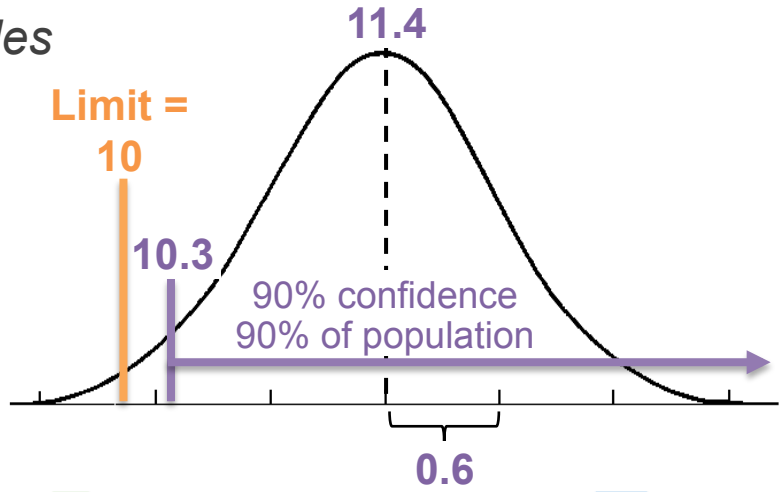
There is 90% confidence that 90% of the population is greater than 10 if:

$10 < (\bar{x} - 1.867\ s)$ with 15 samples

Test result:

$10 < (11.4 - 1.867 * 0.6)$

$10 < 10.3$



Therefore, there is 90% confidence that 90% of the button force population is greater than 10.

But I have a lot of Requirements!

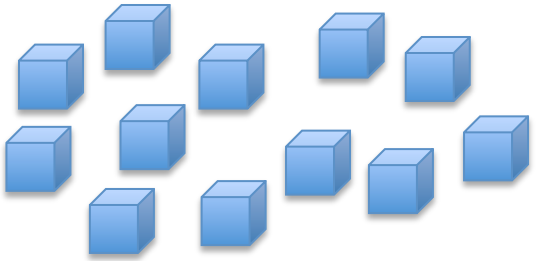
26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016

- So do I!
- Statistical verification is not required for every requirement
- Inspections, demonstrations, and analysis typically do not require it
- Features that are not expected to vary lot to lot may be exempt

Remember

- Statistical verification enables you to:

Know how many samples to test



Have statistical confidence in your results

Conclusion

- Statistical verification techniques are accessible to engineers for common test types
  - We have done the research and math for you
  - Uses straightforward equations and lookup tables
  - Step by step approach
  - Repeatable process
  - Justifiable
  - Defensible

26th annual INCOSE international symposium
Edinburgh, UK
July 18 - 21, 2016

A Practical Guide to Statistical Verification

James Kolozs

syncroness™
complex problems inspired solutions