False Alarm vs. Missed Signal

A Thought-provoking Seminar

ASTM D02 Committee Meeting
Dec / 2018, Atlanta, GA

Presented by:
Alex T.C. Lau, Chairman
D02.CS94 Quality Assurance and Statistics

Seminar Objective

Promote awareness and understanding on the two types 'wrong decisions' and associated probabilities when making decisions based on "statistical inference" from *limited* data

Seminar Motivation

- clear up general ‘fuzzy’ understanding on statistical inference
- provide illumination on how to interpret “not statistically significant” and the need to simultaneously consider “power of detection”

Hopefully, this would provoke more thoughts and questions when voters read ballots “justified” by statistical “claims”, and ask “the missing question”

High level agenda

- Statistical inference with “Imperfect” data
- The missing question
- Two types of “wrong decisions”

Examples:

- Is my test method in statistical control ?
- Is there a bias between two test methods ?
- Is “it” there ?
- Is the product ‘on-spec’ ?

“Imperfect” data

Context:

1. We have a certain “belief” about the true state of affairs (whatever that may be, which, is application dependent)

2. We wish to make a *decision* as to whether this “belief” is reasonable… but, we don’t have the will or resources or the luxury to undertake an exhaustively comprehensive study and/or collect an “infinite” amount of data (i.e.: perfect data)

3. We collect a limited amount of data, do some math, and use a

‘classical statistical inferential approach’

to help decide whether our belief is ‘justified’ (or not…heaven forbid !)

Classical Statistical Inferential Approach

(grossly simplified)

1. collect some data, do some math, obtain an outcome

2. select an “appropriate” probability model that can *reasonably* represent the data *given* the *belief* is true

3. ask the question: if our *belief* is true, what’s the likelihood (probability) of the mathematical outcome

4. use the selected probability model in 2. (Normal model most common) to get the answer, then make a decision as to whether to *accept* or *reject* our *belief* by inference

Statistical Inference – significant outcome

- if the probability (*P*) of obtaining the mathematical outcome is < X % (typically set at 5%) based on the selected probability model, you <u>reject the belief</u> using the following "rationale":

→ there's a less than 5% chance of obtaining this outcome if the *belief* is true

In statistical jargon, this is referred to as:
"a statistically significant outcome at 5%"
usually abbreviated as "5% significance"

Statistical Inference – insignificant outcome

- if the probability (*P*) is >= X % (typically set at 5%) based on the selected probability model, you <u>do NOT reject the belief</u> using the following "rationale":

→ there is no compelling evidence to suggest the belief is not true because there's >= 5% of this outcome due to chance (random sampling) if the belief is true

In statistical jargon, this is referred to as:
"a statistically insignificant outcome at 5%"
usually abbreviated as "not significant at 5%"

Formal name for the classical inferential approach described in the previous slides is:

Hypothesis Testing

The 'belief' is called the "*null* " hypothesis, represented by the symbol H_0

My personal observation

- what conclusion can (or cannot) be drawn from classical inferential Hypothesis Testing is most misunderstood amongst laypersons,

and,

- sometimes *conveniently* 'mis-used' by those that want to 'lead' the laypersons to draw a *desired* conclusion

Why the classical Hypothesis testing approach is most mis-understood

- classical approach is intended to 'disprove', or 'reject' the belief H_0 → 'reject' is viewed as 'success' by the experimenter
- failure to reject H_0 does NOT 'prove' it's true
- most laypersons view 'failure to reject' as a 'success', and *incorrectly* draw the conclusion that an insignificant outcome (H_0 not rejected) 'proves' it's true

The Missing Question

Correct interpretation of classical Hypothesis Testing outcome must take into account the answer to the following question:

If my 'belief' of the true state of affairs (H_0) is wrong, and the alternate truth is "-----", what is the probability that H_0 would *not* be rejected using the classical inferential approach ?

caveat

Answer to the Missing Question may be inconvenient

(for some entities)

Two types of 'wrong' decisions

When you make decision using ‘imperfect’ data in the context of ‘hypothesis test’ described in the previous slides, there is a finite probability of :

<u>making the ‘wrong’ decision</u>

There are two types of ‘wrong’ decisions you can make with regards to whether your ‘belief’ is reasonable:

- you incorrectly reject your ‘belief’ (i.e.: your ‘belief is correct, but you decide to reject it)

or

- you incorrectly accept your ‘belief’ (i.e.: your belief is wrong, but you decide to accept it)

Incorrect “reject” vs. Failure to “reject”

Type I, or, α error

Type II, or, β error

H_0 : stated belief

Two types of decision errors		your decision: accept "belief"	your decision: reject "belief"
True State of affairs	"belief" is correct	correct	Type I error : incorrect reject
True State of affairs	"belief" is not correct	Type II error : failure to correctly reject	correct

cannot be independently set

Terminology Challenge (!!)

In manufacturing:

- Type I error is referred to as *producer’s risk*
- Type II error is referred to as *consumer’s risk*

In SPC (control charts):

- Type I error is referred to as *false alarm*
- Type II error is referred to *missed signal*

In medical screening:

- Type I error is referred to as *false positive*
- Type II error is referred to as *false negative*

In environmental testing for toxins:

- Type I error is referred to as *false detect*
- Type II error is referred to as *false no detect*

The Missing Question has to be associated with a pre-defined "----"

If my 'belief' of the true state of affairs (H_0) is wrong, and the alternate truth is "-----", what is the probability that H_0 would *not* be rejected using the classical inferential approach ?

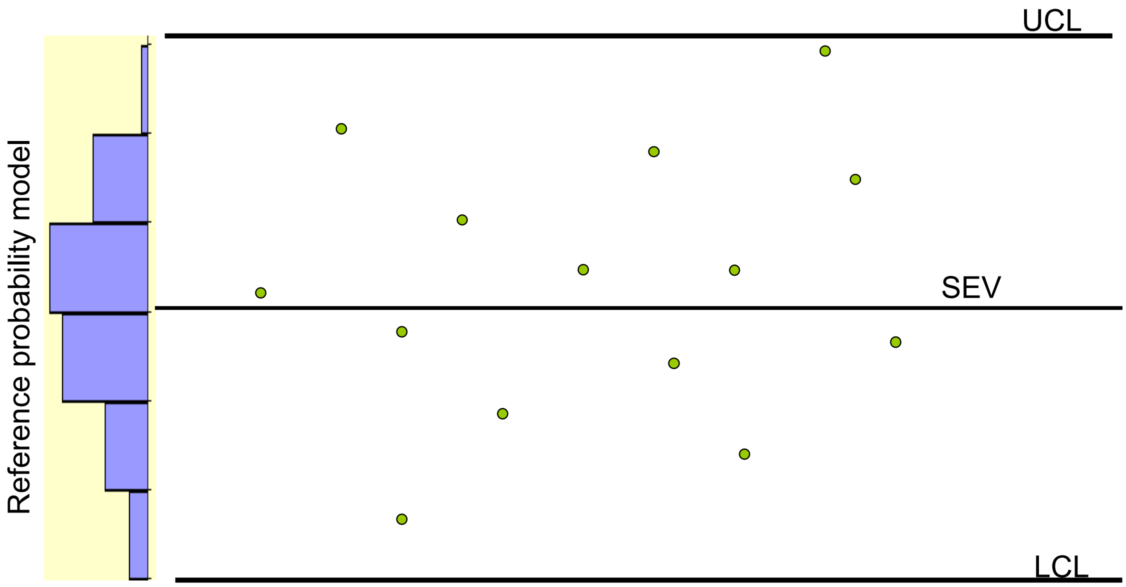
The "----" is a statement articulating the alternate 'truth' to your 'belief', known as the Alternate Hypothesis, represented by the symbol H_a

Example 1

Test Method SQC using the Individual Control Chart

(ASTM D6299)

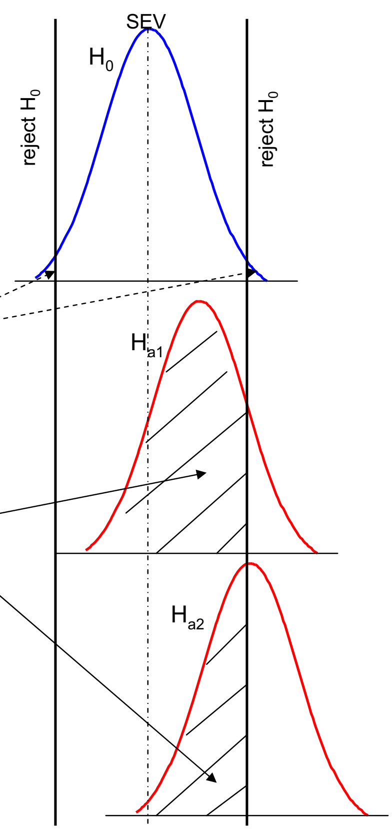
Typical I-chart



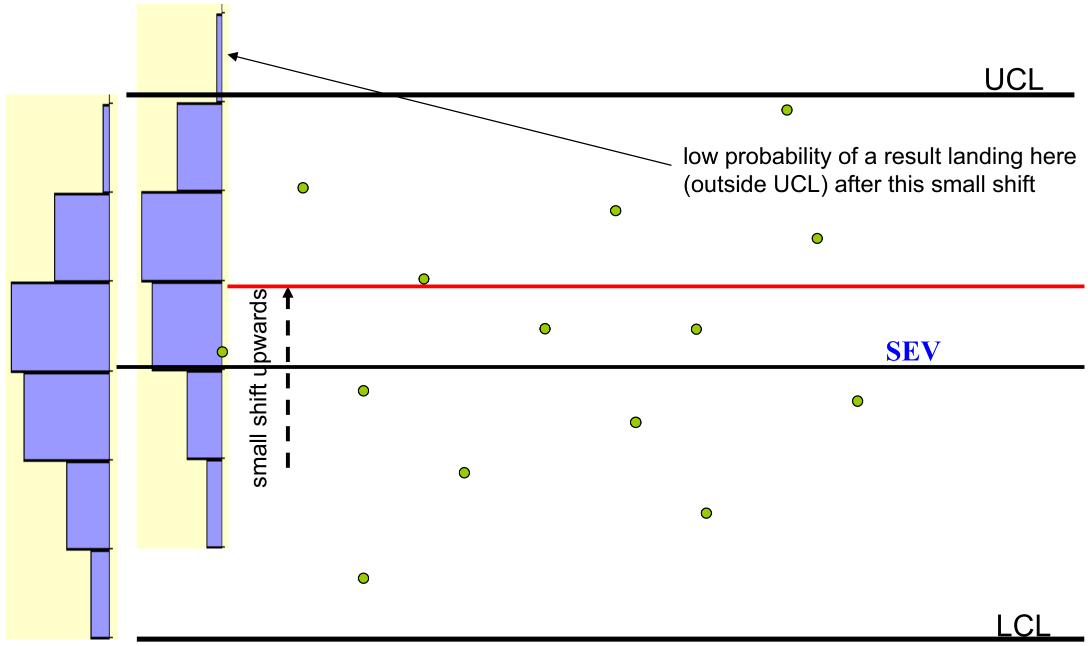
Data: regular SQC results on a Control Sample

H_0: Method is in statistical control

Two types of decision errors		your decision: Declare in control	your decision: Declare out of control
True State of affairs	Method is in statistical control	correct	Type I error : incorrect reject
True State of affairs	Method is not in statistical control	Type II error : failure to correctly reject	correct



Main shortcoming of the I chart => insensitive to small shifts



Site Precision conditions : long time / many operators

Power of Detection

- Type II (β) error decreases as the 'alternate truth' deviates farther and farther from your 'belief'

- (1 − β) is referred to as the *power of detection* at various alternate $H_{a's}$

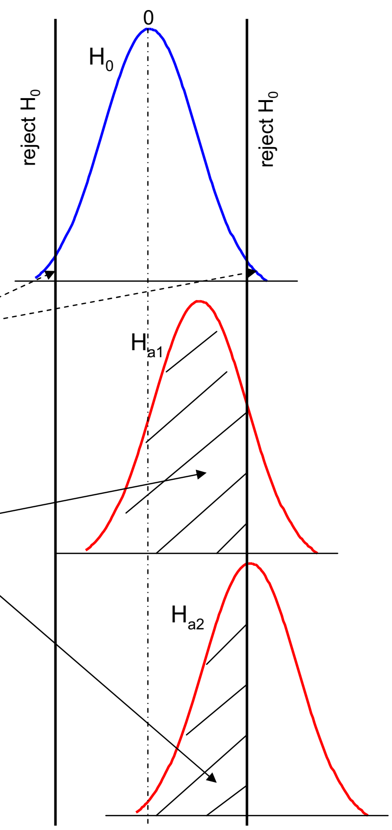
Example 2

Is there a bias between two methods ?

Data: difference between Method A and Method B on multiple comparison samples

H_0: Methods are not biased versus each other

Two types of decision errors		your decision	
		Declare methods not biased	***Declare methods biased***
True State of affairs	Methods are not biased	correct	Type I error : incorrect reject
	Methods are biased	Type II error : failure to correctly reject	correct



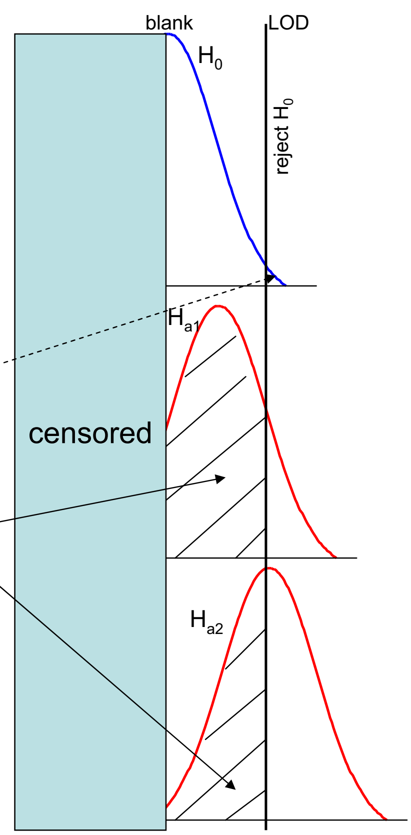
Example 3

Is “it” there ?

Data: test for a Toxin in drinking water

H_0: Toxin not present

Two types of decision errors		your decision	
		Declare Toxin present	***Declare Toxin not present***
True State of affairs	Toxin not present	correct	Type I error : incorrect reject
	Toxin present	Type II error : failure to correctly reject	correct



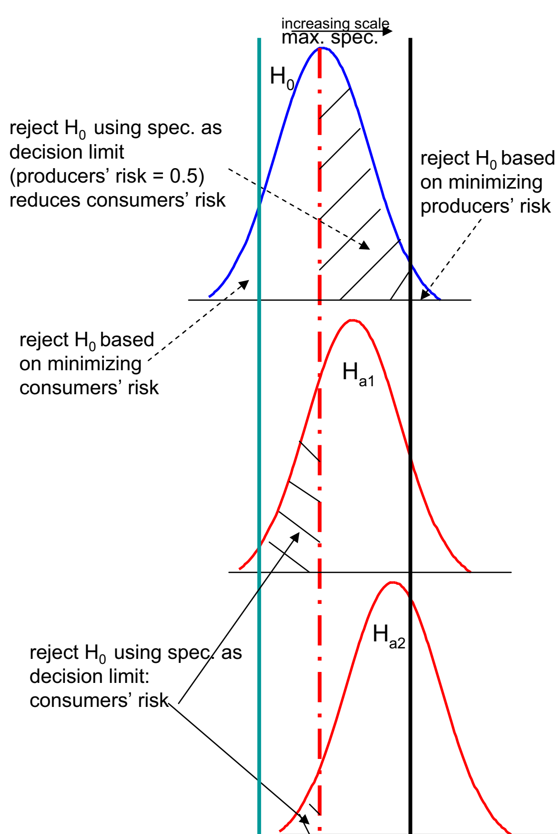
Example 4

Is the product ‘on spec’ ?

Data: test result on product with **max** spec

H_0: True property value <= max spec

Two types of decision errors		your decision	
		Declare on spec	***Declare off spec***
True State of affairs	True property value <= max spec	correct	producer’s risk : incorrect reject
	True property value > max spec	Type II error : failure to correctly reject	correct



A plug for our new Guide

Designation: D8146 – 18 An American National Standard

Standard Guide for
Evaluating Test Method Capability and Fitness for Use[1]

1. Scope

1.1 This guide covers techniques for evaluating the statistical capability and fitness for use of standard test methods used for measuring properties of petroleum products, liquid fuels, and lubricants. Specifically, this guide provides strategies for evaluating the capability of a test method to provide a sufficiently precise estimate of the intended parameter versus a given level or value of that parameter and for assessing, with sufficient confidence, the fitness for use of a test method for determining the acceptability of products versus specification, regulatory, or manufacturing limits.

Closing Messages

- A significant outcome (reject H_0) at a low Type I (α) error (low false reject) is evidence that the 'test' has sufficient power, so it's a good thing → H_o is likely not true

- An insignificant outcome should always be interpreted in context by asking the missing question:
  - → what is the *power of detection* if H_0 is false, and the truth is H_a ; choose H_a at a level that is suitable for your application
  - → if the *power* is low, it's likely that you need more data