# Data Analysis in Software Engineering. Part 4

Classical Hypothesis Testing

- By "classical" we mean the standard "frequentist" approach to hypothesis testing. The "frequentist" approach to probability sees it as the frequency of events in the long run. We repeat experiments over and over and we count the times that our object of interest appears in the sequence.
- The classical approach is usually called **null hypothesis significance testing** (NHST) because the process starts by setting a null hypothesis H_0 which is the opposite about what we think is true.
- The rationale of the process is that the statistical hypothesis should be *falsifiable*, that is, we can find evidence that the hypothesis is not true. We try to find evidence against the null hypothesis in order to support our alternative hypothesis H_A
- Usually, the null hypothesis is described as the situation of "no effect" and the alternative hypothesis describes the effect that we are looking for.
- After collecting data, taking an actual sample, we measure the distance of our parameter of interest from the hypothesized population parameter, and use the facts of the sampling distribution to determine the probability of obtaining such a sample *assuming the hypothesis is true*. This is amounts to a test of the hypothesis.
- If the probability of our sample, given the null hypothesis is high, this provides evidence that the null hypothesis is true. Conversely, if the probability of the sample is low (given the hypothesis), this is evidence against the null hypothesis. The hypothesis being tested in this way is named the *null hypothesis.*
- The goal of the test is to determine if the null hypothesis can be rejected. A statistical test can either reject or fail to reject a null hypothesis, but never prove it true.
- We can make two types of errors: false positive (Type I) and false negative (Type II)
- Type I and Type II errors

		Null hypothesis (H_0) is	
		Valid/True	**Invalid/False**
Judgement of Null Hypothesis (H_0)	**Reject**	Type I error False Positive	Correct inference True Positive
	Fail to reject	Correct inference True Negative	Type II error False negative

- Two-tailed NHST
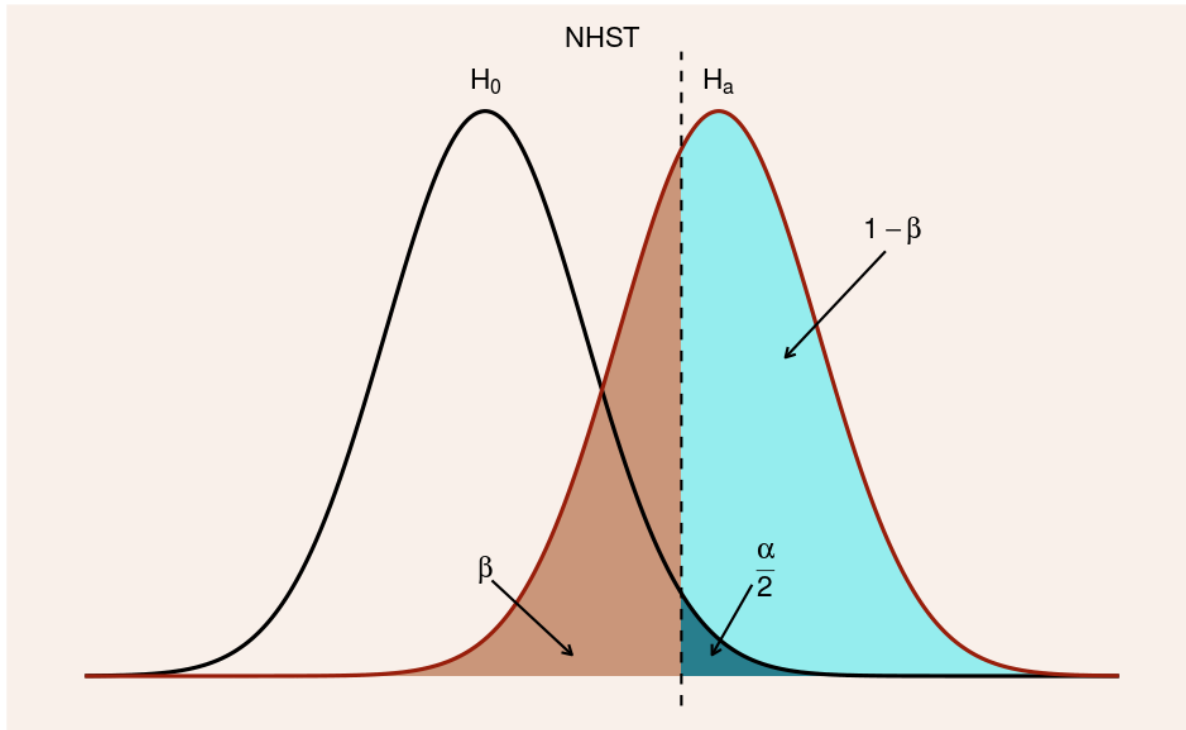
- One-tailed NHST
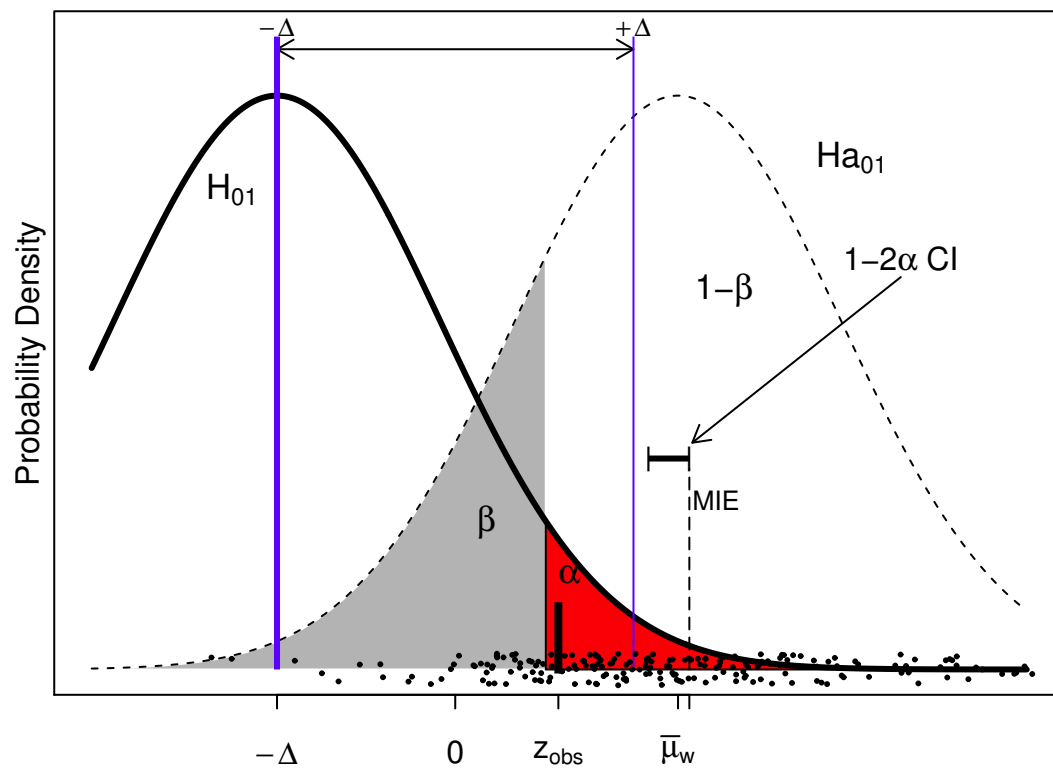
- elementary example

```
data = c(52.7, 53.9, 41.7, 71.5, 47.6, 55.1, 62.2, 56.5, 33.4, 61.8, 54.3, 50.0, 45.3, 63.4, 53.9, 65.5
t.test(data, mu=50, alternative = 'greater')
```

```
##
##  One Sample t-test
##
## data:  data
## t = 2.1562, df = 23, p-value = 0.02088
## alternative hypothesis: true mean is greater than 50
## 95 percent confidence interval:
##  50.88892      Inf
## sample estimates:
## mean of x
##  54.33333
```

- Keeping this simple, we could start hypothesis testing about one sample median with the wilcoxon test for non-normal distributions.

- "ae" is the absolute error in the China Test data

```
median(ae)
```

```
## [1] 866.7193
```

```
mean(ae)
```

```
## [1] 1866.819
```

```
wilcox.test(ae, mu=800, alternative = 'greater') #change the values of mu and see the results
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  ae
## V = 8990, p-value = 0.0007975
## alternative hypothesis: true location is greater than 800
```

- Quick introduction at https://psychstatsworkshop.wordpress.com/2014/08/06/lesson-9-hypothesis-testing/

p-values

- p-value: the p-value of a statistical test is the probability, computed assuming that H_0 is true, that the test statistic would take a value as extreme or more extreme than that actually observed.
- http://www.nature.com/news/psychology-journal-bans-p-values-1.17001
- https://www.sciencenews.org/blog/context/p-value-ban-small-step-journal-giant-leap-science

Psychology journal bans *P* values

Test for reliability of results 'too easy to pass', say editors.

Chris Woolston

26 February 2015 | Clarified:

ScienceNews
MAGAZINE OF THE SOCIETY FOR SCIENCE & THE PUBLIC

Explore ▾

LATEST   MOST VIEWED

NEWS IN BRIEF
Polar bears' 'walking hibernation'
not much of an energy saver
BY SUSAN MILIUS   JULY 16, 2015

NEWS
Good luck outsmarting a mosquito
BY SUSAN MILIUS   JULY 16, 2015

NEWS
Defense hormones guide plant

Context

SCIENCE PAST AND PRESENT
TOM SIEGFRIED

CONTEXT NUMBERS

P value ban: small step for a journal, giant leap for science

Editors reject flawed system of null hypothesis testing
BY TOM SIEGFRIED 3:18PM, MARCH 17, 2015

Search Science News...