# Probability and Statistical Analysis

Stochastic variables

Calculating probability

Normal distribution

Robust statistics


"Half of the numbers are accurate, that's why we're auditing the remaining 56%."

# Overview of statistic variables and methods

Discrete stochastic variable
- flipping a coin
- casting a die
- playing cards (poker)
- quantum states

Probability

$$P = \frac{\#\text{ of accessible states}}{\text{Total }\#\text{ of states}}$$

Continuous stochastic variable
- measurement noise
- experimental error

Least squares

Robust statistics

# Sources of error

I. Gross error

Example. grab the wrong vial.  NOTE: always label your vials.
Documentation is essential.

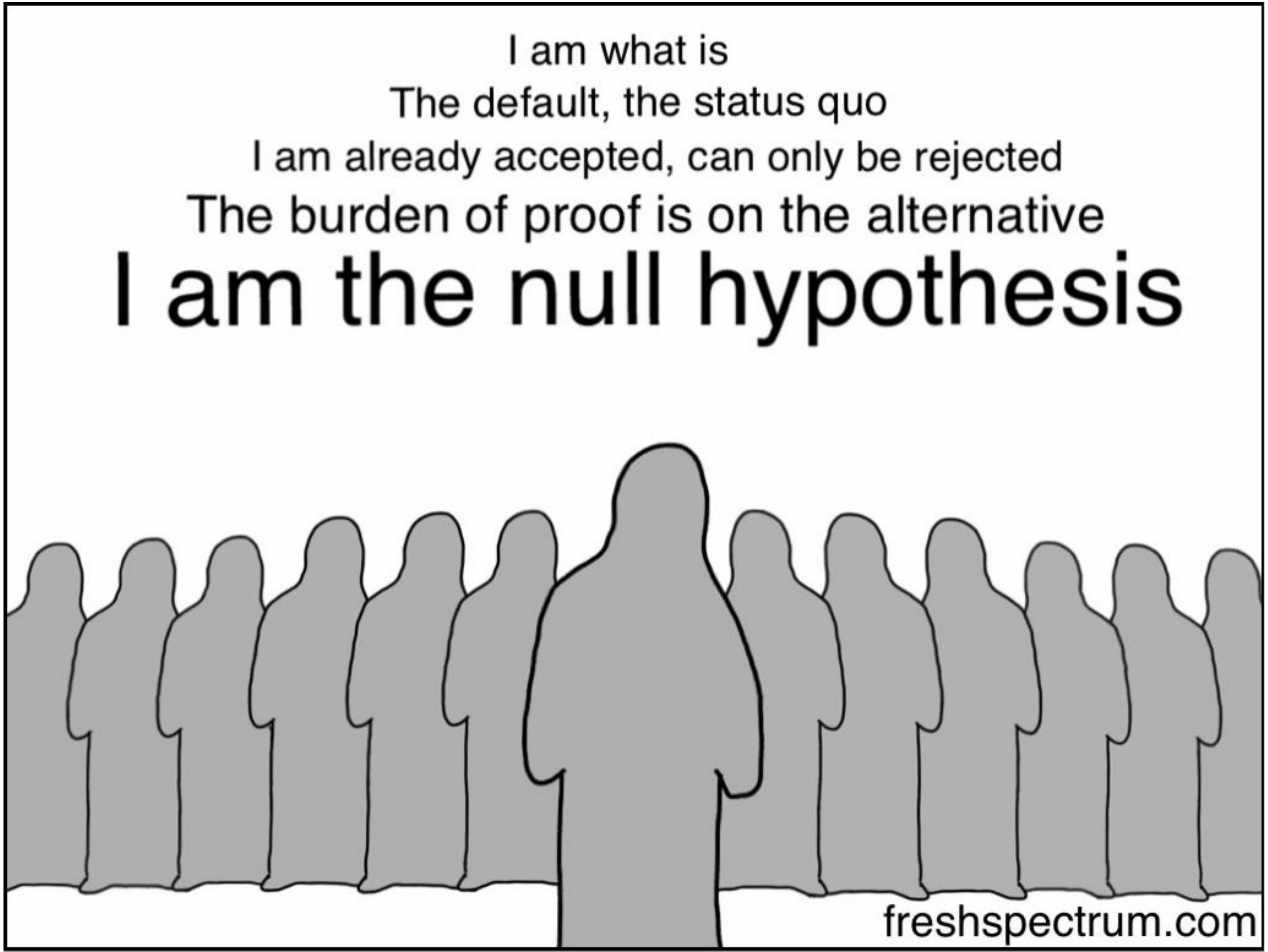
II. Systematic error

III. Random error

IV. Incomplete randomization

Example: weather, there is a short term predictability in the weather.
Incomplete mixing of solutions can give rise to this source of error.
Drift in electronic equipment. Error gets larger over time.
We want to prevent IV, by mixing carefully, using stabilized power supplies etc.

# The null hypothesis


I am what is
The default, the status quo
I am already accepted, can only be rejected
The burden of proof is on the alternative
I am the null hypothesis
freshspectrum.com

# $H_0$ = the null hypothesis

*Example:* $H_0: \mu_1 = \mu_2$

$\mu_1$ = the mean of population 1, and
$\mu_2$ = the mean of population 2.

A stronger null hypothesis is that the two samples are drawn from the same population, such that the variances and shapes of the distributions are also equal.

# $H_1$ = the alternative hypothesis

*Example:* $H_1: \mu_1 \neq \mu_2$

As before $\mu_1$ and $\mu_2$ are the two means. But, when we write $\mu_1 \neq \mu_2$, we must use a criterion to determine that the two means are different by an amount that is greater than the error. One convention is to use 2 $\sigma$. If the means of the two distributions are separated by a value greater than two standard deviations then we can say that the alternative hypothesis is validated and the null hypothesis is falsified. We can reject the null hypothesis.

# Type I and Type II Errors

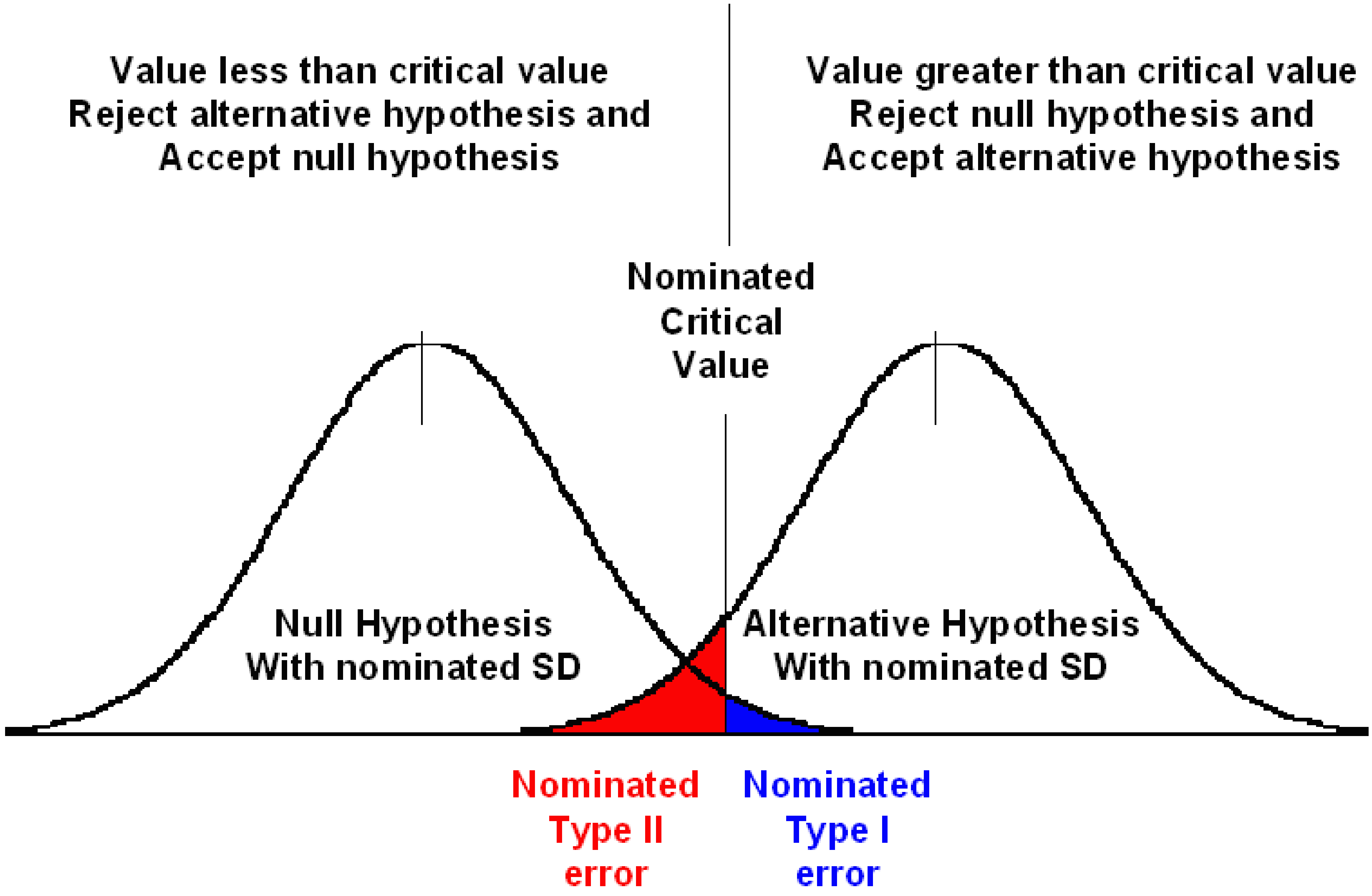
We return to the null and alternative hypotheses are:
Null hypothesis ($H_0$): $\mu_1 = \mu_2$ The two medications are equally effective.
Alternative hypothesis ($H_1$): $\mu_1 \neq \mu_2$ The two medications are not equally effective.

A type I error occurs if the researcher rejects the null hypothesis and concludes that the two medications are different when, in fact, they are not. If the medications have the same effectiveness, the researcher may not consider this error too severe because the patients still benefit from the same level of effectiveness regardless of which medicine they take. Type I error = false positive

A type II error occurs if the researcher fails to reject the null hypothesis when it should be rejected. That is, the researcher concludes that the medications are the same when, in fact, they are different. This error is potentially life-threatening if the less-effective medication is sold to the public instead of the more effective one.
Type II error = false negative

A **type I error** is:
the incorrect rejection of a true null hypothesis
(a "false positive")

A **type** II **error** is:
the incorrect retention of a false null hypothesis
(a "false negative").


Value less than critical value
Reject alternative hypothesis and
Accept null hypothesis
Value greater than critical value
Reject null hypothesis and
Accept alternative hypothesis
Nominated
Critical
Value
Null Hypothesis
With nominated SD
Alternative Hypothesis
With nominated SD
Nominated
Type II
error
Nominated
Type I
error

# Precision and accuracy

A measurement that has a small random error is "precise"

A measurement that has a small systematic error is "accurate"

These are not the same thing.