Concepts in Statistical inference

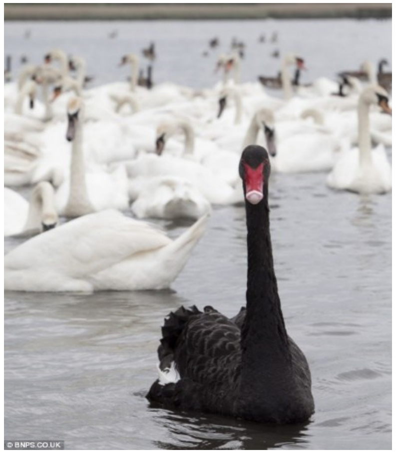
Topic	Further guidance
Principles of statistical inference	▪ Explain hypothesis testing and estimation ▪ Contrast Type I and II errors ▪ Interpret p-values and confidence intervals ▪ Define and identify the difference between statistical and clinical significance

Hypothesis testing

17th Century Europeans assumed that all swans were white.

The hypothesis that "All swans are white" was assumed to be true

Rejected by the sighting of a black swan by Willem de Vlamingh in 1697



The black swan resulted in a rejection of the original

null hypothesis (H_0): *"All swans are white"*

in favour of the

alternative hypothesis (H_1): *"All swans are not white"*.

Every day observation reinforces that which we believe to be true.

It does not ***prove*** what we believe is true.

A theory in the empirical sciences can never be proven,

But if assumed to be true - it can be subject to falsification.

One can think about falsification as the successive chipping away at a block of stone

the more we chip away the closer we get to the underlying form

Medical research

Clinical trials employ hypothesis tests

Determine if a novel treatment is effective in comparison to a control treatment.

Types of clinical trials

Superiority trial

Demonstrate that a new treatment is better than existing treatments.

Equivalence trial

Demonstrate that difference between control and experimental treatments is ***not large*** in either direction.

Non-inferiority trial

Demonstrate that an experimental treatment is ***not substantially worse*** than a control treatment.

Null Hypothesis

The **null hypothesis, H_0** is a statement of '*no difference*' or '*no effect*' which is *assumed* to be true.

Clinical trial of a new drug for hypertension:

null hypothesis: new drug has a similar average effect on blood pressure as another drug in current use

***H_0**: there is no difference in the effect on blood pressure between the two drugs*

Alternative Hypothesis

The **alternative hypothesis** (**H_1**) is the negation of the null hypothesis. It holds if the null hypothesis is not true.

***H_1*:** *the effects of the two anti-hypertensive drugs are not equal*

null hypothesis H_0: *there is no difference in the effect on blood pressure between the two drugs*

alternative hypothesis H_1: *the effects of the two anti-hypertensive drugs are not equal*

Null Hypothesis (examples)

There is no difference in wound infection rates between open prostatectomy and TURP.

The incidence of lung cancer in people who smoke is the same as among those who do not smoke

Measles occur in vaccinated and non-vaccinated children at the same rate.

H_0: *there is no difference in the effect on blood pressure between the two drugs*

H_1: *the effects of the two anti-hypertensive drugs are not equal*

We assume that the null hypothesis is true.

Find the probability that the observed data would be obtained under this assumption.

Which hypothesis do the data support ?

Logic of hypothesis testing is similar to a Court case:

"Innocent until proven guilty"

….. we gather evidence to come to a decision as to whether a person is innocent or guilty

In statistics we use the evidence from the sample to help decide whether to reject the null hypothesis

The evidence comes from the data

To test the null hypothesis.

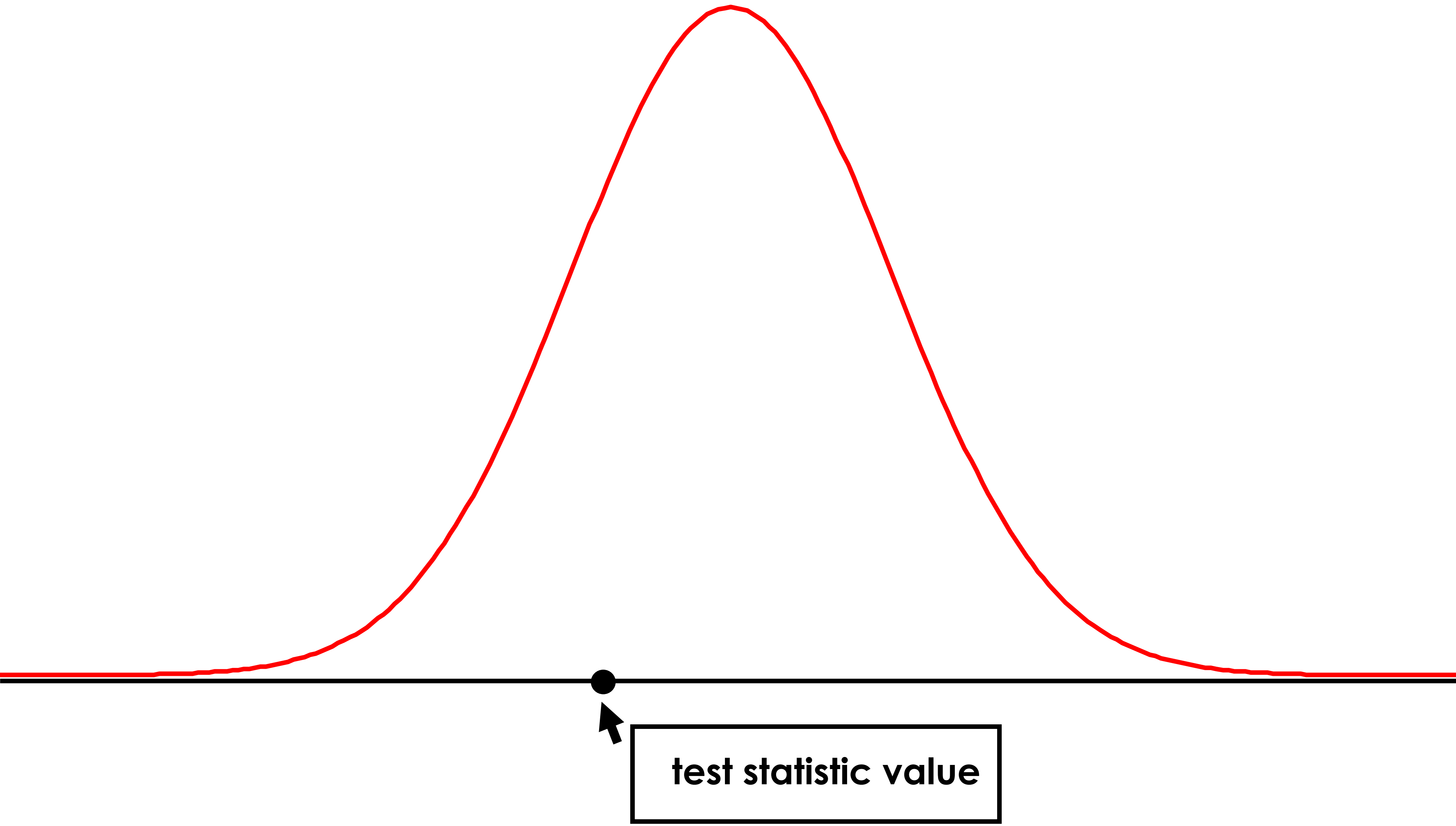
A numerical value is calculated from the sample data

It is called a **test statistic**

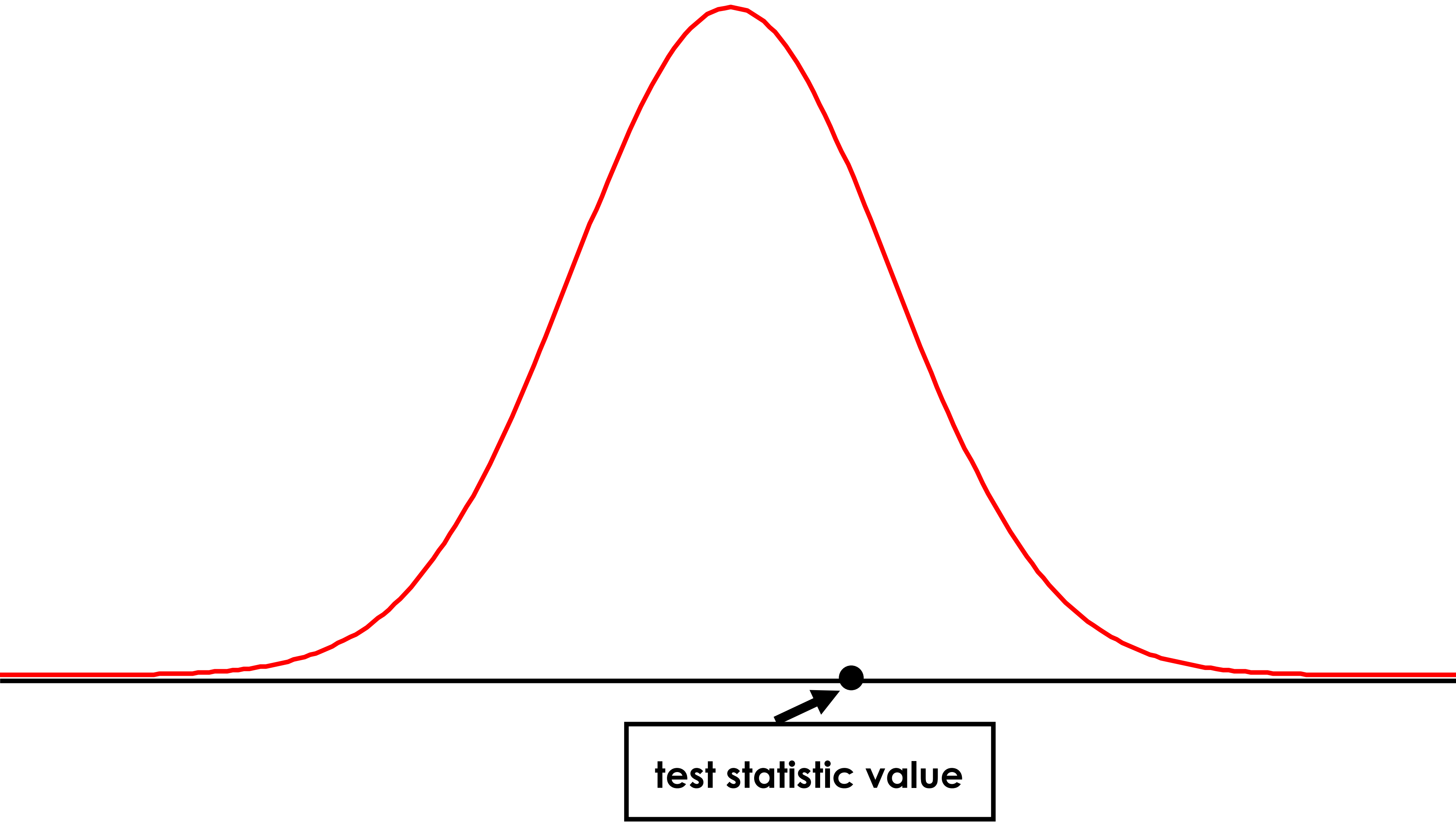
Using probability theory, it is possible to calculate the probability of obtaining a **test statistic value** as large or larger than the value observed from the data, when Ho is true.

The value of the test statistic is used to obtain a **p-value**

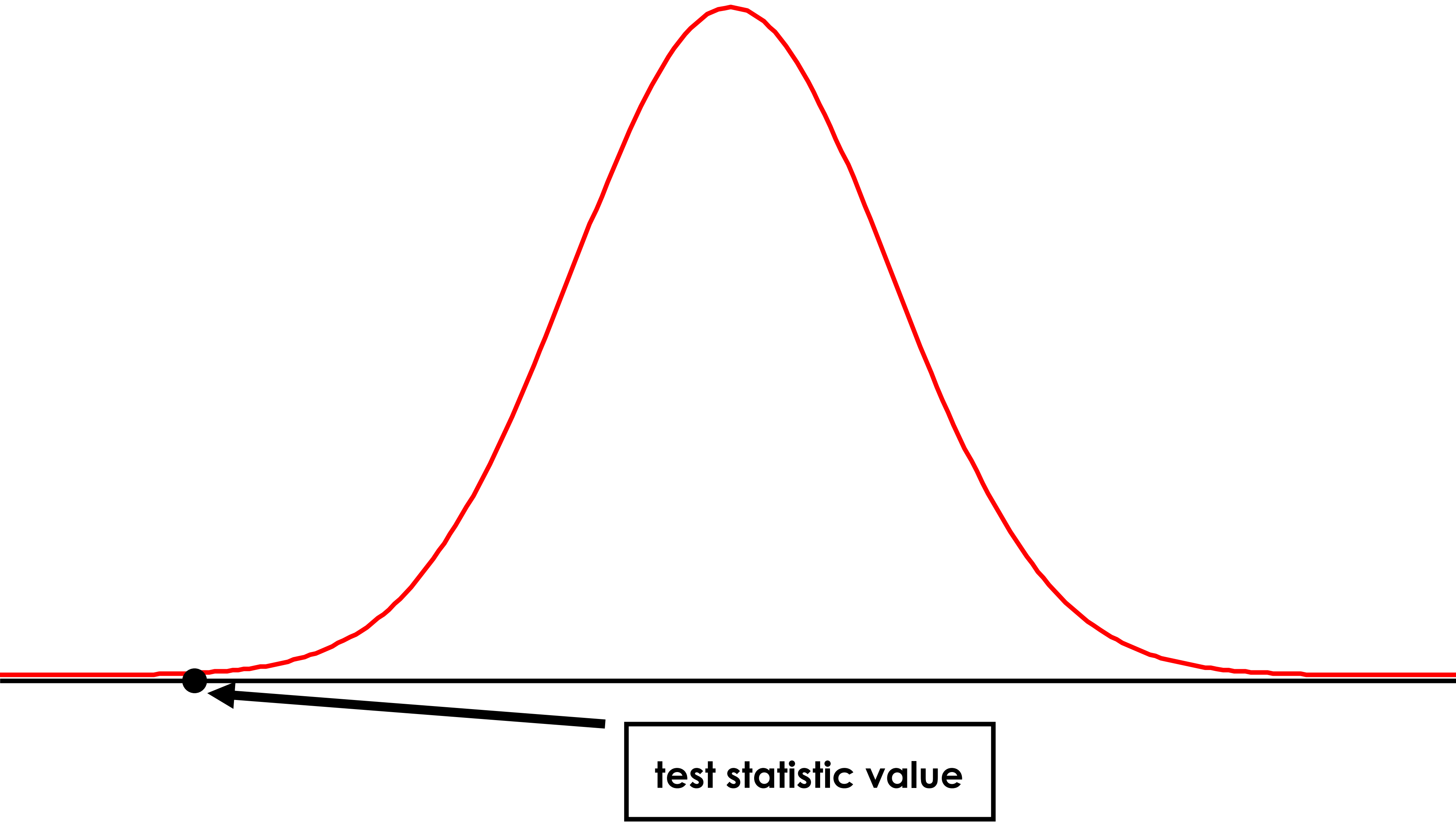
The value of the **test statistic** is related to a probability distribution



The value of the **test statistic** is related to a probability distribution



The value of the **test statistic** is related to a probability distribution

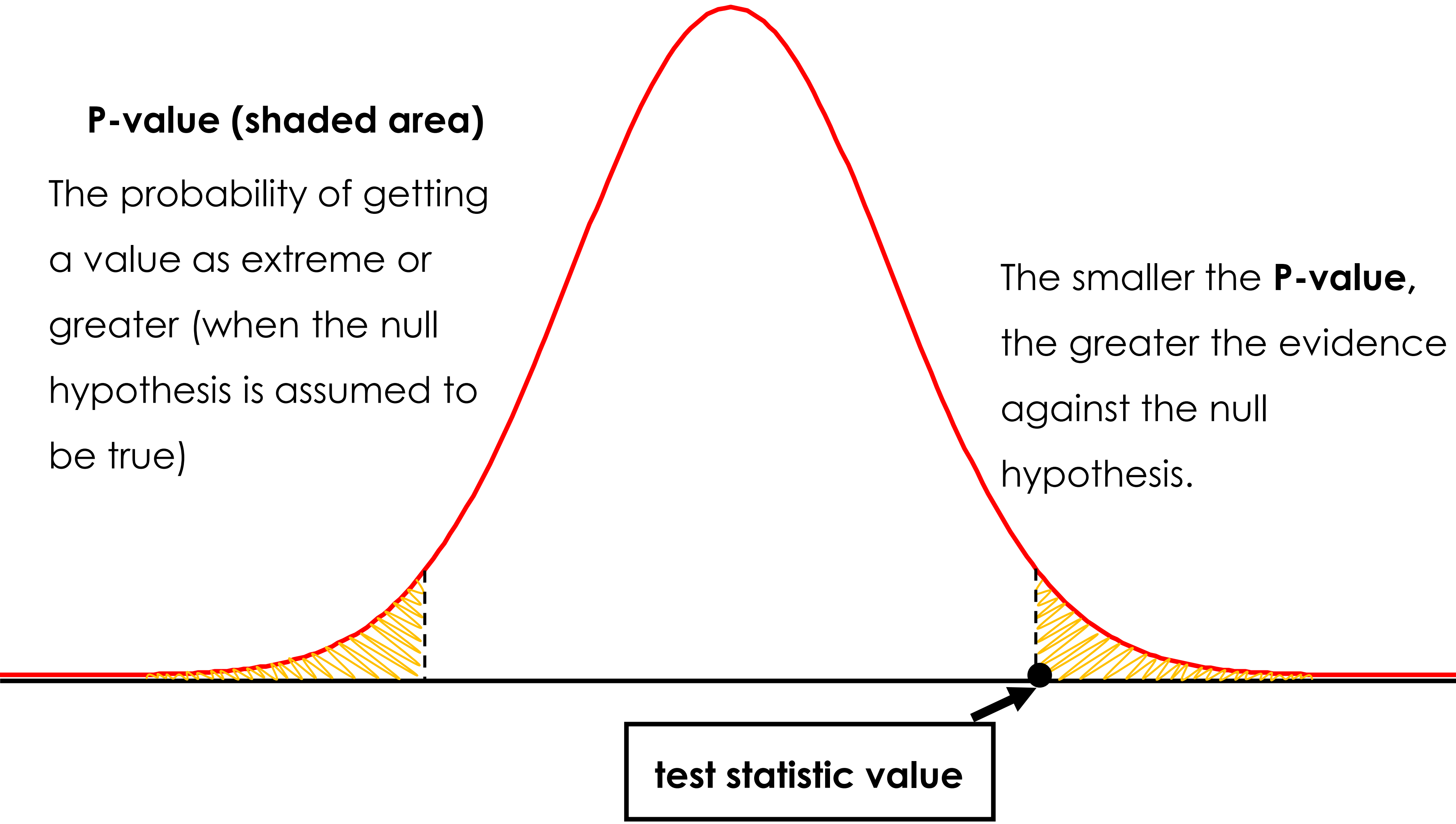


The value of the **test statistic** is related to the probability distribution

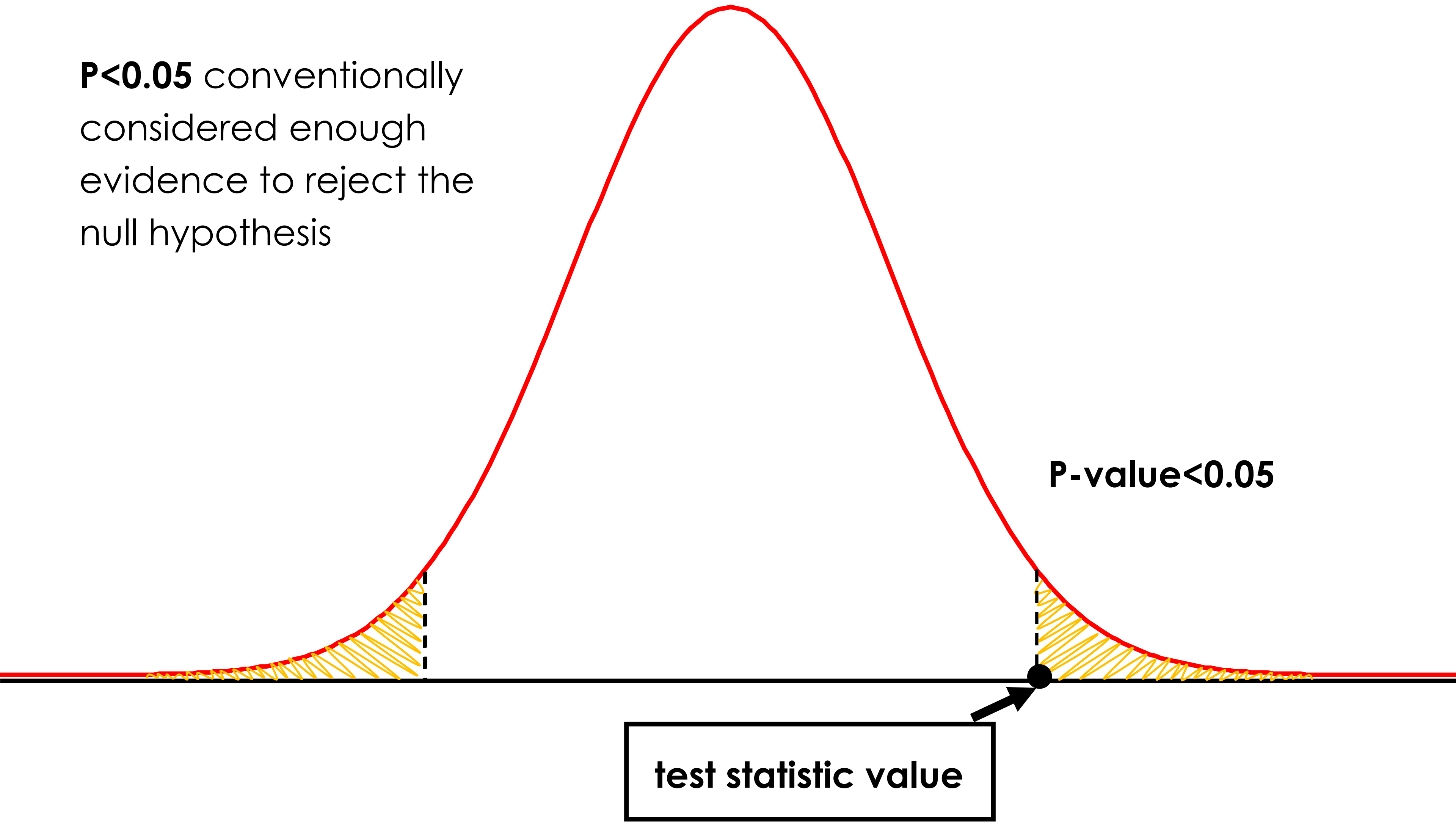
P-value (shaded area)

The probability of getting a value as extreme or greater (when the null hypothesis is assumed to be true)

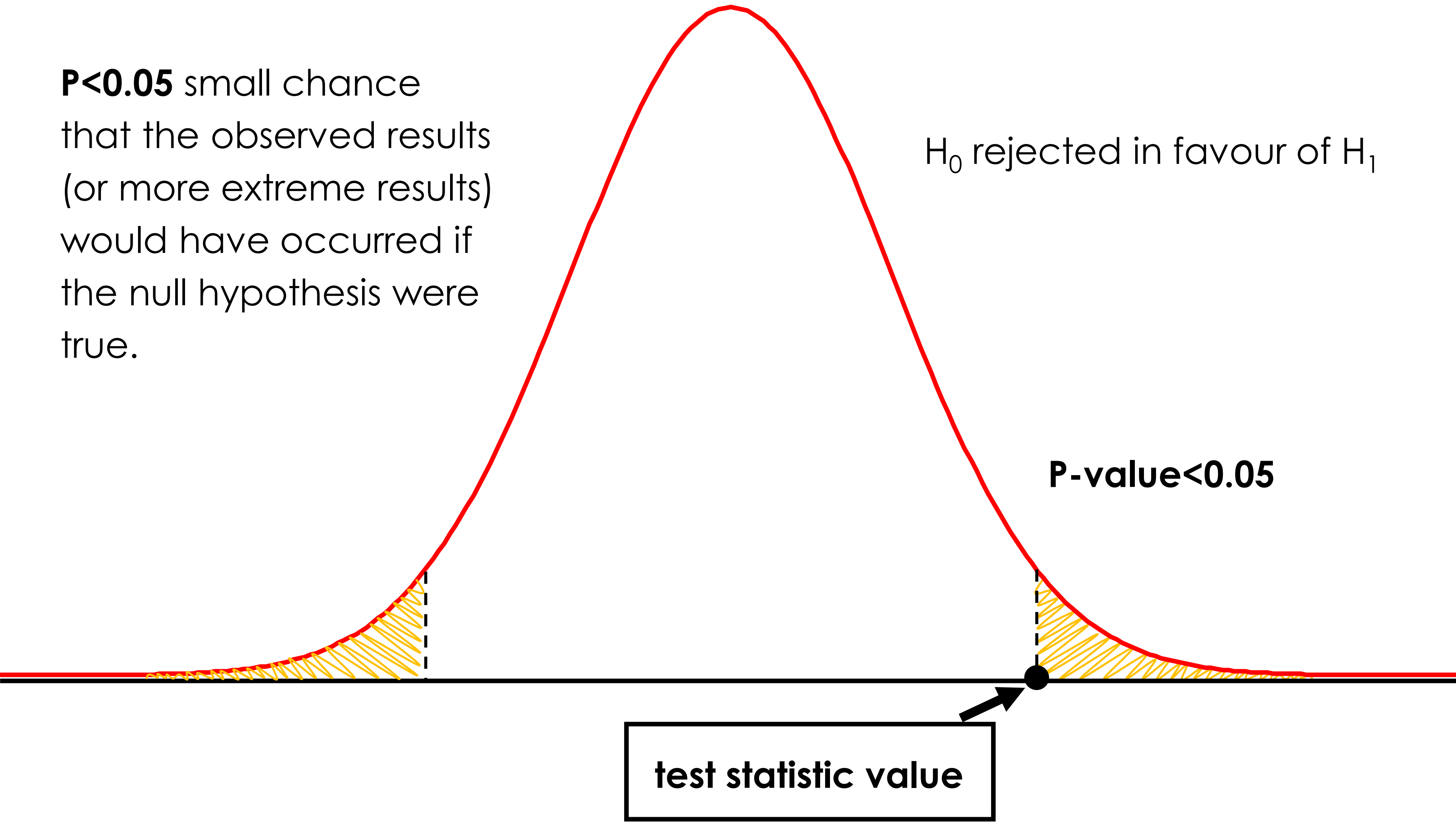
The smaller the **P-value**, the greater the evidence against the null hypothesis.



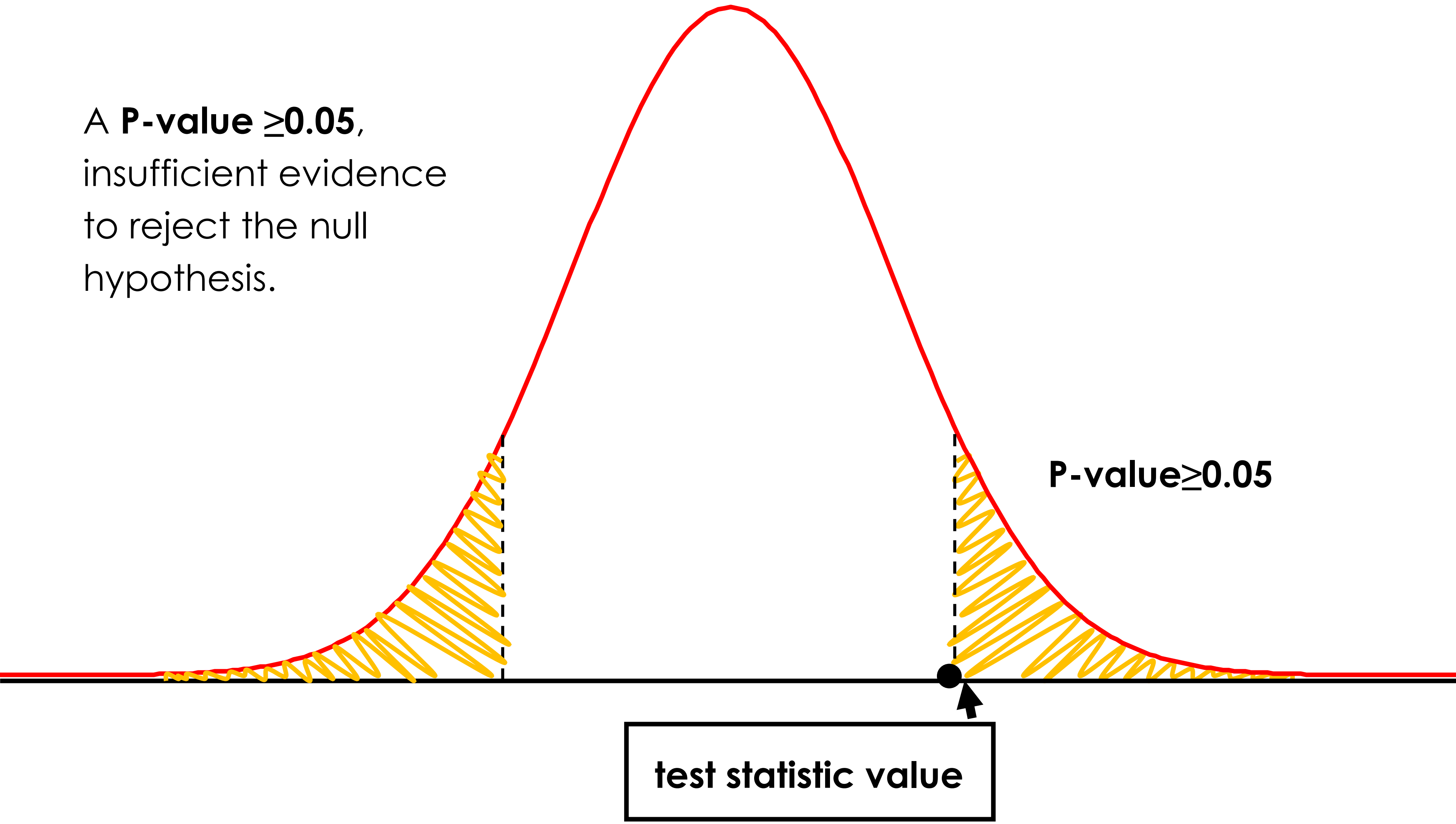
P<0.05 conventionally considered enough evidence to reject the null hypothesis



P<0.05 small chance that the observed results (or more extreme results) would have occurred if the null hypothesis were true.

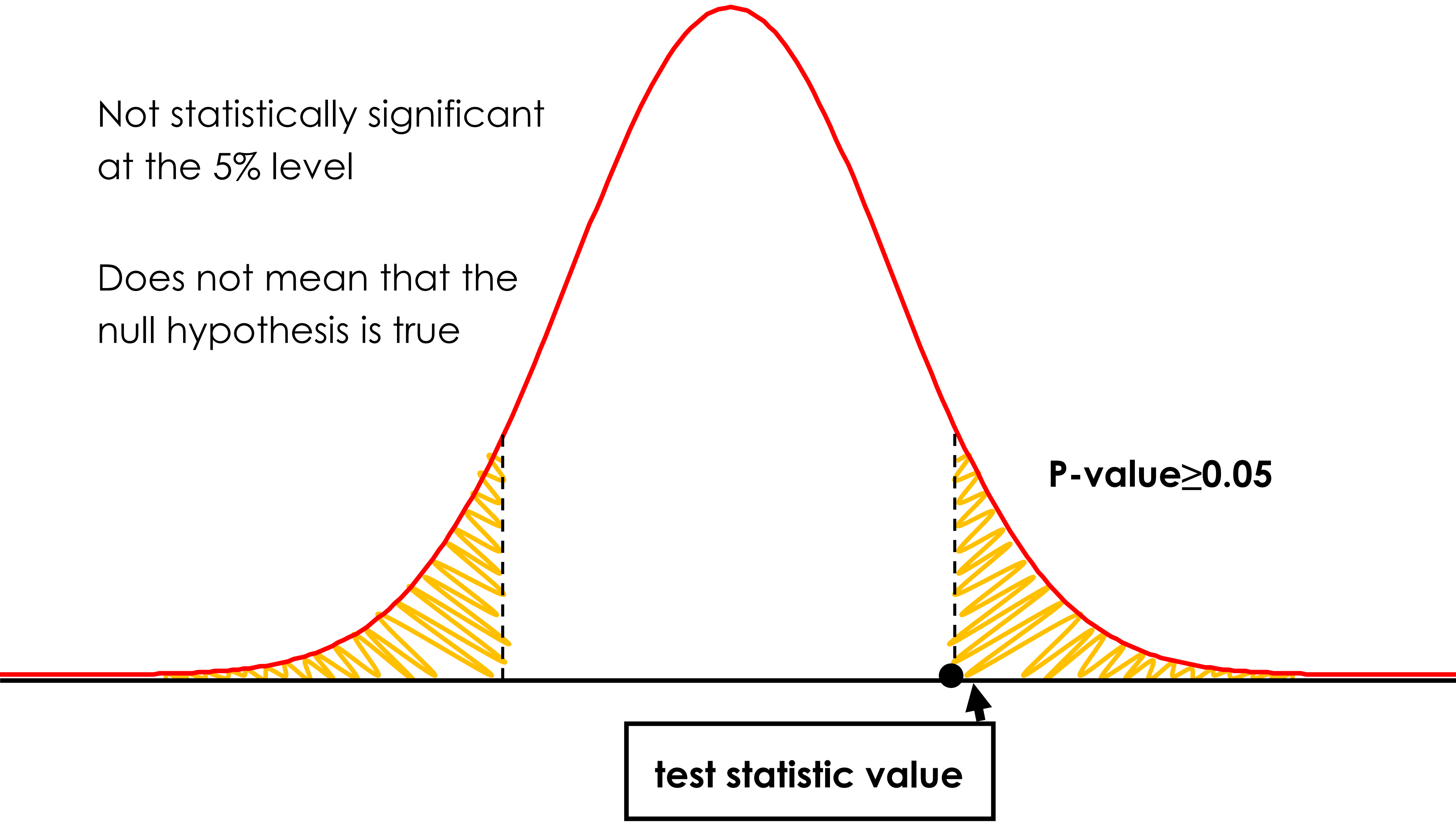


A **P-value ≥0.05**, insufficient evidence to reject the null hypothesis.



Not statistically significant at the 5% level

Does not mean that the null hypothesis is true

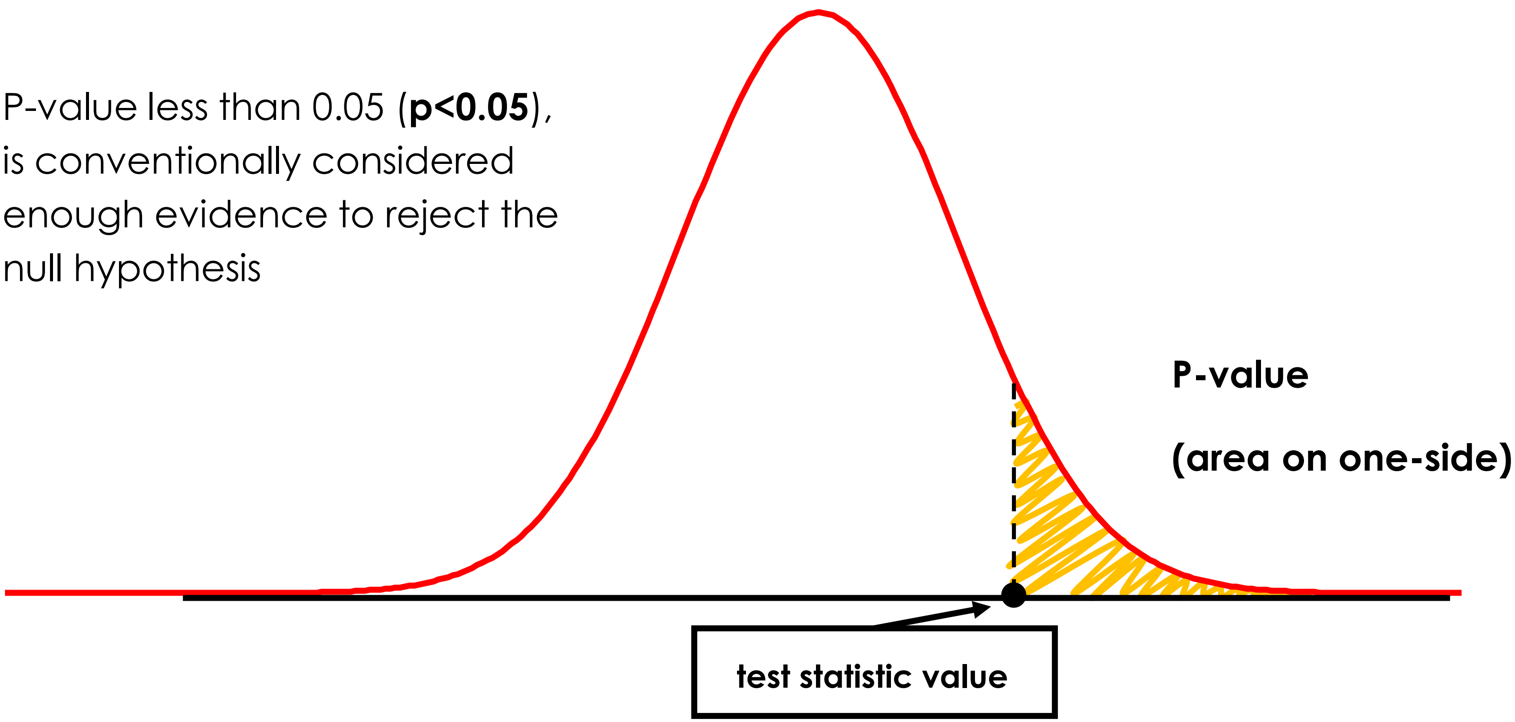


One sided test

***H_0**: there is no difference in the effect on blood pressure between the two drugs*

***H_1**: the effect of anti-hypertensive drug A* **>** *anti-hypertensive drug B*

P-value less than 0.05 (**p<0.05**), is conventionally considered enough evidence to reject the null hypothesis



Common misinterpretations of p-values

The p-value is ***not*** the probability that the null hypothesis is true.

(the null is either true or not)

Also, (1-p-value) is **not** the probability that the alternative hypothesis is true. (the alternative is either true or not true)

Correct interpretation

The p-value is the probability of getting a value of a test statistic as high or more extreme than the value of the statistic computed from the collected data, under the assumption that the null hypothesis is true

Making a decision

Type I & Type II errors, significance level and power.

Making a decision - Type I and type II errors

A **type I** error leads to the conclusion that an effect exists when in fact it does not.

(a "false positive")

A **type II** error is a failure to detect an effect that is present.
(a "false negative").

Type I is more serious than type II – *it could affect change*

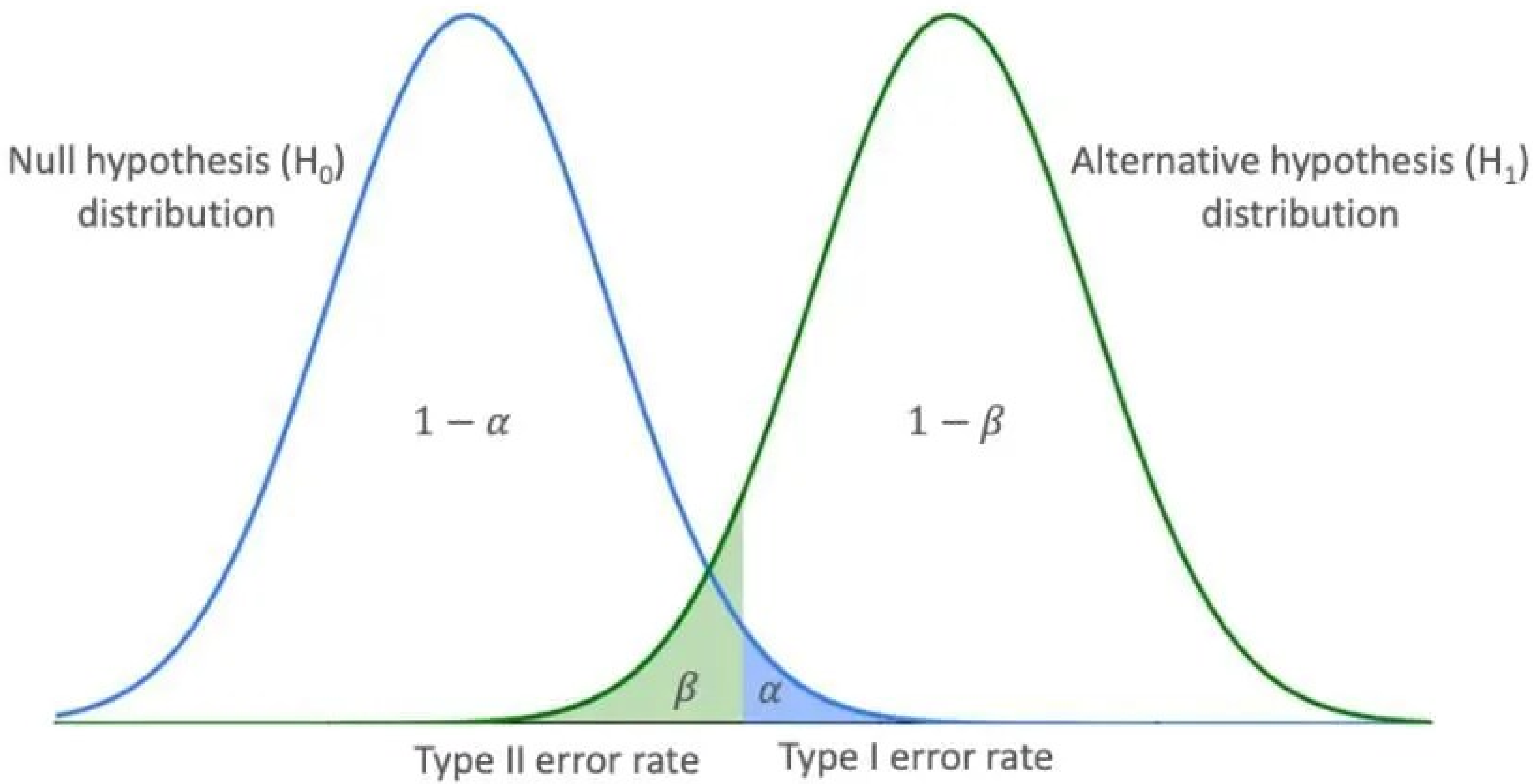
The **power** of the test is chance of detecting, as statistically significant, a real treatment effect.

Power is the probability of rejecting the null hypothesis when it is false;

α is the **significance level.**

It **is the probability of rejecting the null hypothesis when it is true**;

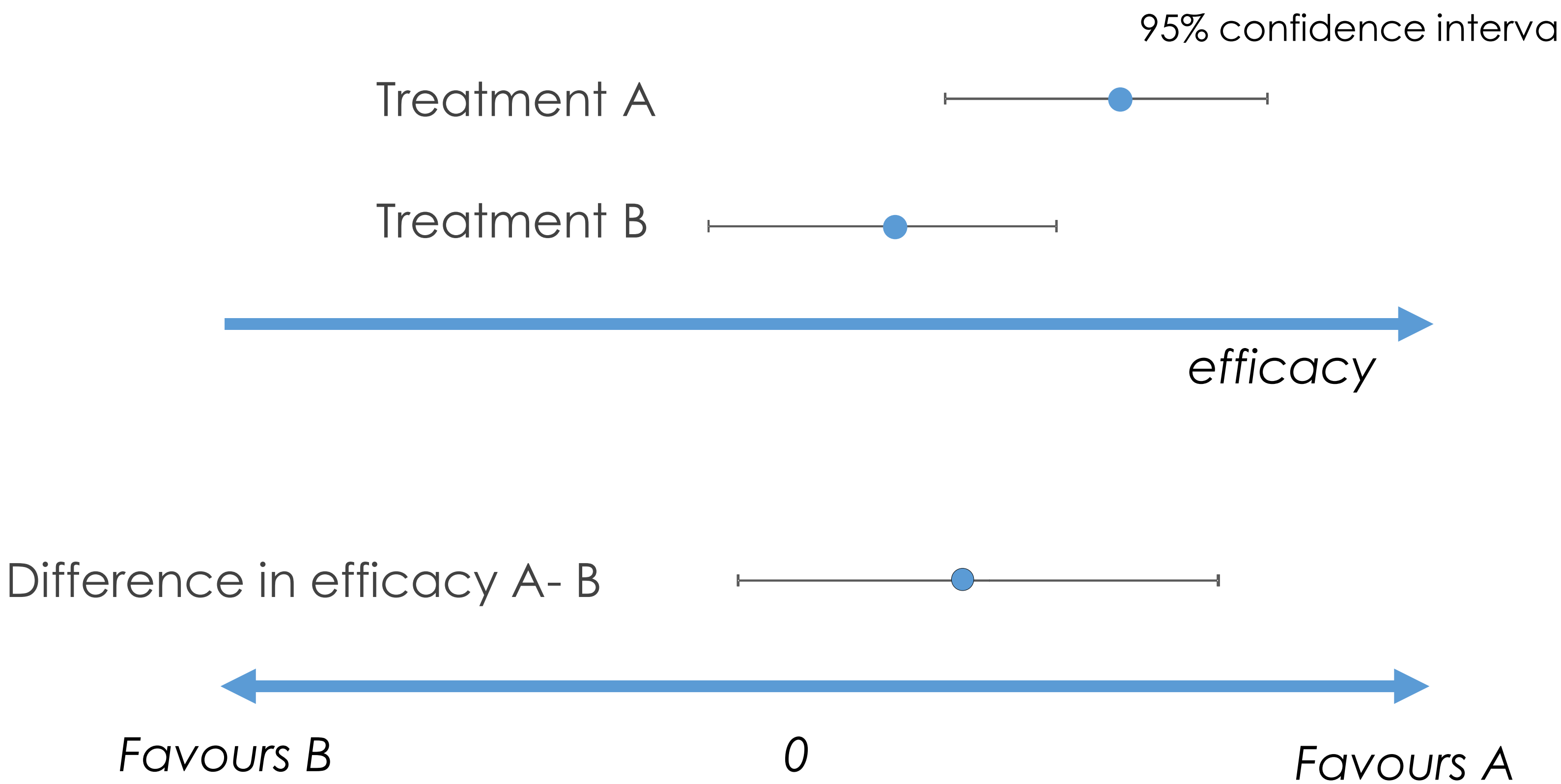
i.e. it is the chance (usually expressed as a percentage) of detecting, as statistically significant, a treatment effect when none exists.


Null hypothesis (H_0) distribution
Alternative hypothesis (H_1) distribution
$1 - \alpha$
$1 - \beta$
β
α
Type II error rate
Type I error rate

P-values or Confidence intervals?

Confidence intervals

Confidence intervals define the range of values within which a treatment effect is likely to lie



P-values or Confidence intervals?

A confidence interval is a range of values that the parameter of interest is likely to lie in the population

Provides information on the imprecision due to sampling variability

Advantages over just giving P values which dichotomies results into significant or non-significant.

P-values or Confidence intervals?

With a confidence interval, we can determine whether a parameter is or is not likely to be different to something

If the confidence interval contains a specific number - there is no evidence that the parameter is different from that number.

If the number is not within the interval, then there is evidence that the parameter is different from that number.

What does $p=0.001$ mean?

Highly statistically significant

(1 in a thousand chance of happening if null hypothesis is true)

But, doesn't necessarily imply practical relevance

(in health care, clinical, or laboratory settings)

It could be a very small effect,
estimated very precisely (i.e. Very narrow CI)

What does p=0.6 mean?

Not statistically significant

(6 in 10 chance of happening if null hypothesis true)

Does that mean nothing is happening?

It could be, but **not necessarily**

It could be a large effect,

but we've estimated it very **imprecisely** (i.e. very wide CI)

Statistical significance and clinical significance

Statistical significance can always be achieved with very large sample sizes. The larger the sample size the smaller the minimum effect that will be detected.

Clinical significance is the practical importance of a treatment effect - whether it has a real genuine, palpable, noticeable effect on daily life.

2. A p-value for a test-statistic from a particular statistical test

A. gives the probability that the null hypothesis is true
B. gives the probability of obtaining a test-statistic value as large or larger when the null hypothesis is assumed to be true
C. gives the probability that the alternative hypothesis is true
D. gives the probability of obtaining a test-statistic value as large or larger when the null hypothesis is assumed to be false
E. gives the probability of obtain a test-statistic value as large or larger when the alternative hypothesis is assumed to be true

The power of a statistical test

A. is the probability that the null hypothesis is true
B. is the probability of not rejecting the null hypothesis when it is false
C. is the probability of rejecting the null hypothesis when it is true
D. is the probability of rejecting the null hypothesis when it is false
E. is the probability of detecting a clinically significant difference