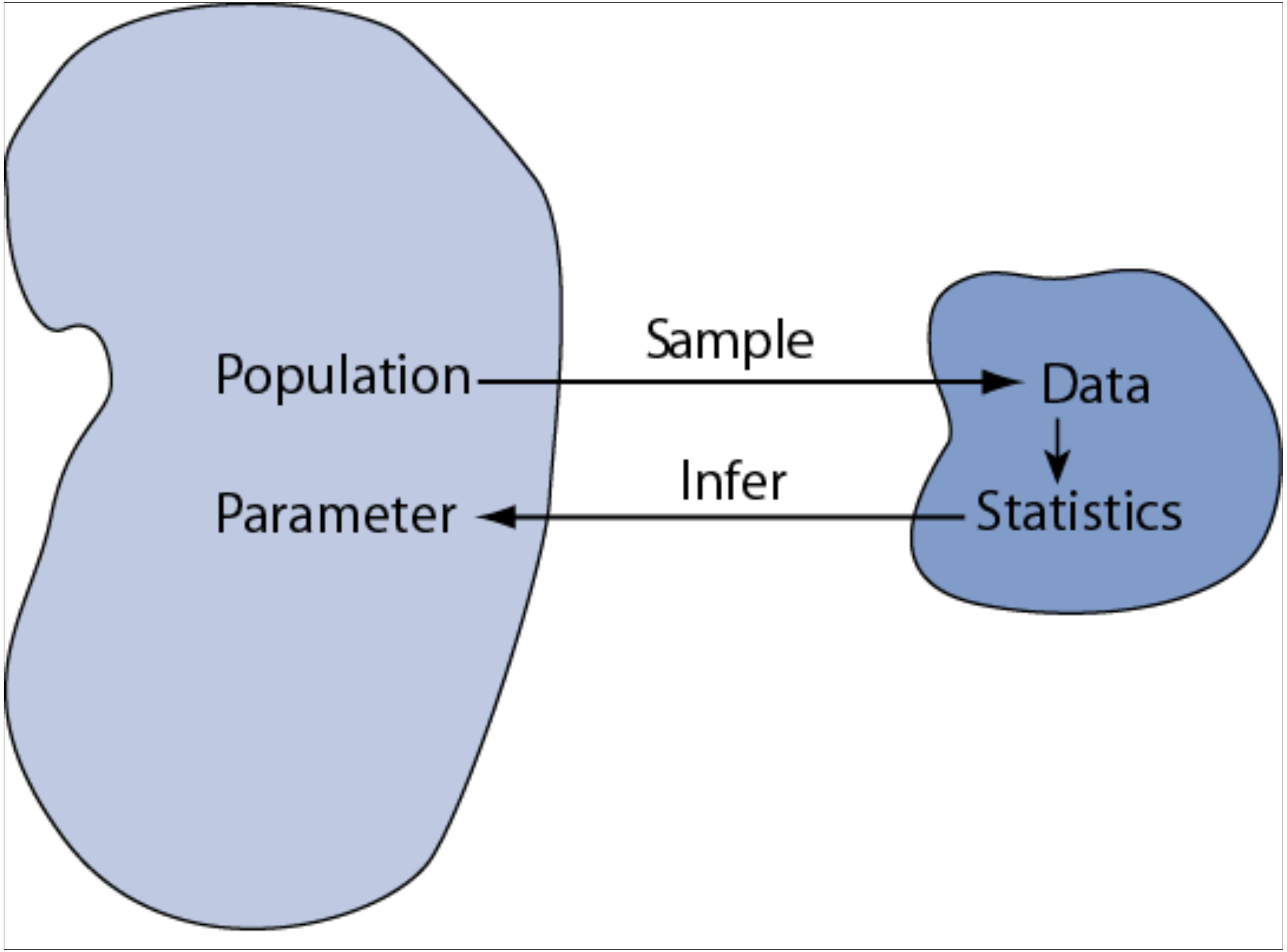
HYPOTHESIS TEST

CONTENTS


Population
Sample
Data
Infer
Parameter
Statistics

HYPOTHESIS TEST?

Hemoglobin (g/dl)		
Population A (all the male students in your university) (μ_a=170.00, σ=43.10)	**Sample a1** ($\overline{X}_{a1}$=171.21, n=10)	**Sample a2** ($\overline{X}_{a2}$=160.50, n=10)
Population B (all the female students in your university) (μ_b=160.00, σ=45.20)	**Sample b1** ($\overline{X}_{b1}$=160.30, n=10)	

Because we already know μ_a and μ_b

a1-a2 → Sampling error

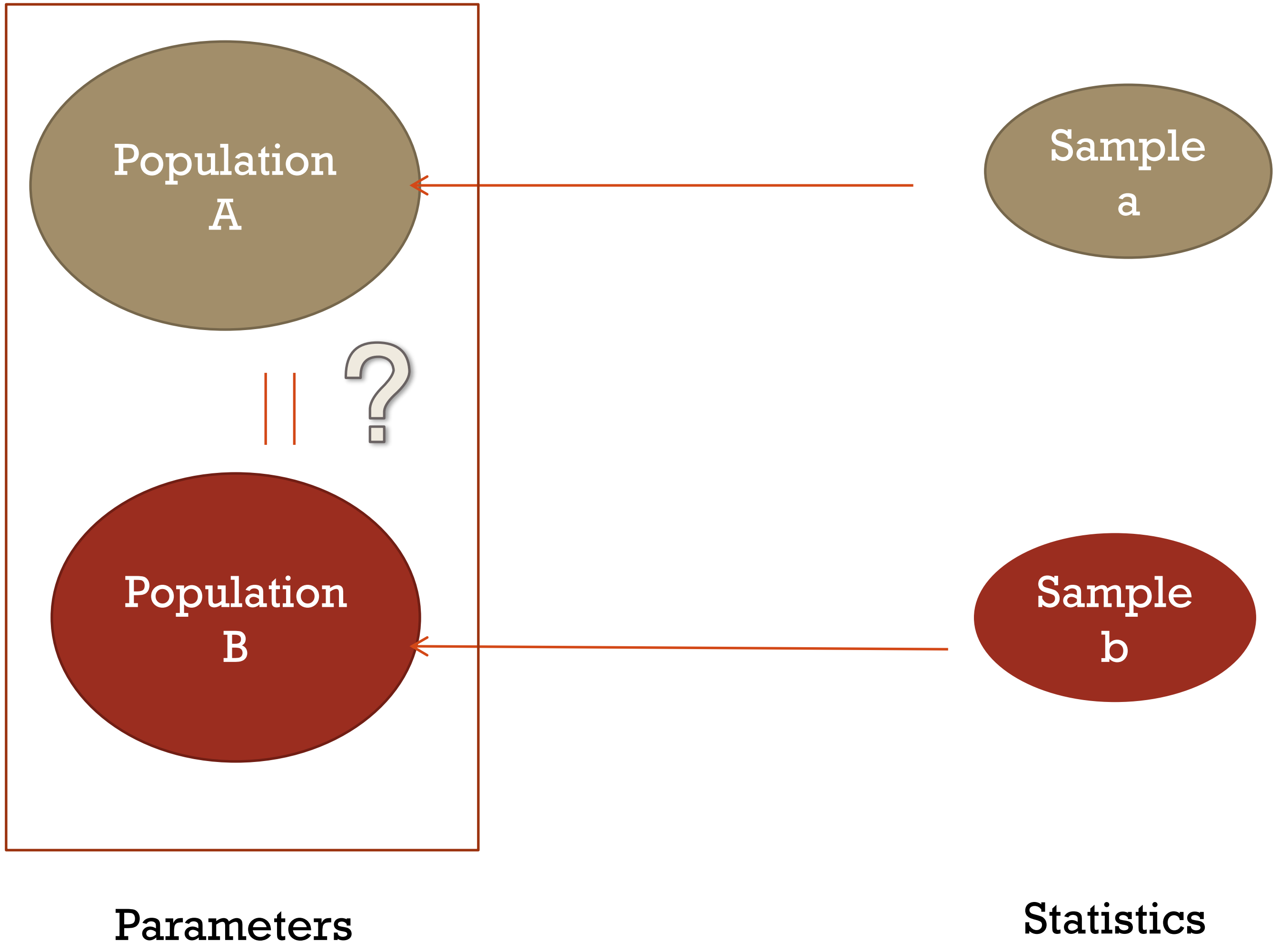
a1-b1 → True difference of their original μ

Example

128.6

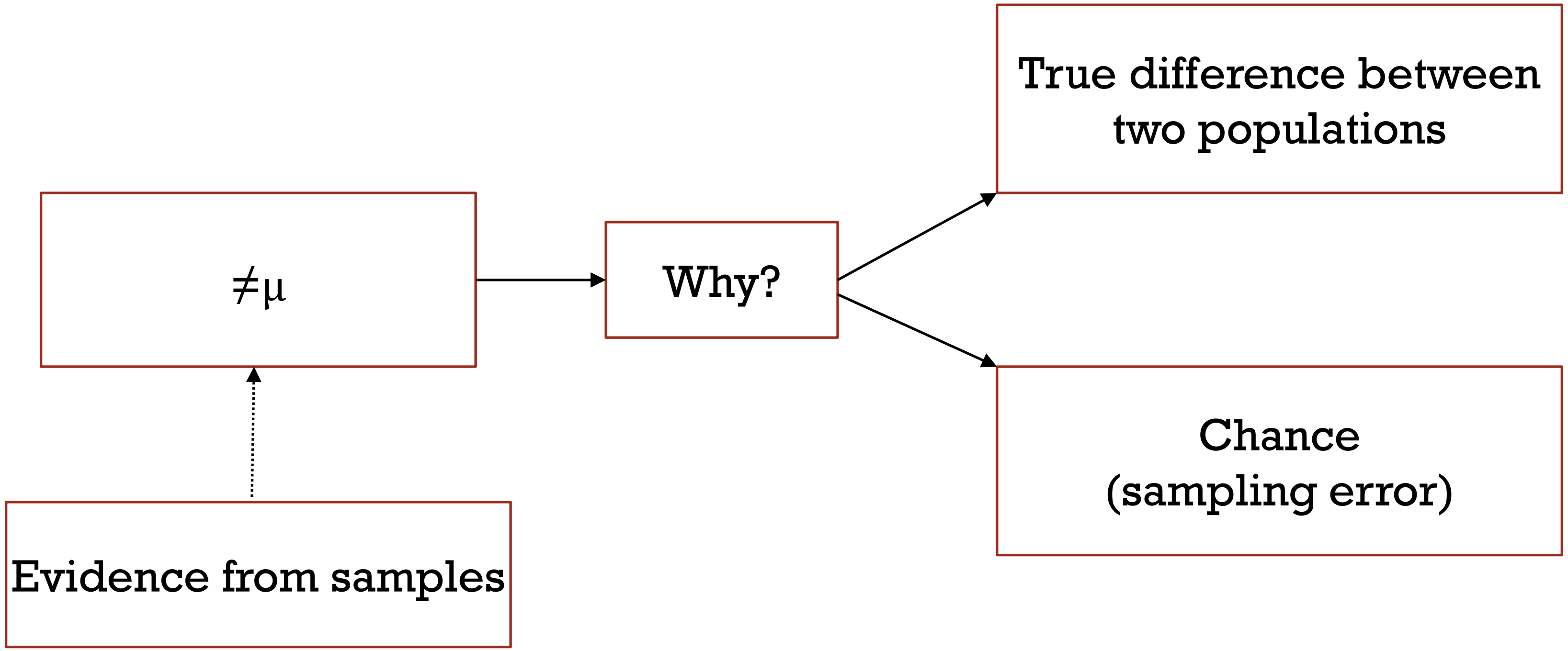
✓Suppose that we are interested in investigating the effect of race on level of blood pressure. One research group enrolled 100 white males and 100 black males. The means of systolic blood pressure among white males and black males were reported as 128.6 mmHg and 135 mmHg, respectively. Can we conclude that SBP are the same for whites and blacks?

135


Population
A
Population
B
Sample
a
Sample
b
Parameters
Statistics

1. LOGIC OF HYPOTHESIS TEST

HYPOTHESIS TEST TASK



So the hypothesis task is to differentiate that the difference between the samples is from the true difference between two populations or from the chance!

HYPOTHESIS TEST

Hypothesis test, is also called significance testing.

It tests a claim about a parameter using evidence/data from samples(s).

At the heart of hypothesis testing is "proof by contradiction" and "small probability event".

Small probability event

- Also call rare event
- If the probability of some special event is very low (close to 0), then it is a rare event.
- In one experiment, the event with very small probabilities will not occur.

- **PROOF BY CONTRADICTION**

- Proposition: Females don’ t smoke.

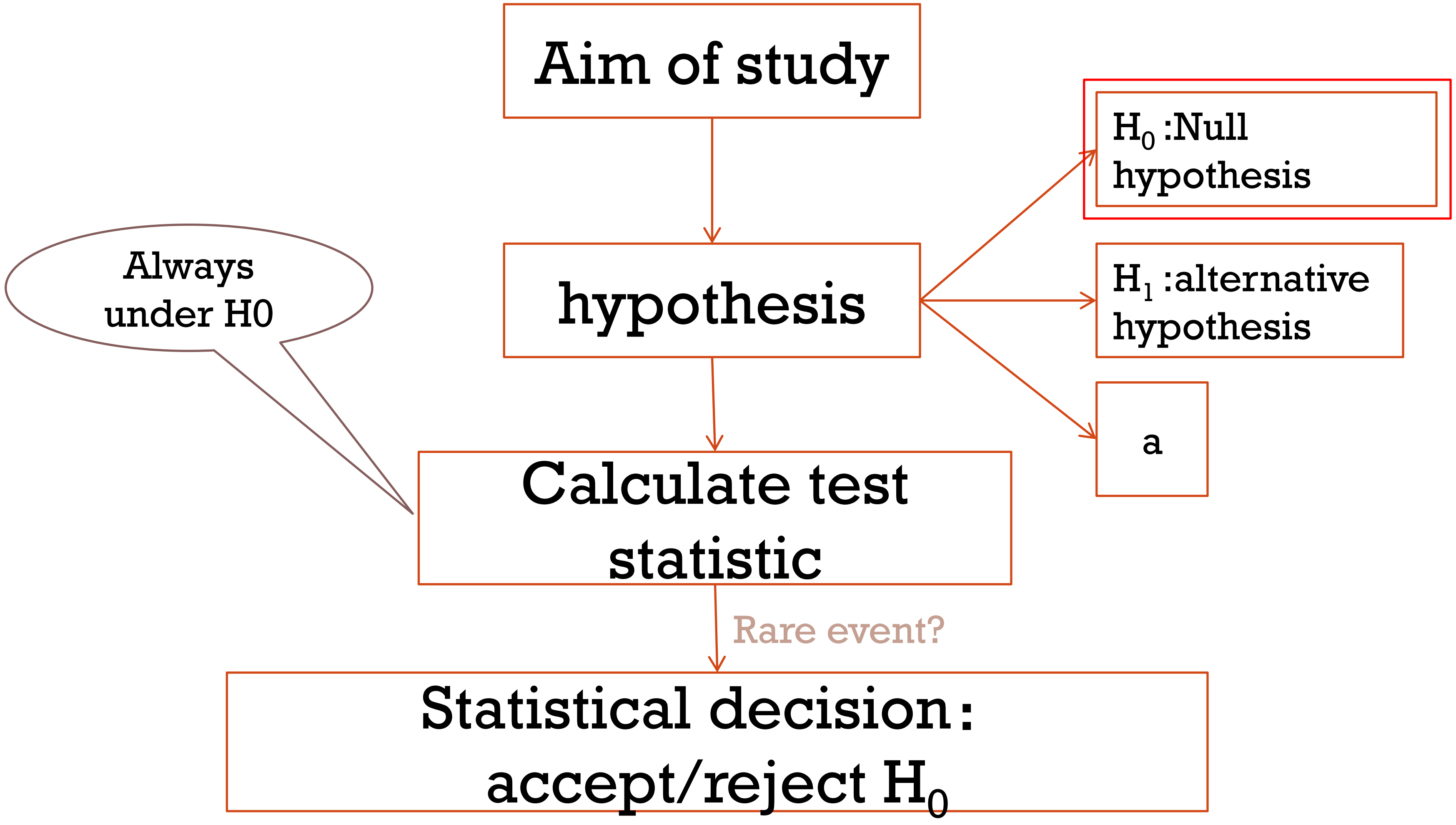
  All the swans are white.

A conclusion drawn on the basis of an inductive method can never be fully proven. Proof by contradiction is relatively easy to form the negation of the proposition.

- Analogies: Trial by jury

In criminal court, the accused is “presumed innocent” until “proved guilty beyond all reasonable doubt.” In other words, an innocent defendant is sometimes convicted by overwhelming circumstantial evidence.

LOGIC OF HYPOTHESIS TEST



NULL HYPOTHESIS, H_0

- A null hypothesis is often concerned with a parameter or parameters of population(s).
- The hypothesis to be test.
- It is always stated in the null form, indicating no difference or no relationship between distributions or parameters.

- If there is no difference between parameters, or the data follows a specific distribution, then there will be easy to adopt some established rules/laws/distributions.
- Researchers often want to reject H_0, and reveal there are true difference.

ALTERNATIVE HYPOTHESIS, H_1/H_A

- A hypothesis that in some sense contradicts the null hypothesis H_0.
- Claims H0 is false
- $P(H_0)+P(H_1)=1$

● Compare among three population means (ANOVA)

H_0 : $\mu 1=\mu 2=\mu 3$

H_1: $\mu 1\neq\mu 2\neq\mu 3$; $\mu 1=\mu 2\neq\mu 3$; $\mu 1\neq\mu 2=\mu 3$

LEVEL OF SIGNIFICANCE

- Your definition: Probability $< \alpha$ is a rare event ;
- Determines the reject region.

Common choice: α=0.05

2. STEPS OF HYPOTHESIS TEST

Example

In previous large-scale survey, the birth length of male newborns was normally distributed, with an average of 49 cm. A gynecologist in a hospital randomly selected 25 male babies born between 2010 and 2013, and found that the average birth length of the 25 babies was 50 cm, and the standard deviation was 5 cm. The gynecologist wants to know whether the birth length of male babies born between 2010 and 2013 is different from 49cm?

$\mu = 49$

$n = 25$

$\bar{X} = 50$

$SD = S = 5$

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

- **aim:**

μ_0=49cm

All male newborns

? $\neq$

μ=?

Born between 2010 and 2013

n=25

$\bar{x}$ =50,
S=5

Classis 3 Steps

Step 1. State hypothesis and α

H_0: $\mu = \mu_0$ — The μ of birth length of babies born between 2010 and 2013 is identical to 49cm

H_1: $\mu \neq \mu_0$ — The μ of birth length of babies born between 2010 and 2013 is not identical to 49cm

$$\alpha = 0.05$$

1-sided and 2-sided hypothesis

two-sided/nondirectional test :

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

one-sided/directional test :

$$① \begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu > \mu_0 \end{cases} \quad or \quad ② \begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu < \mu_0 \end{cases}$$

Is there a difference between the cholesterol levels of men and women?

Does the mean age of a target population differ from that of the general population?

Is a new drug superior to a standard drug?

Does the air pollution exceed safe limits?

Has the death rate been reduced for those who quit smoking?

Step 2. Choose method and calculate test statistics

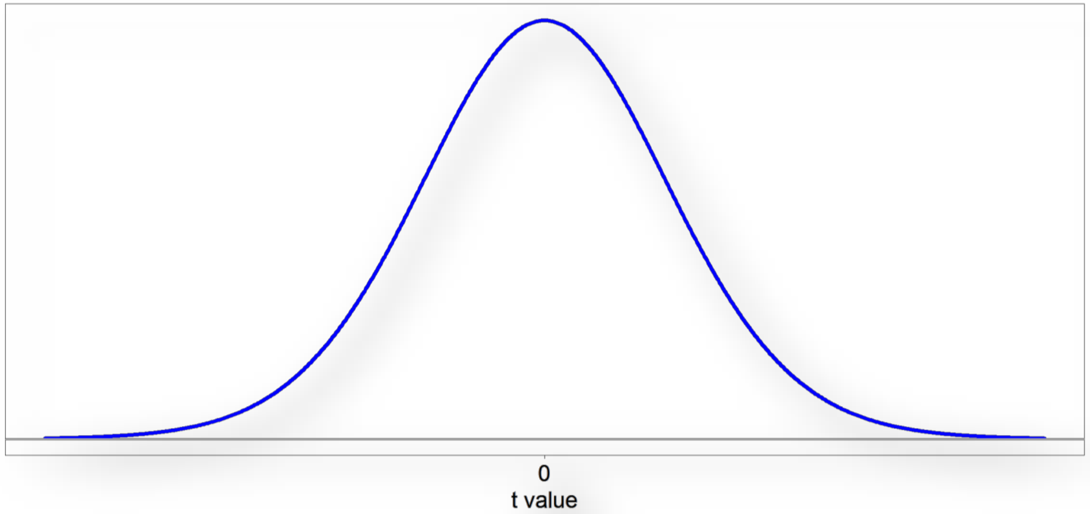
Choose method: depends on data type, study design, study aims and conditions. The most difficult step!

The sampling distributions of a mean follows t distribution (σ is not known)

Under H0

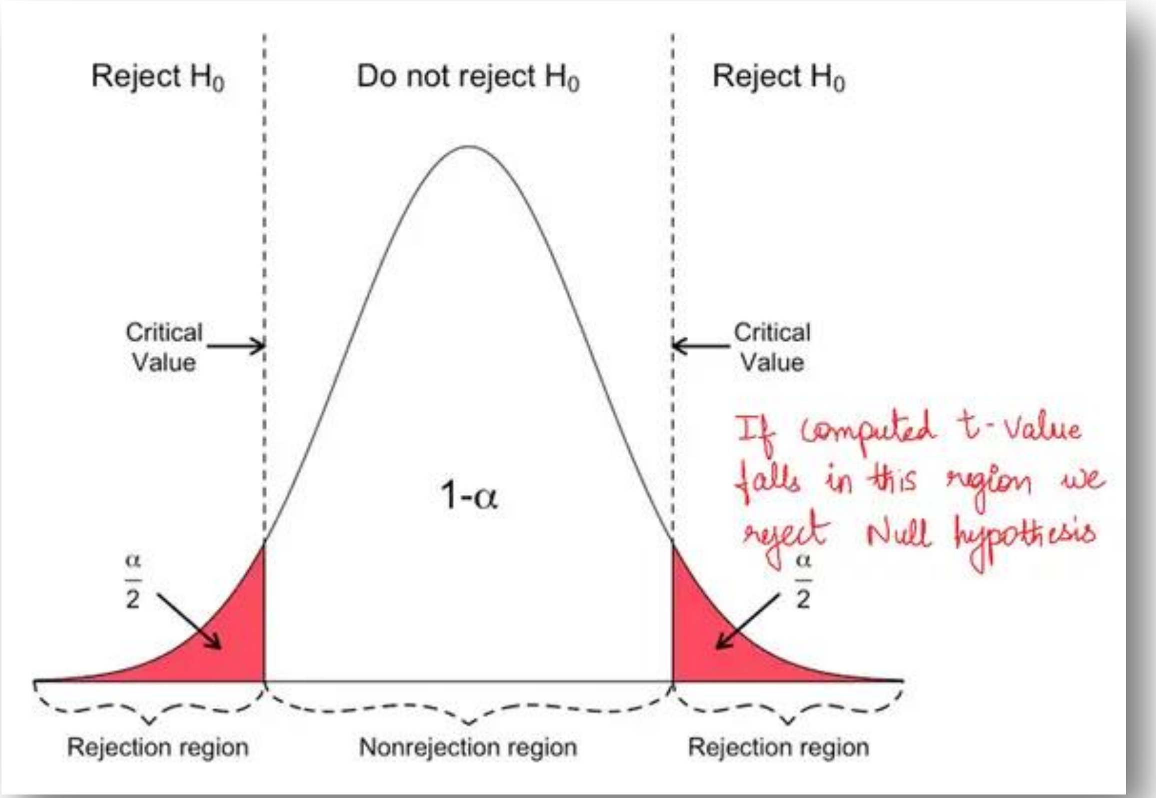
$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

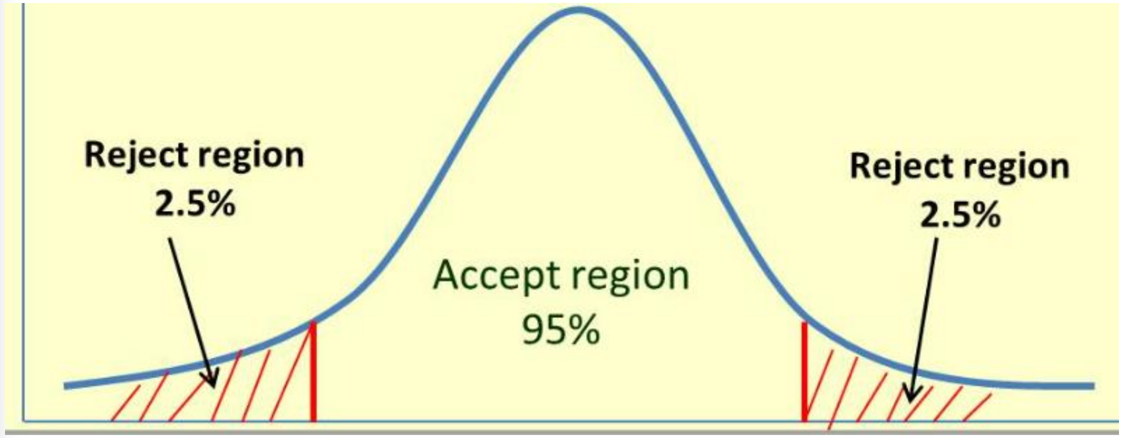
$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{50 - 49}{5/\sqrt{25}} = 1.000$$



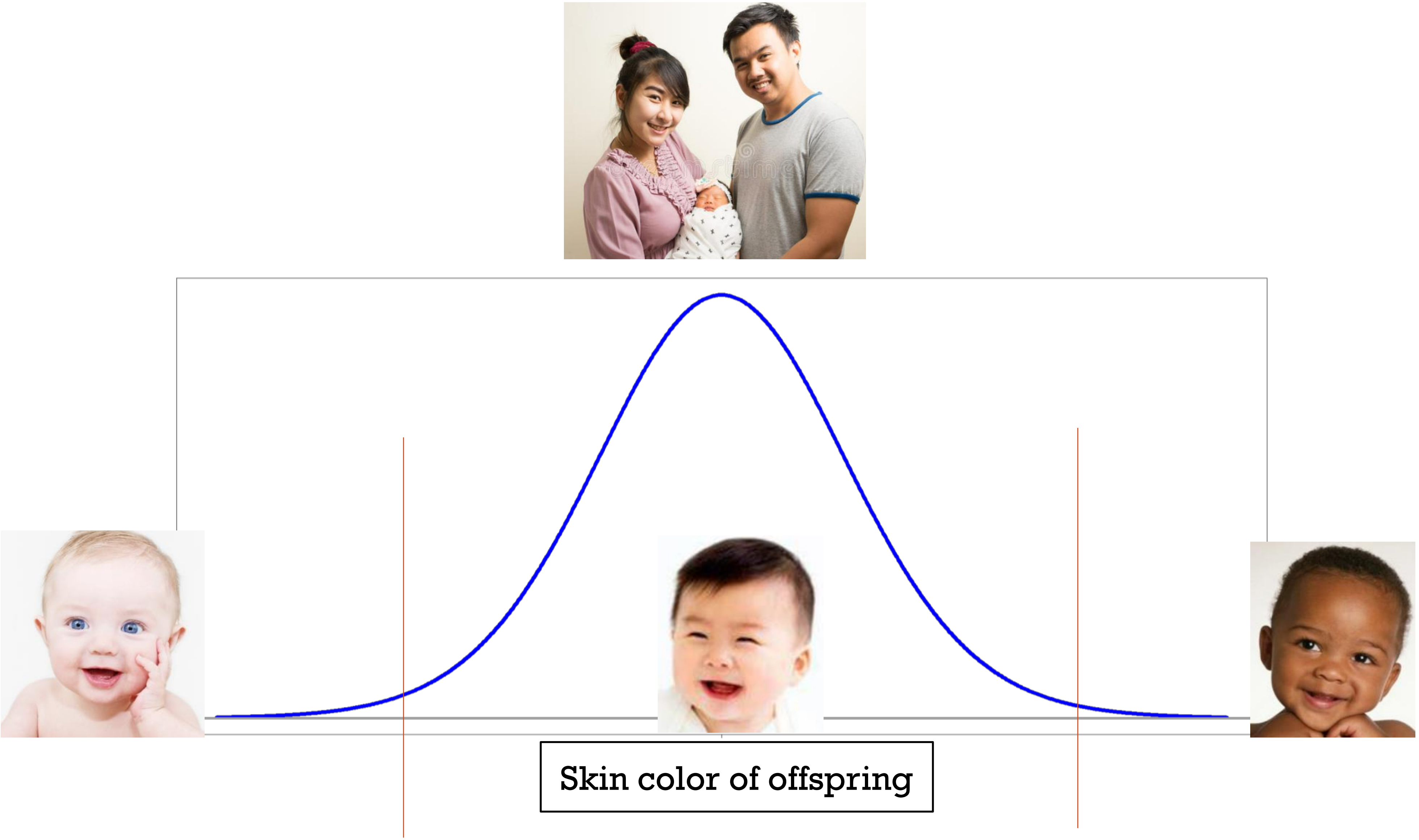
Step 3. Make a statistical decision

Our example is a two-sided test





Analogy of reject region

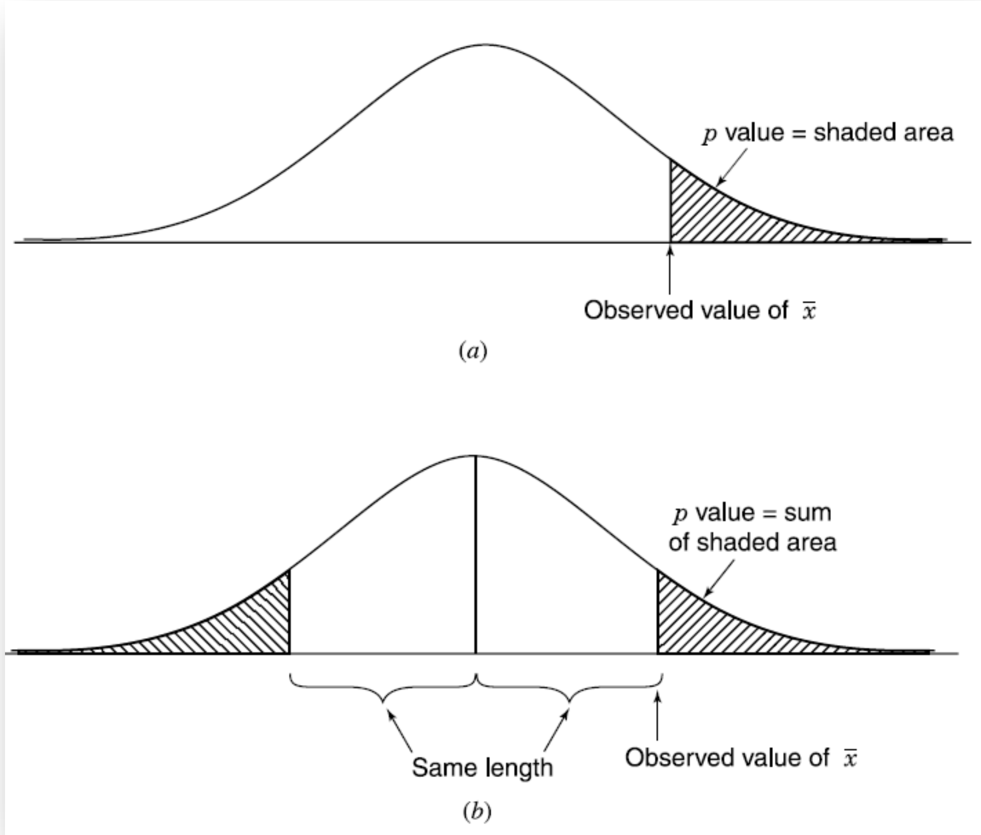


P value

The P value is the probability of getting values of the test statistic as extreme as, or more extreme than, that observed if the null hypothesis is true.

One-sided

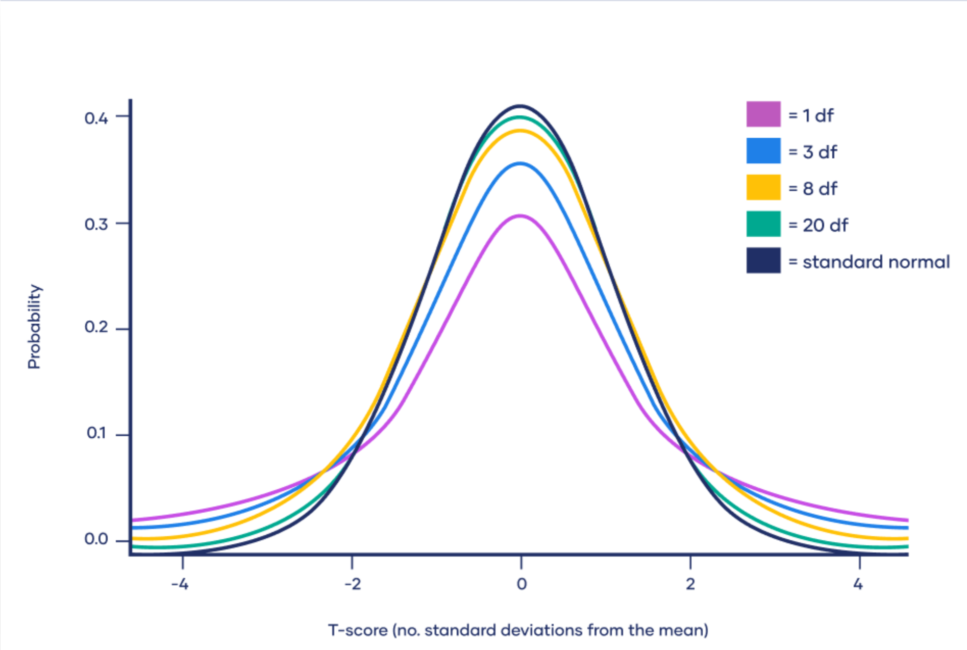
Two-sided

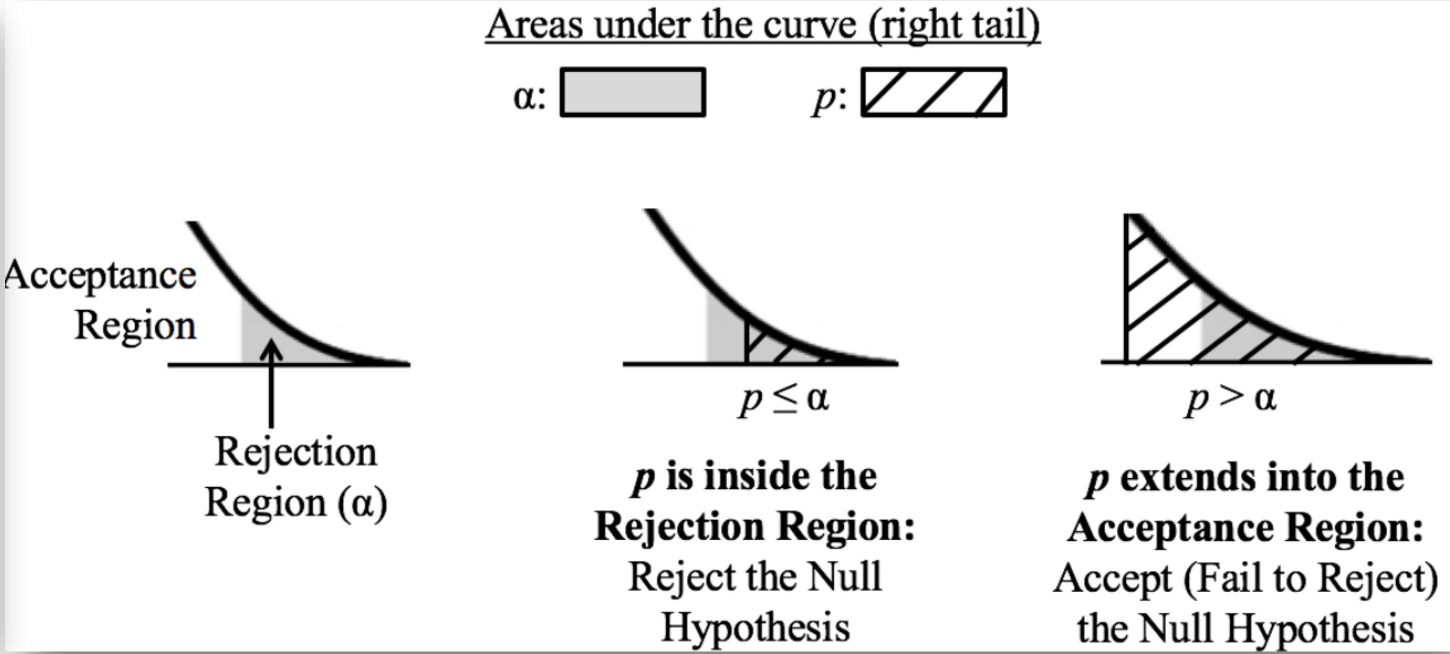


P value

Why we need to calculate a *P* value

More informative than a test statistic





Can't tell how far is the statistic from the critical value

Reporting the p value associated with a test lets the reader know how common or how rare is the computed value of the test statistic given that H0 is true.

P value and a

***P* value:**

calculated based on your test statistic

H_0 is taking an exam/test!
The score is present as P value!

$P > \alpha$,

H_0 passes the test,

it is NOT rejected.

α

Set in advance

$P \leq \alpha$,

H_0 dose not pass the test,

it is rejected.

p Value	Interpretation
$p > 0.10$	Result is not significant
$0.05 < p < 0.10$	Result is marginally significant
$0.01 < p < 0.05$	Result is significant
$p < 0.01$	Result is highly significant

Summary: 3 steps

$$H_0: \quad \mu = \mu_0$$
$$H_1: \quad \mu \neq \mu_0$$
$$\alpha = 0.05$$

(1)State hypothesis and α

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{50 - 49}{5/\sqrt{25}} = 1.000$$

(2) Choose statistic method,
Calculate test statistic

$$t < t_{0.05/2,24} = 2.064, \quad P > 0.05$$

Don't reject H0

$\nu = n - 1 = 24$

(3) Make statistical decision

3. ERRORS

Type I and Type II error

Four possible outcomes

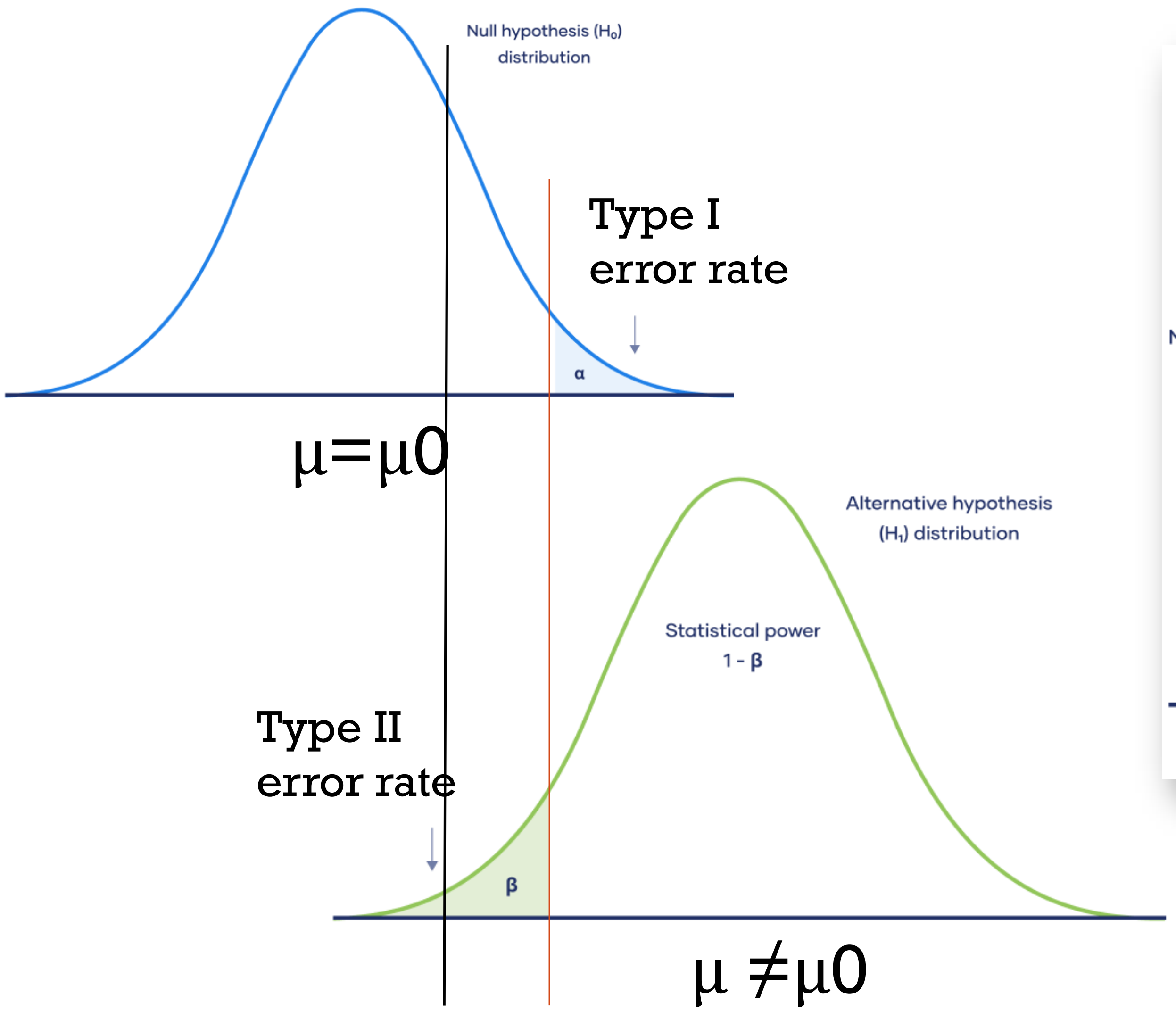
	Reject H_0, accept H_1	Not reject H_0
True H_0	Type I error (α)	**Correct decision ($1-\alpha$)**
False H_0	**Correct decision (statistic power, $1-\beta$)**	Type II error (β)

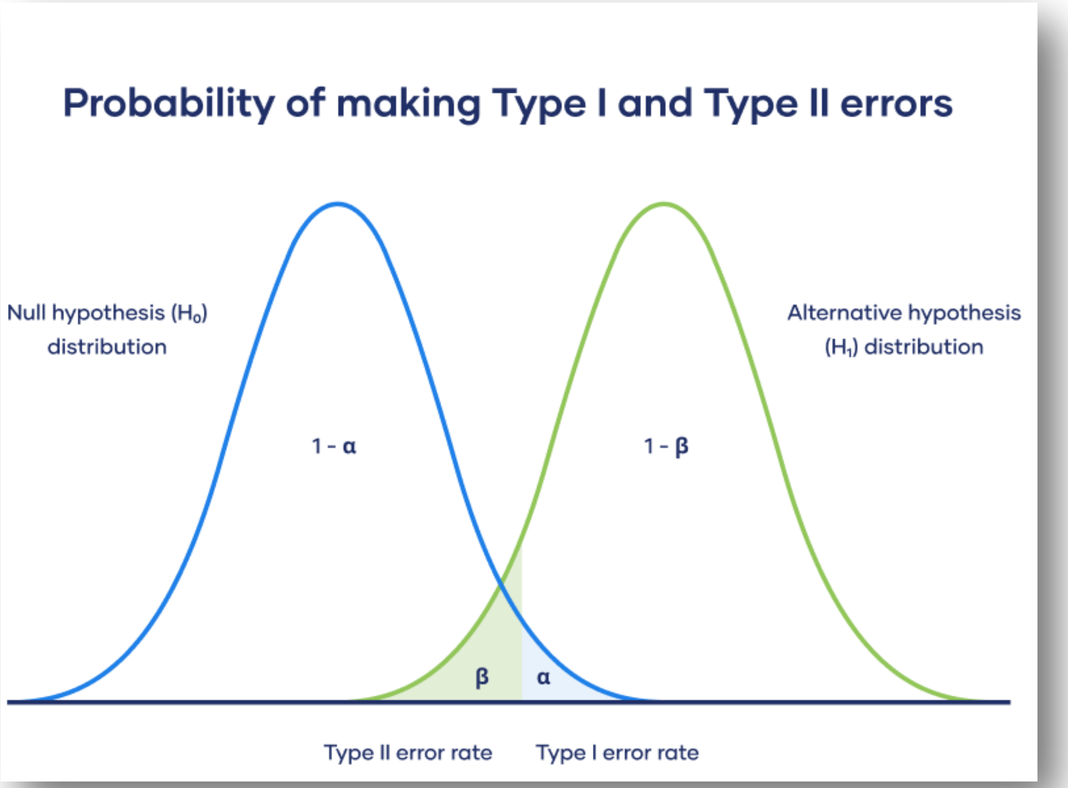
A Type I error means rejecting the null hypothesis when it's actually true.

A Type II error means not rejecting the null hypothesis when it's actually false.

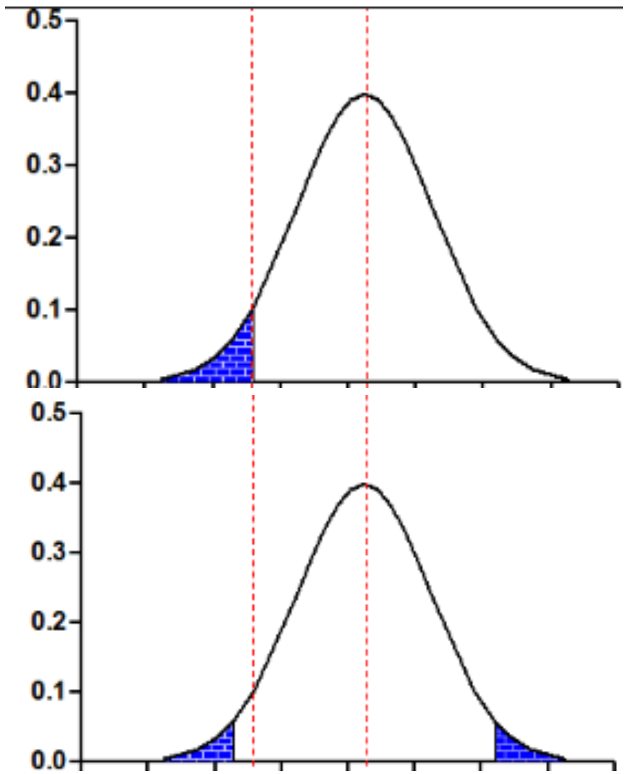
Statistics power is the probability of rejecting a false null hypothesis.

Trade-off between Type I and Type II errors





- Under the same α， for the same dataset， one-sided test is easier to reject H0 than two-sided test
- Wrongly use one-sided test， easily to make type I error
- Wrongly use two-sided test， easily to make type II error



One-sides,α=0.05

Two-sides,α=0.05

Four primary factors affecting power

Significance level (or alpha)

Sample size

Magnitude of the effect of the variable

Standard deviation

Analogies

Test of significance	Court trial
Null hypothesis	Every defendant is innocent until proved guilty
Research design	Police investigation
Data/test statistics	Evidence/exhibits
Statistical principles	Judge’s instruction
Statistical decision	Verdict
Type I error	Conviction of an innocent defendant
Type II error	Acquittal of a criminal

4. NOTES

Notes

① Rigorous research design

- ✓ homogeneity
- ✓ representative

② Understand the meaning of P value

The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.

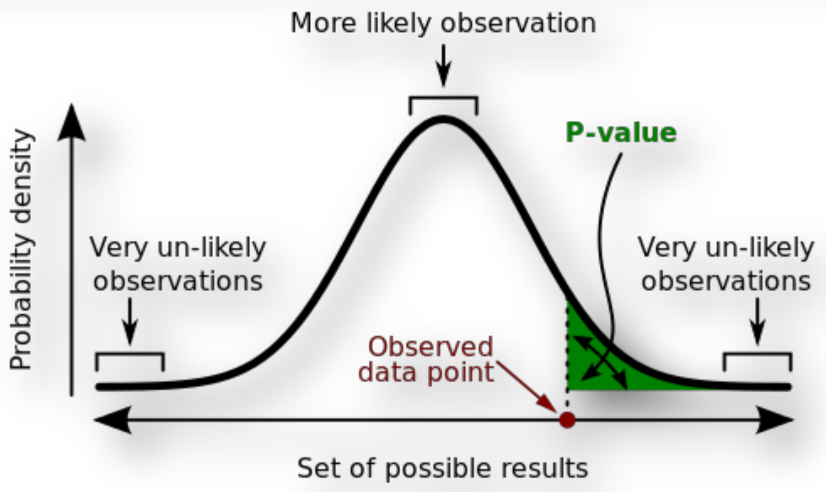
$P1<0.05$，$P2<0.001$; dose not mean the latter parameters have bigger difference.

Important:

Pr (observation | hypothesis) ≠ Pr (hypothesis | observation)

The probability of observing a result given that some hypothesis is true is *not equivalent* to the probability that a hypothesis is true given that some result has been observed.

Using the p-value as a “score” is committing an egregious logical error: **the transposed conditional fallacy.**



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

Notes

③ Statistically significant may not be practically significant

When *n* is large, even a trivial difference may show up as a statistically significant result

- eg menu choice:
  mean selection time of menu a is 3.00 seconds;
  menu b is 3.05 seconds

Statistical significance **does not imply** that the difference is important!

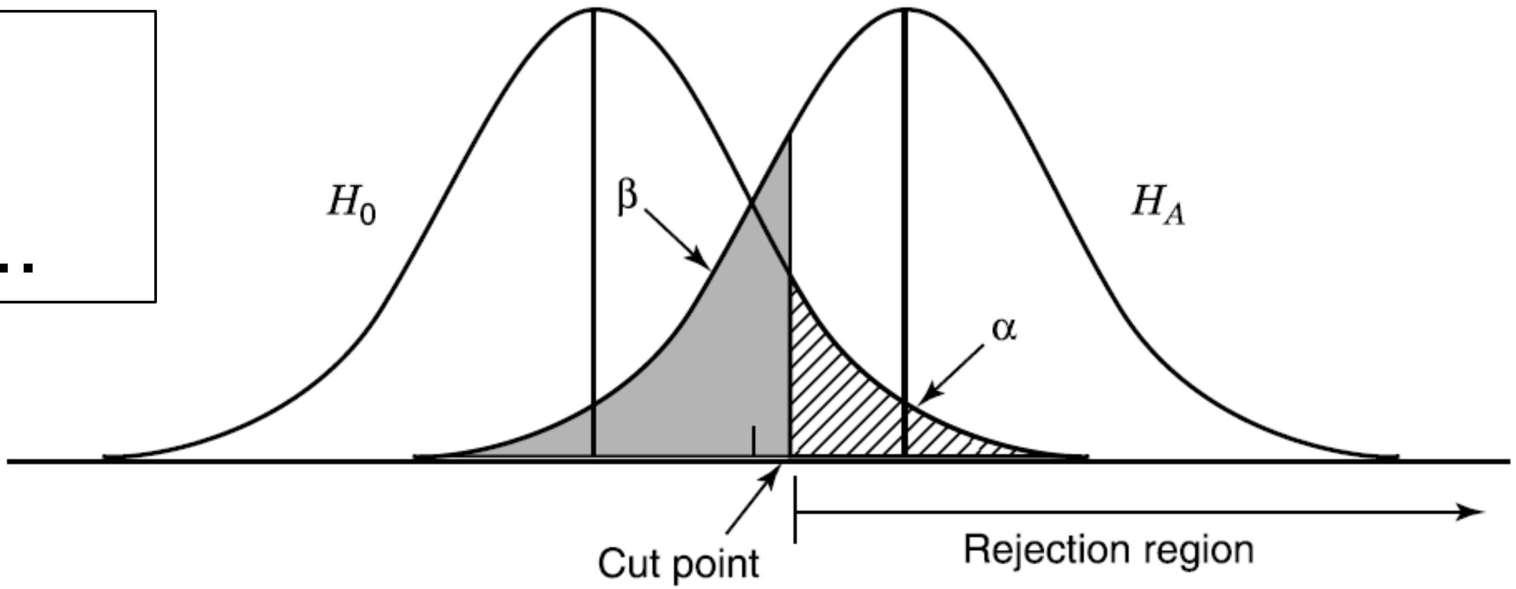
- a matter of interpretation
- statistical significance often abused and used to misinform

A new hypertensive drug could make a 10 mmHg decrease of SBP, while the traditional drug could make a 9.5 mmHg decrease ($P<0.05$).

Notes

④ Not guaranteed to be reproducible

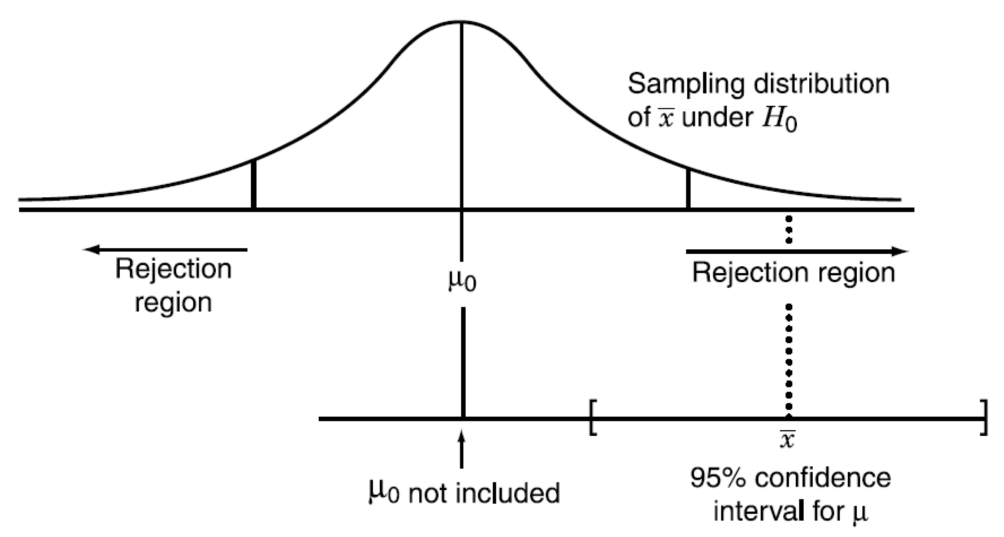
Change α,
Change sample size,
Change covariables……



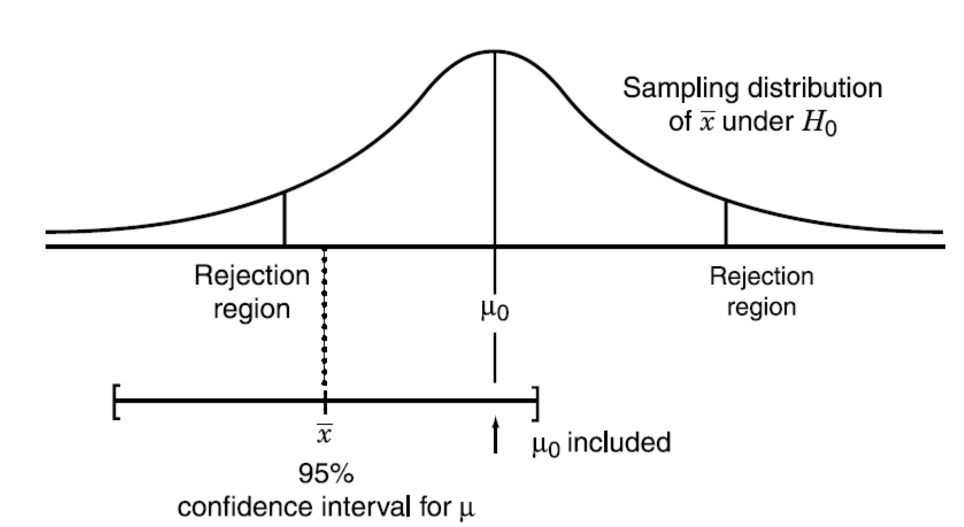
5. RELATIONSHIP TO CONFIDENCE INTERVALS

① If μ0 is <u>not included</u> in the 95% confidence interval for μ, H0 should be <u>rejected</u> at the 0.05 level.

② If μ0 is <u>included</u> in the 95% confidence interval for μ, H0 <u>should not be rejected</u> at the 0.05 level.



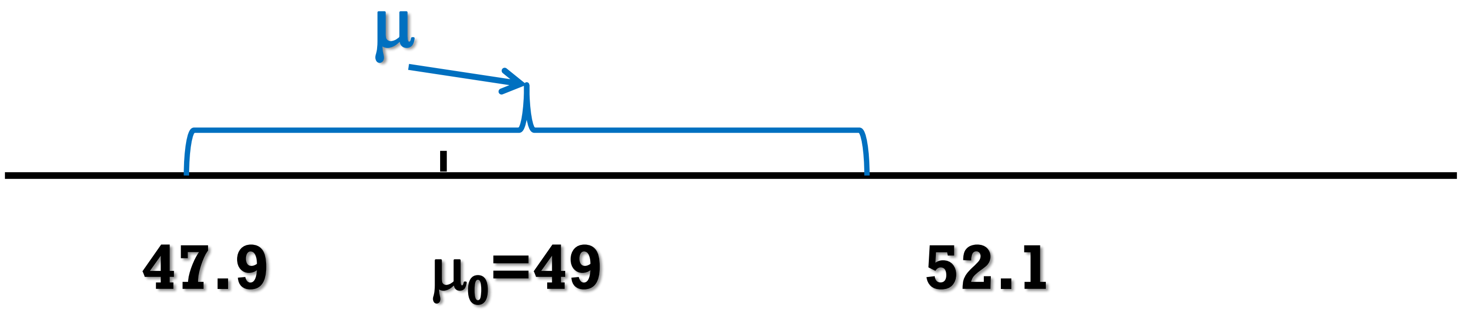
μ_0 not included at 95% confidence interval for μ.



μ_0 included in 95% confidence interval for μ.

Estimate the population mean of birth length of male babies born between 2010 and 2013

$$\bar{x} \pm t_{0.05/2,\nu} \cdot s_{\bar{x}} = 50 \pm t_{0.05/2,24} \times (5/\sqrt{25}) = (47.9, 52.1)$$



Don’t reject H0


Have a Lovely Weekend