Hypothesis Testing: Understanding Significance Level, Type 1 and Type 2 Error

Problem:
Imagine you've developed a COVID test.
- Sometimes, it says someone is **infected** when they **actually are not** — false alarm,
and
- Sometimes it says someone is **not infected** when they **actually are** — it misses real cases.

Both are mistakes, but of different kinds: represented by α and β.

So how much risk of being wrong are you willing to accept?
How would you define above risk in statistics ?
Among the types of risks: α / β, how would you decide which ones to use ?

Hypothesis Testing: Understanding Significance Level, Type 1 and Type 2 Error

When we do a hypothesis test, we generally follow 2 steps:

Step1) We start with two statements about population:

Null hypothesis (H_0): There is no effect or no difference.
Example: “The new drug has no effect on population.”

Alternative hypothesis (H_1): There is an effect or difference.
Example: “The new drug works better than the old one on population.”

Step2) Then we collect sample data, run a test (like a t-test or z-test), and make a decision — either reject H_0 or fail to reject H_0.

Here 4 different scenarios are possible based on our decision

Hypothesis Testing: Understanding Significance Level, Type 1 and Type 2 Error

	H0 is true	H0 is false
Reject H0	ERROR (TYPE 1) α	Correct Decision
Do not reject H0	Correct Decision	ERROR (TYPE 2) β

P(type I error) = α

Type I error occurs when you **reject the null hypothesis (H_0) when it's actually TRUE**.
It's a "false positive" - you conclude something is happening when it really IS NOT happening. For e.g. Person testing "positive" for COVID-19, when he is actually not: false alarm.

α (alpha) is the **significance level** you choose **before** conducting your test (commonly 0.10, 0.05 or 0.01).

If you set **α = 0.05**, you are saying:
"I am okay with being wrong 5% of the time when I say there is an effect."

P(type 2 error) = β

A Type II error occurs when you **fail to reject the null hypothesis when it's actually FALSE**.
It's a "false negative" - you conclude nothing is happening when something really IS happening. For e.g. Person testing "negative" for COVID-19, when he is actually positive: Missed case

If **β = 0.20**, you're saying:
"I'm okay with being wrong 20% of the time **when I say there is *no* effect (missed detection).**"

Hypothesis Testing: Understanding Significance Level, Type 1 and Type 2 Error

So, if you set α = 0.05 and β = 0.20, you are saying:
"I accept a 5% risk of a false alarm (Type I error)
and a 20% risk of missing a true effect (Type II error)."

	H0 is true	H0 is false
Reject H0	ERROR (TYPE 1) α	Correct Decision
Do not reject H0	Correct Decision	ERROR (TYPE 2) β

α: This is chosen by user
β : This is dependent on size of sample n and population variability σ:
As n increases, β decreases
As population standard deviation σ increases, β increase
User does not have direct control on this.

Power of test = 1 – β
It measures the probability of correctly rejecting a false null hypothesis (i.e., detecting a real effect).
High power (close to 1) means the test is good at detecting true effects.

Hypothesis Testing: Understanding Significance Level, Type 1 and Type 2 Error

Problem:
Imagine you've developed a COVID test. Sometimes, it says someone is infected when they're not — and sometimes it misses real cases. Both are mistakes, but of different kinds.
So how much risk of being wrong are you willing to accept?
That's exactly what the significance level and Type I and II errors help us define in statistics.

Ans:
The amount of risk you're willing to accept when being wrong is defined by your

1. **significance level (α)** — the chance of a false alarm (Type I error) — and
2. by **β**, the chance of missing a real effect (Type II error).

When **α** is more important (you want fewer false alarms)

Example: Medical drug approval
You don’t want to approve a drug that doesn’t really work (or is harmful).

Type I Error → Approving a useless/harmful drug.
Type II Error → Failing to approve a helpful one.

Here, **α is more important** — safety first.

→ That’s why Drug Companies often uses **α = 0.01 or 0.001**.

When **β** is more important (you want fewer missed detections)

Example: Disease screening tests
H_0: Person is healthy.
H_1: Person is infected.
Type I Error = False alarm (healthy person flagged).
Type II Error = Missed case (infected person not detected).

Missing a real case can be deadly → **β** is more important.
→ That's why early COVID or cancer screening tests are designed for **high power (low β)**, even if α is higher

Example: Quality control in manufacturing
H_0: Machine producing within tolerance.
H_1: Machine is faulty.

Here, Type II Error, which is letting defective products pass → costly/dangerous.
So **β** is more critical here..

Balanced Scenarios for α and β

When both types of errors are equally important (e.g., routine scientific studies), researchers typically choose:
α = 0.05
Power = 0.80 (β = 0.20)
This balances being too strict (missing effects) and too lenient (false claims).

Extra

α significance level	Test Type	Critical z-value(s) (Found from z-table)
0.10	Two-tailed	+- 1.645
0.10	One-tailed	±1.28 (+ for Right Tailed and – for Left Tailed)
0.05	Two-tailed	±1.96
0.05	One-tailed	±1.645 (+ for Right Tailed and – for Left Tailed)
0.01	Two-tailed	±2.576
0.01	One-tailed	±2.33 (+ for Right Tailed and – for Left Tailed)