Qualitative and Quantitative Data

Qualitative → includes descriptions that don't contain numeric values

Quantitative → contains numbers obtained by counting or measuring

- Discrete data
  - certain exact value
- Continuous data
  - can have any value
  - → heights

Types of Data

↓

levels of measurement

- four
  - Nominal , ordinal , Interval , Ratio
  - refer to different ways That values can be @uantified or categorized

Metric level

Why we need level of measurement?

→ it tells us how our data can be collected , analyzed and interpreted

1. Nominal variables

↳ most basic level

Examples ⟶ gender , Types of animals

⟶ it is not possible to rank categories

⟶ we can categorize and count Responses but cannot infer any order

2. Ordinal Variables

↳ can be categorized

possible to have a meaningful ranking of the categories

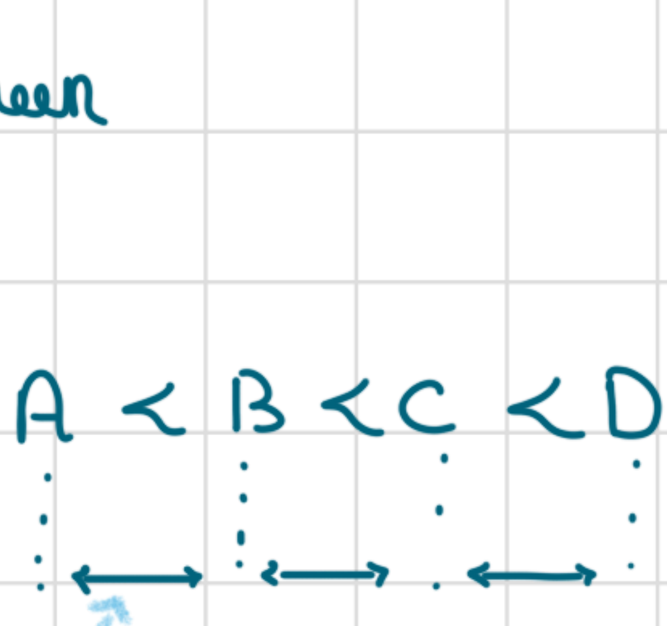
Examples : Rankings , level of education

→ allows us to rank responses but not to measure precise differenes between ranks

3. Metric variables

↳ highest level of measurement

→ it is like ordinal but intervals between values are equally spaced



Examples : Income , weight . Age

→ enables us to measure exact differences between data points

interval

have no true

zero point.

Ratio

have a true zero

point

represent the

absense of

the quantity

being measured

Steps in data-driven decision-making.

1- Formulate a hypothesis
2- Find the right test
3- Execute The test
4. Make a decision

→ what is hypothesis ?

↓ is an idea that can be tested.

something that can be compared

→ Null hypothesis H_0

. The one to be tested

. There is no difference between two things

→ Alternative hypothesis H_1 or H_A

. everything else

.

Two Types of errors

α Type I error → reject the hypothesis test

↳ bad

Ho True

β Type II error → fail to reject hypothesis test

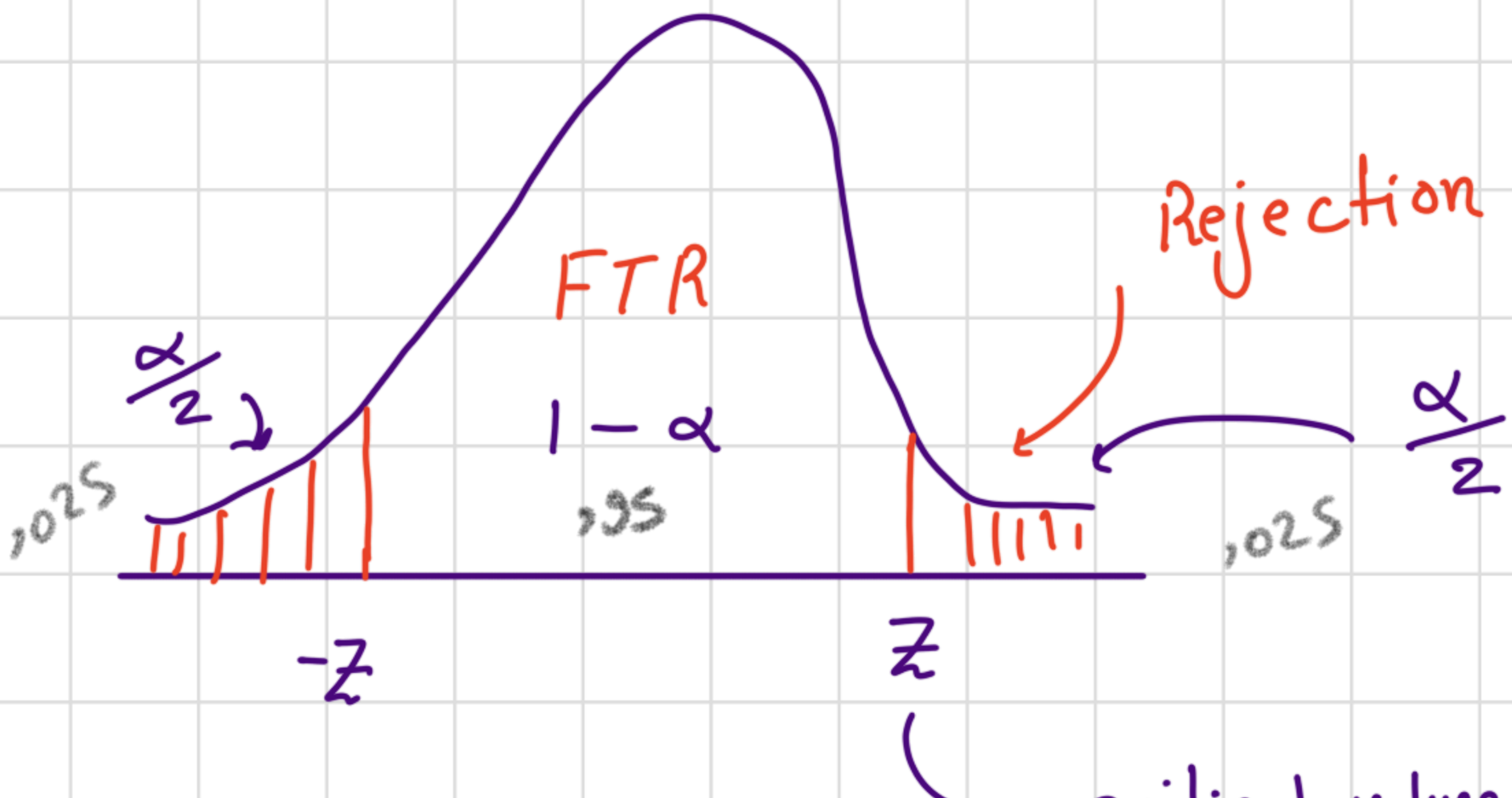
Ho False

bad decision

One tailed and two tailed Tests.

$H_0 : \mu = 1009$

$H_a : \mu \neq 1009$



↳ critical values

⟶ $C = ,95$ ⟶ confidence level

$\alpha = 1 - C$ ⟶ significance level

$= ,05$

in order to determine whether you should reject or not reject the null hypothesis

→ Calculate Z value and compare to critical value

$H_a: \mu < 100g$

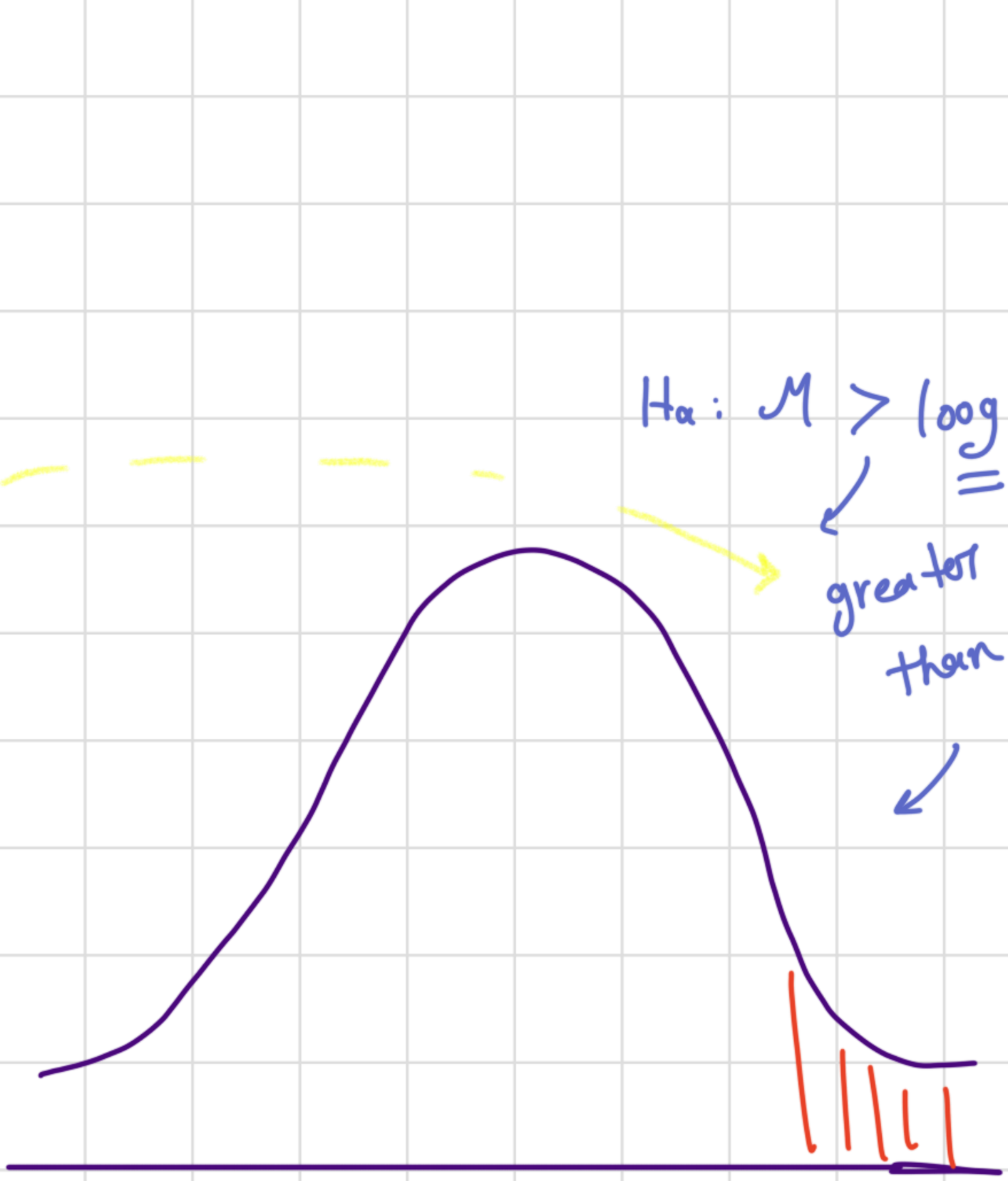
less than

left tail test

right tail test

$H_a: \mu > 100g$

greater than

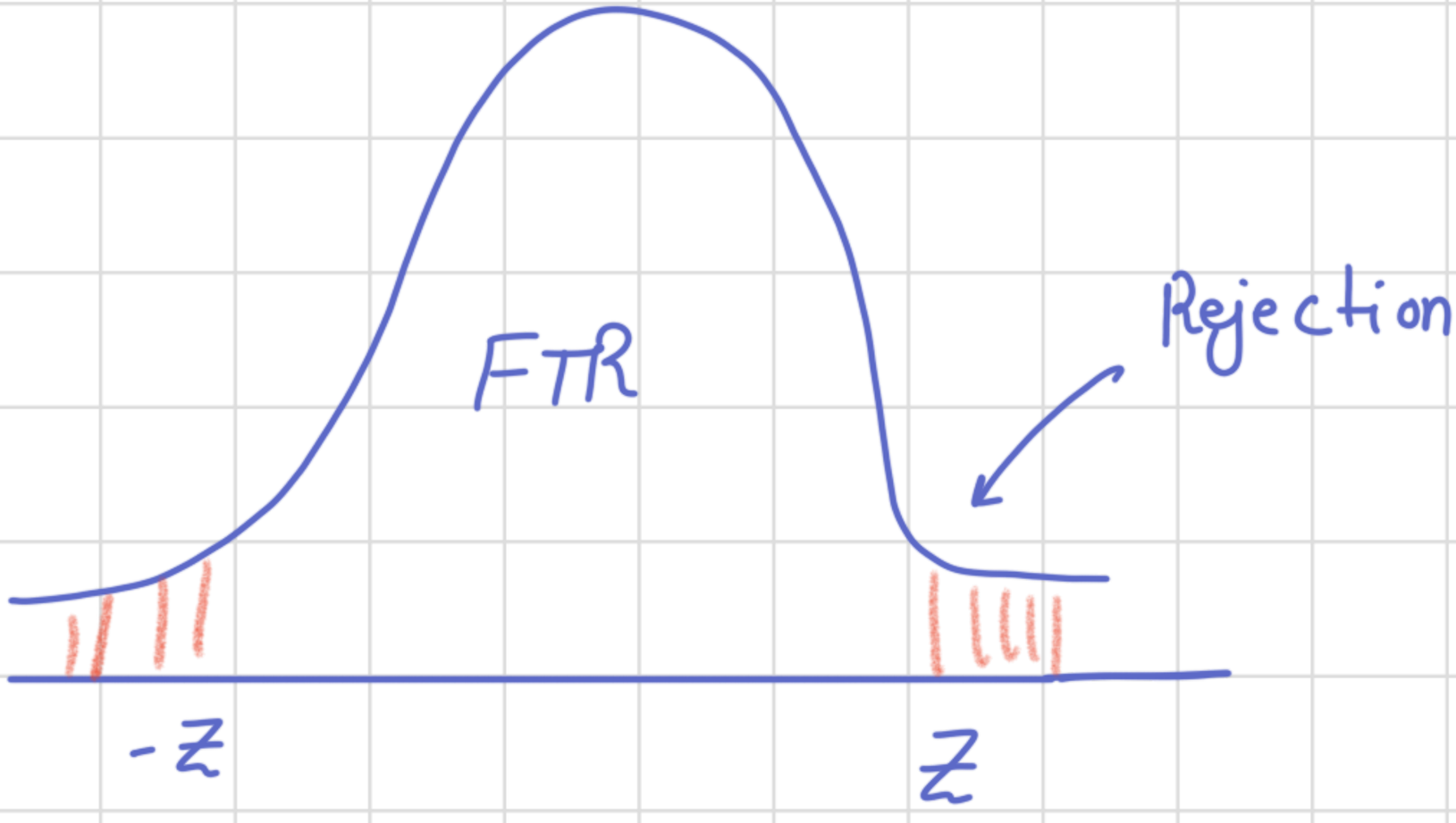


Test statistic

$C = 95\%$

$H_0 : \mu = 50$

$H_a : \mu \neq 50$



a) → bad X H_0 True

type 1 error

reject

b) → good ✓ H_0 False

c) → bad X H_0 False

d) → good ✓ H_0 True

type II error

fail to reject

p values

p value > ,05

fail to see a difference

p value < ,05

There is a difference.

small p value with no difference

↳ False positive

Confidence intervals

range in which true mean is likely to be found with certain probability

For normally distributed

$$CI = \hat{x} \pm z \cdot \frac{s}{\sqrt{n}}$$

Regression Analysis

→ method for modeling relationships between variables

→ infer or predict a variable.
↓
dependent variable

→ The variable we use for prediction
↓
independent variables

Regression analysis can be used to achieve two goals:

→ measure the influence of one variable or more on another variable

→ prediction of one variable by one or more other variables

Types
- simple linear
- multiple linear
- logistic