MISSING VALUES - MCR, MAR, MNAR

🎯 1. Missing Completely At Random (MCAR)

Definition (simple):

Missingness has nothing to do with the data — neither observed nor missing.

Meaning:

- The data is missing for purely random reasons.
- Everyone has the same probability of missing data.

Example:

- A computer glitch deletes 5% of rows randomly.
- A lab machine occasionally fails regardless of the patient's condition.

Analogy:

Imagine dropping a stack of papers and losing 3 random sheets. The lost pages have no pattern.

Why it matters:

✔️ Best-case scenario

✔️ You can delete missing rows with no bias

✔️ Most statistical methods stay valid MCAR is ideal, but rare in real life.

🎯 2. Missing At Random (MAR)

Definition (simple):

Missingness depends on observed data, but not on the missing value itself.

Meaning:

- The reason data is missing is related to something you can measure.
- If you account for those variables, the missingness becomes explainable. Example:
- Older people are more likely to skip income questions.(So missing income depends on age, which is observed.)
- Weight is missing more often for women than men.(Missingness depends on gender, which is known.)

Analogy:

You lose pages on a windy day.If you knew the wind direction (observed variable), you could understand which pages were more likely lost.

Why it matters:

✔️ MAR is very common

✔️ Multiple imputation, regression, and ML methods work well

✔️ Bias can be corrected if you include variables related to missingness

🎯 3. Missing Not At Random (MNAR)

Definition (simple):

Missingness depends on the missing value itself (even after considering observed data).

Meaning:

- The reason data is missing is tied to what the value would have been.
- Even if you know all observed variables, missingness still depends on the unobserved truth. Example:
- People with very high income refuse to report it.
- Depressed people skip mental health surveys.
- Patients with severe symptoms don't come to follow-up visits.

Analogy:

You lose the embarrassing pages of your diary because you hid them yourself.The missingness is caused by the content of the missing page.

Why it matters:

❌ Hardest type

❌ No statistical technique can magically fix it

❌ Requires modeling the missingness mechanism explicitly(Example: selection models, sensitivity analysis)

🧠 Quick Summary Table

Type	Why data is missing	Can we fix it?	Example
MCAR	Pure randomness	Easy	Random system glitch
MAR	Depends on observed variables	Usually yes	Older people skip income
MNAR	Depends on missing values	Hard	Rich people hide income

📝 Rule of Thumb

✔️ MCAR → Safe to delete rows

✔️ MAR → Use imputation methods

✔️ MNAR → Need special modeling (not easy)