

Missing Data

MCAR

Missing Completely At Random -
missingness unrelated to data



MAR

Missing At Random - related to observed
variables

Group A:

Group B:

Depends on group

MNAR

Missing Not At Random - related to
unobserved values

Low:

High:

Missing if high value

Imputation Strategies

Mean/Median

Simple but biased

K-NN

Uses similar samples

MICE

Multiple Imputation

MissForest

Random Forest based

EM Algorithm

Maximum Likelihood

Deep Learning

Autoencoders, GANs