Handling Missing Data

Structurally Missing Data

Structurally Missing Data is data that is expected to be missing.
For example, there are structurally missing data in the 'Litters' and 'Pups/Litter' columns for all the male dogs in the table below because we would not expect male dogs to have puppies.

ID#	Name	Breed	Sex	Litters	Pups/Litt
1	Gnasher	ACD	M		
2	Cassie	Collie	F	1	3
3	Pepper	French Bulldog	F	4	2
4	Jed	Golden Retreiver	M		
5	Henry	Spaniel	M		
6	Ruby	ACD	F	1	6

Data Missing Not at Random

Missing Not at Random (MNAR) data is missing for reasons that cannot be inferred. These data are systematically missing, meaning their missingness may be predictable by the value of another variable, but there is no clear explanation as to why. In the table below, the sales data for bananas is missing, but try as you might, you cannot figure out why it is missing. Bananas were stocked and sold every week that data was collected! The missing banana data is MNAR data.

Week	Fruit	TotalSales
1	Apple	300
1	Banana	
1	Lemon	100
2	Apple	330
2	Banana	
2	Lemon	110
3	Apple	200
3	Banana	
3	Lemon	60

Missing at Random Data

Missing at Random (MAR) data is missing because of some random characteristic about the person or thing being studied. Often, this type of data is reliably missing based on the value of another variable in the dataset.
In the table below, the bacterial cell counts for all the stool samples are 'NaN'. If we looked into this, we might find that there were too many bacterial cells to count in all those samples. Therefore, the bacterial cell counts for stool samples would be MAR data.

Sample ID	Sample Type	Bacterial Cell Counts
1	Hand Swab	1008
2	Stool	NaN
3	Mouth Swab	7876
4	Hand Swab	657
5	Stool	NaN
6	Hand Swab	2442
7	Mouth Swab	5444
8	Stool	NaN
9	Hand Swab	4654
10	Stool	NaN

Data Missing Completely at Random

Missing Completely at Random (MCAR) data has no detectable underlying reason causing the values to be missing.
The table below has MCAR data. The # of fruits is missing for some plants, but the missing fruit data seems unrelated to the height of the plant. Short and tall plants are both missing fruit data. In addition, we are missing the height for one of our plants!

Plant	Height (cm)	# of Fruits
1	65	10
2		87
3	987	
4	44	
5	105	35
6	547	74
7	876	
8	55	
9	875	95