

Simpson's Paradox:-

When aggregated data lead us to draw different conclusion than we would from the original data

* Aggregated two or more tables *

$$\text{Percentage} = \frac{F}{N} * 100 \Rightarrow [0: 100]$$

$$\text{Proportion} = \frac{F}{N} \Rightarrow [0: 1]$$

Ex:-

| | Male | | | | Female | | |
|----------|---------|----------|------|--|---------|------|------|
| | Applied | Admitted | rate | | Applied | Adm- | rate |
| Major(A) | 900 | 450 | 50% | | 100 | 80 | 80% |
| Major(B) | 100 | 10 | 10% | | 900 | 180 | 20% |

⇒ When we look in data alone it seems Female rate is bigger than male rate

~~when~~ When aggregated two tables

| | APPLIED | ADMITTED | Rate | | APPLIED | ADMITTED | Rate |
|----------|---------|----------|------|--|---------|----------|------|
| Major(A) | 900 | 450 | 50% | | 100 | 80 | 80% |
| Major(B) | 100 | 10 | 10% | | 900 | 180 | 20% |

⇒ When we look in data alone it seems Female rate is bigger than male rate

⇒ ~~when~~ when aggregated two tables

Male → 1000 460 46%

Female → 1000 260 26%

We find Male rate is bigger