Statistics Fundamentals

Types of Data
- Categorical
- Numeric

→ features that have text values
eg gender (M/F), size (L, M, S) . . .

→ features are numbers (int, float)
eg qty sold, sales price
# of children . . .
- discrete (by counting) eg # of children
- continuous (by measure) eg temp

Types of Statistical Analysis

Measures of Central Tendency

Measures of Dispersion

Measures of Lack of Symmetry
- Skewness
- Kurtosis

Important Terms in Statistics
- Population
- Sample
- Population Parameter
- Sample Statistic
- Variable
- Distribution

Types of Statistical Analysis

✓ ① DESCRIPTIVE STATISTICS
(what happened?)

DIAGNOSTIC STATS/ANALYTICS

PRESCRIPTIVE STATS

② PREDICTIVE STATISTICS ⟶ ML ALGOS
(what will happen?)

✓ ③ INFERENTIAL STATISTICS
(Estimate something about the population using a sample collected from the popn)

Measures of Central Tendency

eg Average (mean), median
numeric features

mode eg gender → M → 452, F → 548
↓
used for categorical features

Measures of Dispersion

Extent of fluctuation in numeric values

- RANGE : max − min
  eg Age 92 − 18 = 74
- IQR $Q_3 - Q_1$
  75 percentile − 25 percentile
- STANDARD DEVIATION = $\sqrt{\text{variance}}$
- VARIANCE

Measures of Lack of Symmetry

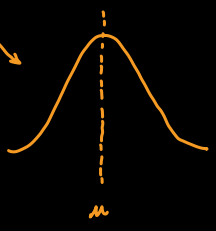
- Skewness
- Kurtosis

SKEWNESS :

↓

Lack of Symmetry

$0 \longrightarrow$ perfectly symmetric

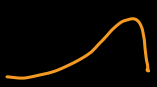


$> 0 \longrightarrow$ positive skew

Right skew

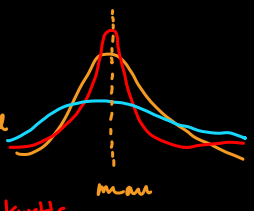
$< 0 \longrightarrow$ negative skew

Left skew



KURTOSIS

kurtosis = 3 $\longrightarrow$ Perfect normal dist (mesokurtic)



kurtosis > 3 $\longrightarrow$ Leptokurtic

kurtosis < 3 $\longrightarrow$ platykurtic

Important Terms in Statistics

- **Population** (people, transactions, ... products)
- **Sample** (subset of the popⁿ) size sample < size popⁿ.
- **Population Parameter**
- **Sample Statistic**
- **Variable** → Entity which takes on different values.
- **Distribution**

eg (mean) average adult wt → context of popⁿ → popⁿ parameter

context of Sample → Sample Statistic

eg temp

↓

the pattern of variation of a variable eg adult wt, IQ, income