Histogram

To create histogram do :

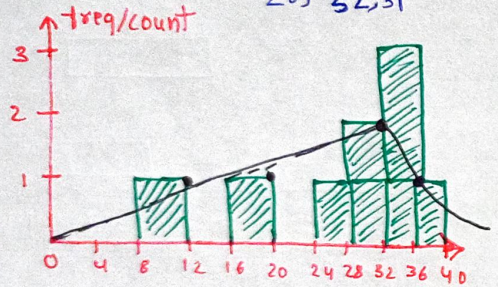
① Sort the values (asec)

② Create bins (groups)

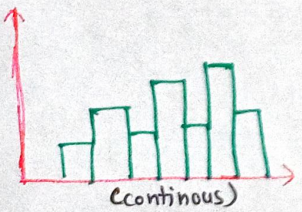
No. of Bins = My choice

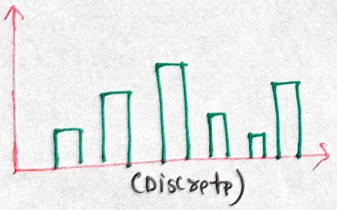
$$\text{No. of bins} = \frac{\text{Last value (Max)}}{\text{First value (Min)}} \text{ or No. of Bins}$$

Ex→ Values = [10, 20, 25, 30, 35, 40, 26, 32, 31]



- Plotting continous values as well as discrete.





PDF: Probability density function

PMF: Probability mass function

- PDF use to smooth discrete histogram and PMF used to smoothen continous histogram.

Measure of Central Tendency (MCT)

MCT is single value that attempts to describe a set of data identifying the central tendancy.

① Mean/Average:

$$\text{Population mean } (\mu) = \sum_{i=1}^{N} \left[\frac{x_i}{N}\right]$$

N >> n

$$\text{Sample mean } (\bar{x}) = \sum_{i=1}^{n} \left[\frac{x_i}{n}\right]$$

But, $\bar{x} > \mu$ or $\mu > \bar{x}$ depending on sampling technique.

➡ In practical, we use mean to replace NAN values [Although not efficient] due to outliers]

② Median:

To calculate it do

↳ sort the values

↳ Find central number

↳ if no. of values even, find average of central numbers

↳ if no. of values odd, find the central number

→ Central number is value in mid of the all values.

→ If no outliers use mean, if outliers use median to replace the NAN values.

③ Mode:

most frequent occuring element

→ Used to replace the NAN values in categorical data within the most frequent value.

④ Measure of Dispersion:

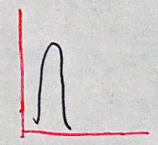
4.1 Variance (σ^2)

- spred of data

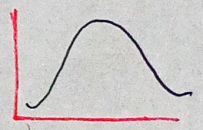
$$\text{Population variane } (\sigma^2) = \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{N}$$

$$\text{Sample variance } (s^2) = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}$$

Variance ∝ Spread of data



Less variance, less spread



More variance, more spread

4.2 Standard Deviation ($\sqrt{\sigma^2}$) (SD) $= \sigma$

- Tells how many standard deviation a value is away from mean.

Ex→ {1, 2, 3, 4, 5}, $\mu = 3$, $\sigma^2 = 2$, $\sigma = 1.41$