Date.

cheat sheet for prob and statistics

Topic 1: Frequency Distribution

- Discrete Frequency Distribution → single value grouping

x	frequency

- Continuous Frequency Distribution → limit value grouping

width = $\frac{\text{range}}{\text{\# of classes}}$, range = max − min

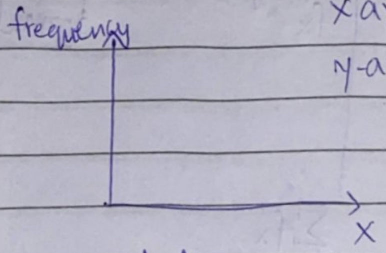
width → always rounded up.

class interval	frequency

Topic 2: Histograms → no spaces between boxes

- For single value / discrete data

x	frequency

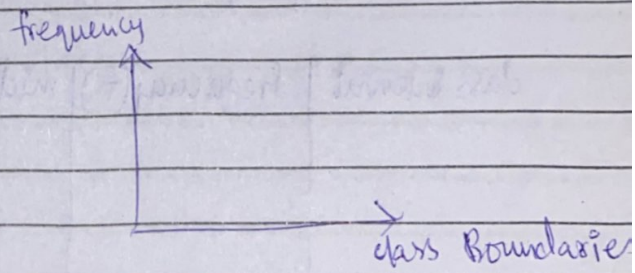


x axis = x
y-axis = frequency

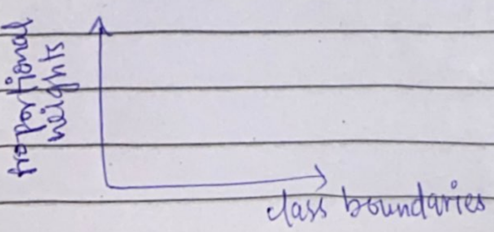
- For limit grouping / continuous data

1- find the class boundaries

interval	class boundaries	frequency



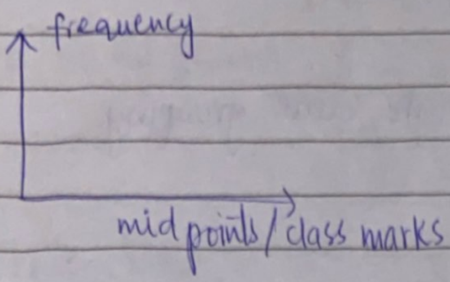
- for unequal intervals histogram:



proportional heights = $\frac{\text{frequency}}{\text{class interval / width}}$

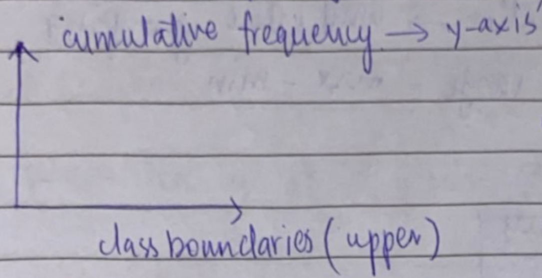
Date..........

Construction of frequency Polygon → closed figure



steps:
1- Take 2 additional groups (start & end) with frequency = 0
2- calculate mid points
3- plotting

Cumulative frequency Distribution / ogive



steps:
1- add one group before the 1st group.
2- make class boundary and (less than) class boundary column
3- calculate cumulative frequency

The Mean

- single value grouping data

x	f	fx
	$\sum f$	$\sum fx$

formula: mean $= \dfrac{\sum fx}{\sum f}$

- mean of grouped data

mean $= \dfrac{\sum fx}{\sum f}$

class interval	frequency (f)	midpoint (x)	fx

steps:
1- calculate (x) midpoint
2- calculate fx
3- calculate mean.

Date.

The Median

- Single-value grouping

x	f	cf(<)

median class
↓
median pos $= \frac{\Sigma f}{2}$

Steps:

1. find cumulative frequency
2. calculate median class
3. look in the cf(<) range and median = x in median class

- Median of grouped data

formula: $\text{median} = l + \frac{h}{f}\left(\frac{\Sigma f}{2} - cf(<)\right)$

class interval	boundaries	f	cf(<)

l = LCB of median class
h = width of median class
f = frequency of median class
$cf(<)$ = cf(<) of previous class to median class

Steps:

1. Fill and find the upper table
2. Find median class using formula $\frac{\Sigma f}{2}$
3. Use formula to find median value

The Mode:

- single-value grouping type data

x	f

steps:

1. mark the class in which highest frequency occur.
2. x of that class = ~~median~~ mode

- Mode of grouped data: $\boxed{\text{modal class} = \frac{\Sigma f}{2}}$

formula: $\text{mode} = l + h \times \frac{(f_m - f_1)}{2f_m - f_1 - f_2}$

class interval	class boundaries	frequency	cf(<)

f_2 = freq. of next class to modal class
f_1 = freq of prev class to modal class
f_m = freq of modal class

Quartiles

$$Q_i = \frac{i(n+1)}{4}^{th} \text{ value}$$

- Single-value grouping type data

Q_1, Q_2, Q_3, Q_4 ??

x	f	cf(<)

steps:

1. Find cf(<)
2. Find quartile class: $\frac{i \sum f}{4}$
3. look and mark by observing in cf(<) ranges
4. x of that class = Q_i

- Quartiles of grouped data

class interval	class boundaries	frequency	cf(<)

Quartile class $(Q_3) = \frac{3 \sum f}{4}$

steps:

1. Find cf (find and fill table)
2. mark Qi class
3. use formula.

Formula:

$$Q_i = l + \frac{h}{f}\left(\frac{i \sum f}{4} - c.f(<)\right)$$

Foundation for Advancement of Science & Technology

Lower and Upper Limits

$$\text{lower limit} = Q_1 - 1.5\,IQR$$
$$\text{upper limit} = Q_3 + 1.5\,IQR$$

Five Number Summary

Min, Q_1, Q_2, Q_3, Max

BoxPlots

To construct a Boxplot

step 1 : Determine the boxplot

step 2 : Determine potential outliers, and the adjacent values

step 3 : Draw a horizontal axis on which the numbers obtained in steps 1 and 2 can be located. Above this axis, mark the quartiles and the adjacent values with vertical lines.

step 4 : Connect the quartiles to make a box, and then connect the box to the adjacent values with lines.

step 5 : Plot each potential outlier with an asterisk.

Foundation for Advancement of Science & Technology

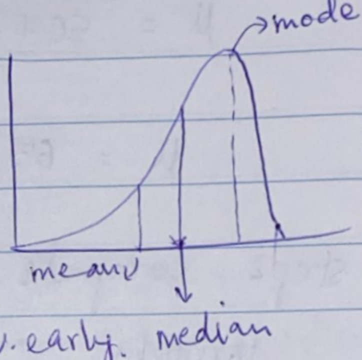
Symmetrical Distribution

- if the frequencies are equally distributed on both the sides of central value.
- a symmetrical distribution may be either bell-shaped or u-shaped.
- In symmetrical distribution, the values of mean, median and mode are equal i.e Mean = Median = Mode

Skewed Distribution

Skewed Distribution < positively skewed / negatively skewed

- skewness is used to measure the level of asymmetry in our data. It is the measure of asymmetry that occurs when our data deviates from the norm.



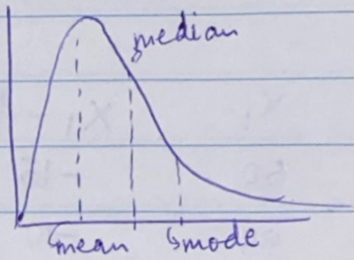
Negatively Skewed:

- The distribution is skewed to the left
- Mode exceeds Mean and Median

example: most people retire at 60-65, but some retire v. early.

Positively skewed:

- The distribution is skewed to the right
- Mean exceeds Mode and Median.



example: income distribution (few v. high incomes pull the mean up).

Population Skewness

measures the asymmetry for the entire population.

formula:

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

where,

$$\mu_3 = \frac{1}{N}\sum (x_i - \mu)^3$$

third central moment

σ = population sd

Sample Skewness

estimates skewness based on the sample from the population.

$$G_1 = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

n = sample size
$\bar{X}$ = sample mean
s = sample sd.

Population Skewness Calculation

Q. Given population dataset: 50, 55, 60, 65, 70, 75, 80

Step 1: Calculate Mean (μ)

$$\mu = \frac{50+55+60+65+70+75+80}{7}$$

$$\mu = 65$$

Step 2: Compute Standard Deviation (σ)

Formula:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

x_i	$x_i - \mu$	$(x_i - \mu)^2$
50	-15	225
55	-10	100
60	-5	25
65	0	0
70	5	25
75	10	100
80	15	+ 225
		700

$$\sigma = \sqrt{\frac{700}{7}}$$

$$\sigma = \sqrt{100} = 10$$

step 3: compute the Third Central Moment (μ_3)

Formula: $\mu_3 = \frac{1}{N} \sum (x_i - \mu)^3$

x_i	$x_i - \mu$	$(x_i - \mu)^3$
50	-15	-3375
55	-10	-1000
60	-5	-125
65	0	~~1250~~ 0
70	5	~~1000~~ 125
75	10	1000
80	15	3375
		0

$\mu_3 = \frac{0}{7} = 0$

Step 4: compute skewness

$\gamma_1 = \frac{\mu_3}{\sigma^3} = 0$

skewness = 0
the data is perfectly symmetric.

Sample Skewness Calculation.

Q. Given sample dataset
5, 7, 9, 11, 13

step 1: Compute Mean ($\bar{x}$)

$$\bar{x} = \frac{5+7+9+11+13}{5}$$

$$\boxed{\bar{x} = 9}$$

Step 2 : Compute the Sample Standard Deviation (s)

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
5	-4	16
7	-2	4
9	0	0
11	2	4
13	4	16

$$s = \sqrt{\frac{40}{4}} = \sqrt{10}$$

$$\boxed{s \approx 3.16}$$

Step 3 : Compute Sample Skewness (G_1)

$$G_1 = \frac{n}{(n-1)(n-2)} \Sigma \left(\frac{x_i - \bar{x}}{s}\right)^3$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^3$
5	-4	-64
7	-2	-8
9	0	0
11	2	8
13	4	64

$$\Sigma \frac{(x_i - \bar{x})^3}{s^3} = \frac{-64 - 8 + 0 + 8 + 6}{(3.16)^2}$$

$$= 0$$

So, $G_1 = \frac{5}{(5-1)(5-2)} (0) = 0$

Since sample skewness = 0, the data is perfectly symmetric.

- if skewness was positive ($\gamma_1 > 0$ or $G_1 > 0$) → long right tail
- if skewness was negative ($\gamma_1 < 0$ or $G_1 < 0$) → long left tail

Kurtosis

measures the tailedness or sharpness of the peak of a probability distribution. It describes how the tails of the distribution compare to a normal distribution.

Types:

1- Leptokurtic ($\beta_2 > 3$) → heavy tailed distribution → more outliers (extreme values) → peaked shape

2- Platykurtic ($\beta_2 < 3$) → lightly tailed distribution → fewer outliers → flatter shape

3- Mesokurtic ($\beta_2 = 3$) → normal-tailed distribution → similar to a normal distribution → standard shape.

Population Kurtosis

$$\beta_2 = \frac{\mu_4}{\sigma^4}$$

μ_4 = 4th central moment

$$\mu_4 = \left(\frac{1}{N}\sum_i (x_i - \mu)^4\right)$$

σ = sd

steps:

1. compute mean (μ)
2. compute sd (σ)
3. compute the 4th central Moment (μ_4)
4. compute Kurtosis

Sample Kurtosis:

$$G_2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$