Date / /
Page No.

Choosing an Appropriate Bin width
The bin width (class interval width) determines the number of intervals and affects how the histogram looks.
- Too many intervals → Too detailed, looks 'choppy'
- Too few intervals → Oversimplified, lacks detail

Two common rules help determine bin width:

1 Sturges' Rule
- Number of intervals ≈ $1 + \log_2(N)$ (or $1 + 3.3 \log_{10}(N)$)
- This rule works best for small datasets but may underestimate intervals for larger datasets

2 Rice Rule
- Number of intervals $= 2 \times$ (cuberoot of N)
- Preferred for larger data sets

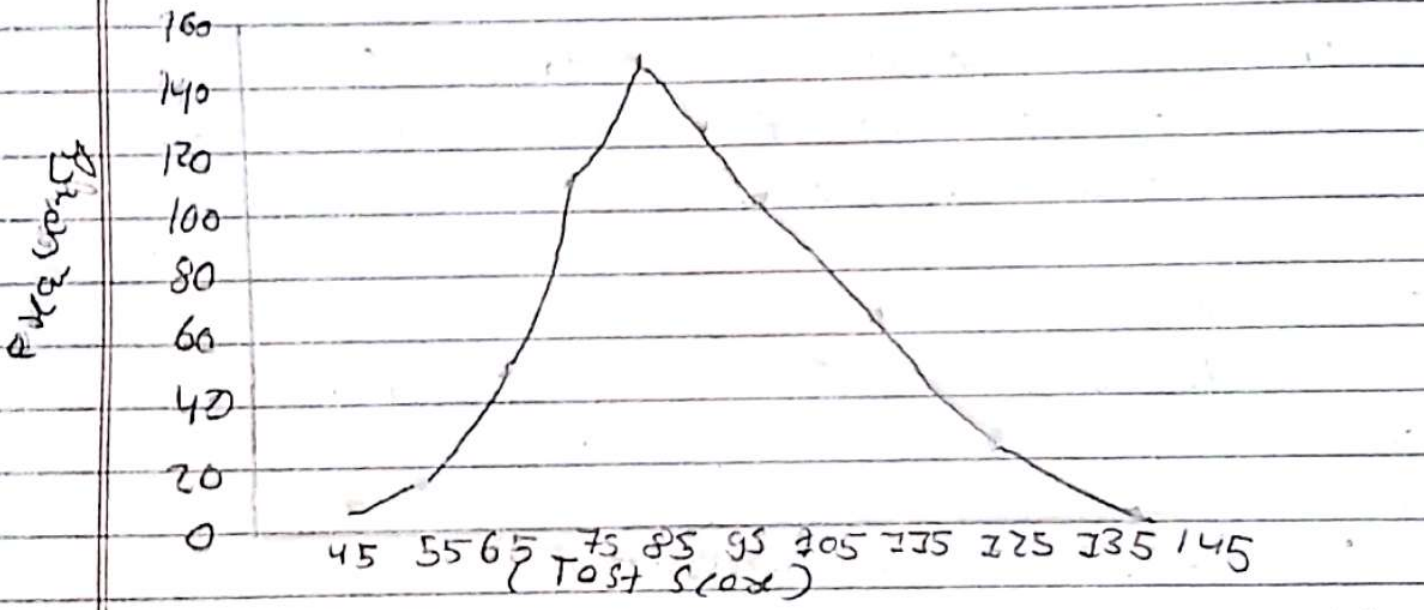
~~Relative frequency Histograms~~

Summary of frequency polygons
A frequency polygon is a graph that helps visualize the shape of a dataset's distribution. It works similarly to histogram but connects data points with line instead of using bars.

How to create a frequency polygon
Begin by grouping data into class intervals, similar to histograms. Then calculate
Class frequencies (how many value fall into each interval)
Midpoints of class intervals (used for plotting)

Example Psychology Test scores

Interval (Lower limit - Upper limit)	MidPoint	Frequency
39.5 - 49.5	45	3
49.5 - 59.5	55	18
59.5 - 69.5	65	53
69.5 - 79.5	75	107
.....	85	147



The psychology test scores skew right because high scores trail off more gradually.

Cumulative frequency Polygons

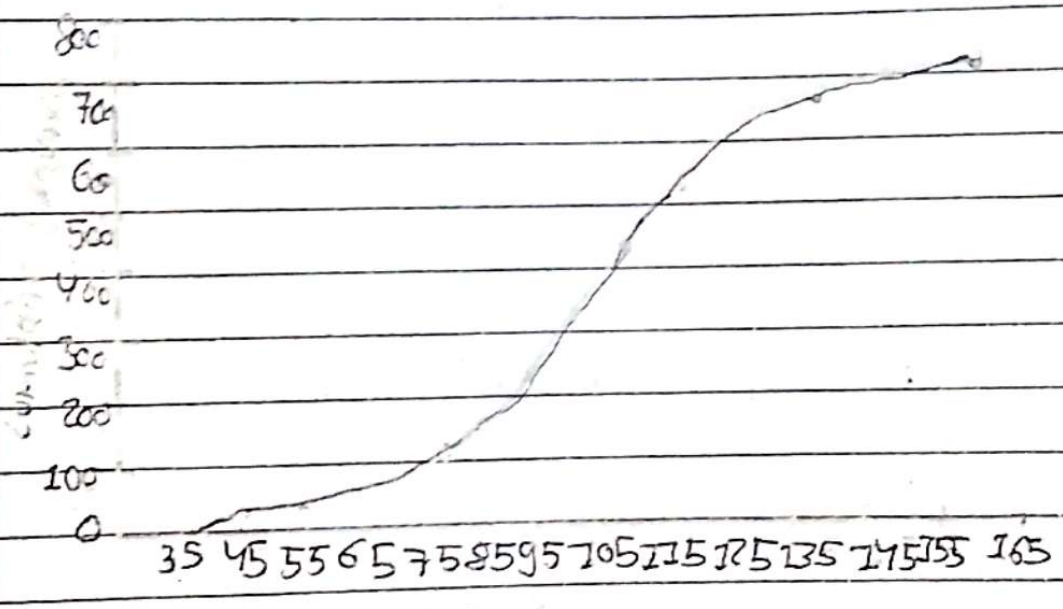
A cumulative frequency polygon (also called an Ogive) represents the total count of observations up to a given class interval

How to Create a cumulative frequency polygon

Instead of plotting raw frequencies, each y-value is the cumulative total of all previous frequencies

- Instead of plotting raw frequencies, each Y-value is the cumulative total of all previous frequencies
- The final cumulative frequency equals the total number of observations

Interval (Lowerlimit-Upperlimit)	Midpoint	Frequency	Cumulative frequency
39.5 - 49.5	45	3	3
49.5 - 59.5	55	10	13
59.5 - 69.5	65	53	66
69.5 - 79.5	75	107	173
79.5 - 89.5	85	147	320
---	---	---	---



Key insights from a cumulative frequency polygon

- Helps determine percentiles and medians easily

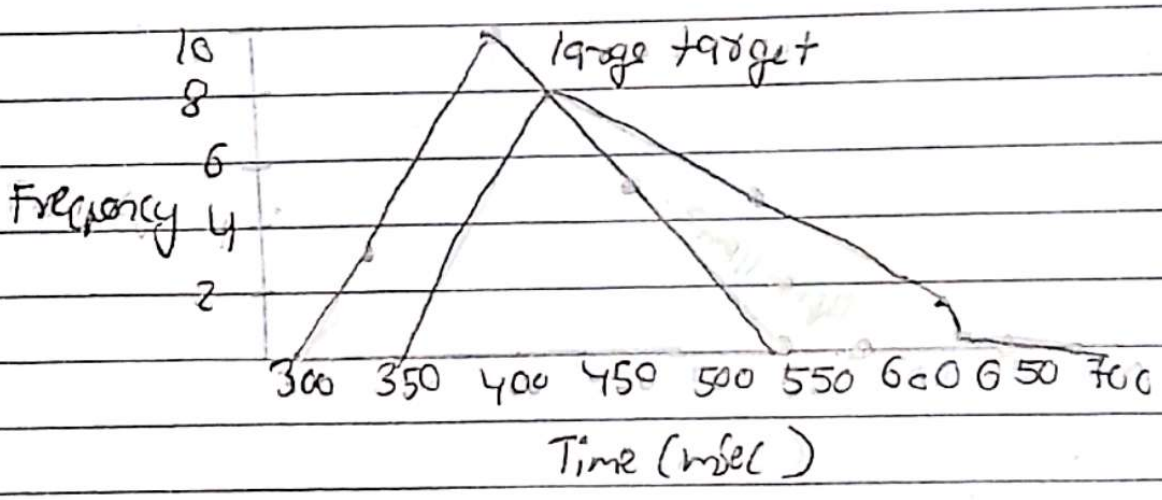
The cumulative frequency of 55 corresponds to 13 students, meaning 13 students scored 55 or lower

Overlaid Frequency Polygons

Comparing two or more datasets is easier with frequency polygons than histograms, since multiple distribution can be plotted together

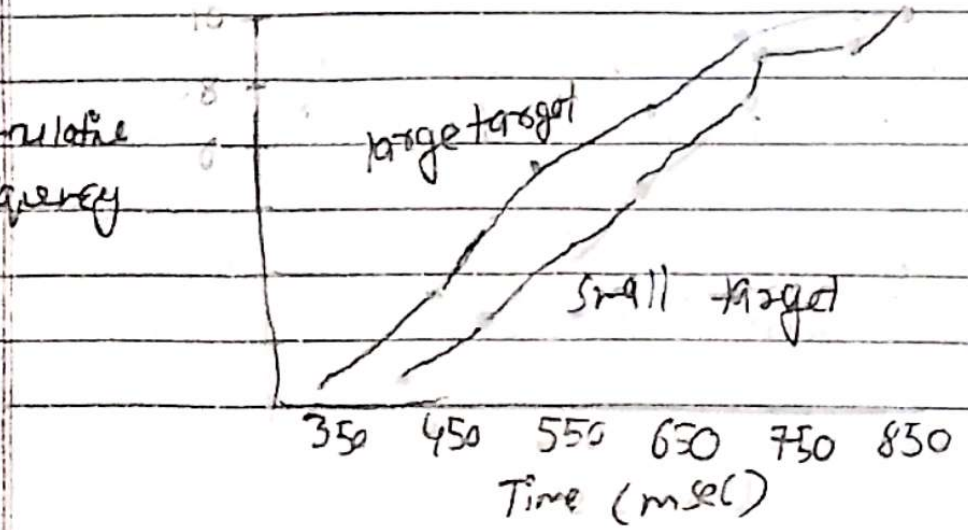
Two datasets Cursor Movement Speeds
small target (harder, takes longer)
large target (easier, takes less time)

A frequency polygon overlays both distributions.



The small target curve is shifted right, meaning longer reaction times.

Overlaid Cumulative frequency polygon



Date / /
Page No.

The small target polygon is below the large target polygon, indicating longer completion times for smaller targets.

Key Takeways

- Frequency polygons help visualize distributions
- Cumulative frequency polygons help analyze percentile rankings
- Overlaid frequency polygons make it easy to compare multiple datasets