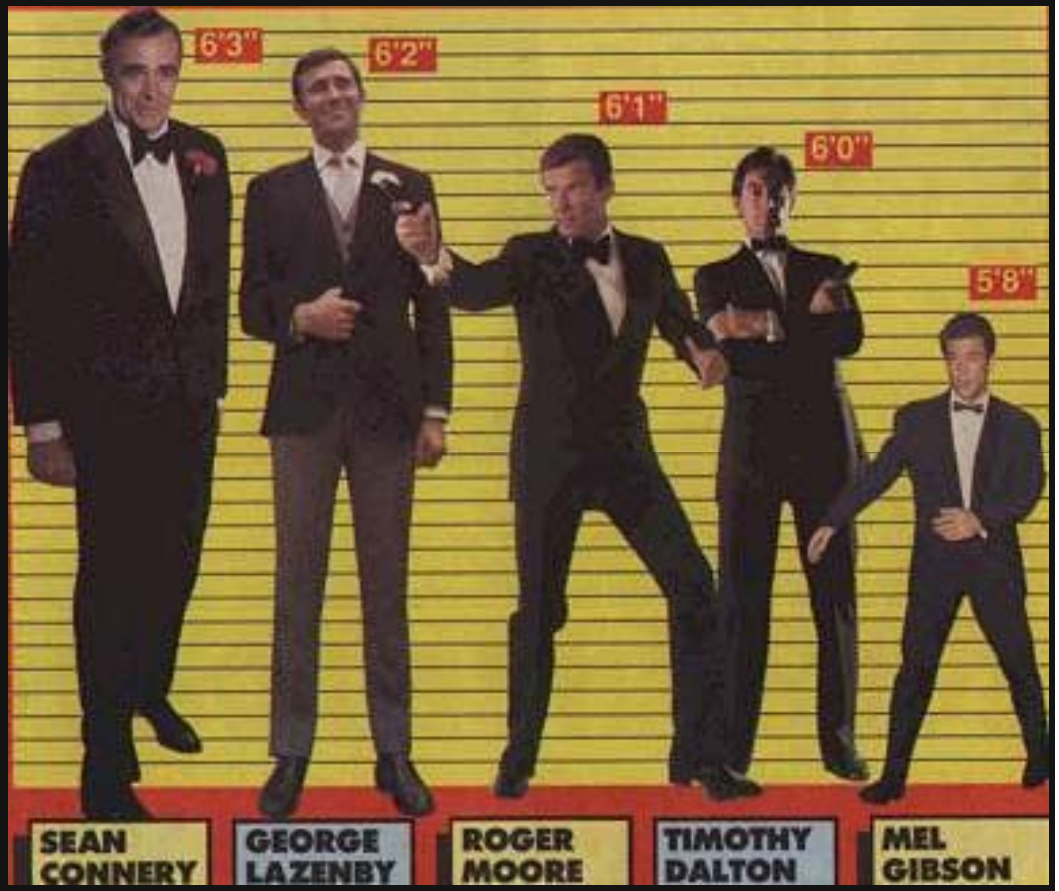
FREQUENCY DISTRIBUTIONS AND PERCENTILES



New Statistical Notation

- **Frequency (*f*):** the number of times a score occurs
- ***N*:** sample size

Simple Frequency Distributions

Raw Scores

- The scores that we have directly measured.
  - number of correct answers on a test
  - people's height
  - temperature measured during the day

Raw Scores

- Here is a data set of some raw scores:

14	14	13	15	11	15
13	10	12	13	14	13
14	15	17	14	14	15

14	14	13	15	11	15
13	10	12	13	14	13
14	15	17	14	14	15

How to construct a simple frequency table:

Score	*f*

15 16 13 16 15

17 16 15 17 15

Example 1

How to construct a simple frequency table:

Score	*f*

7 9 6

6 9 7

7 6 6

Example 2

How to construct a simple frequency table:

Score	*f*

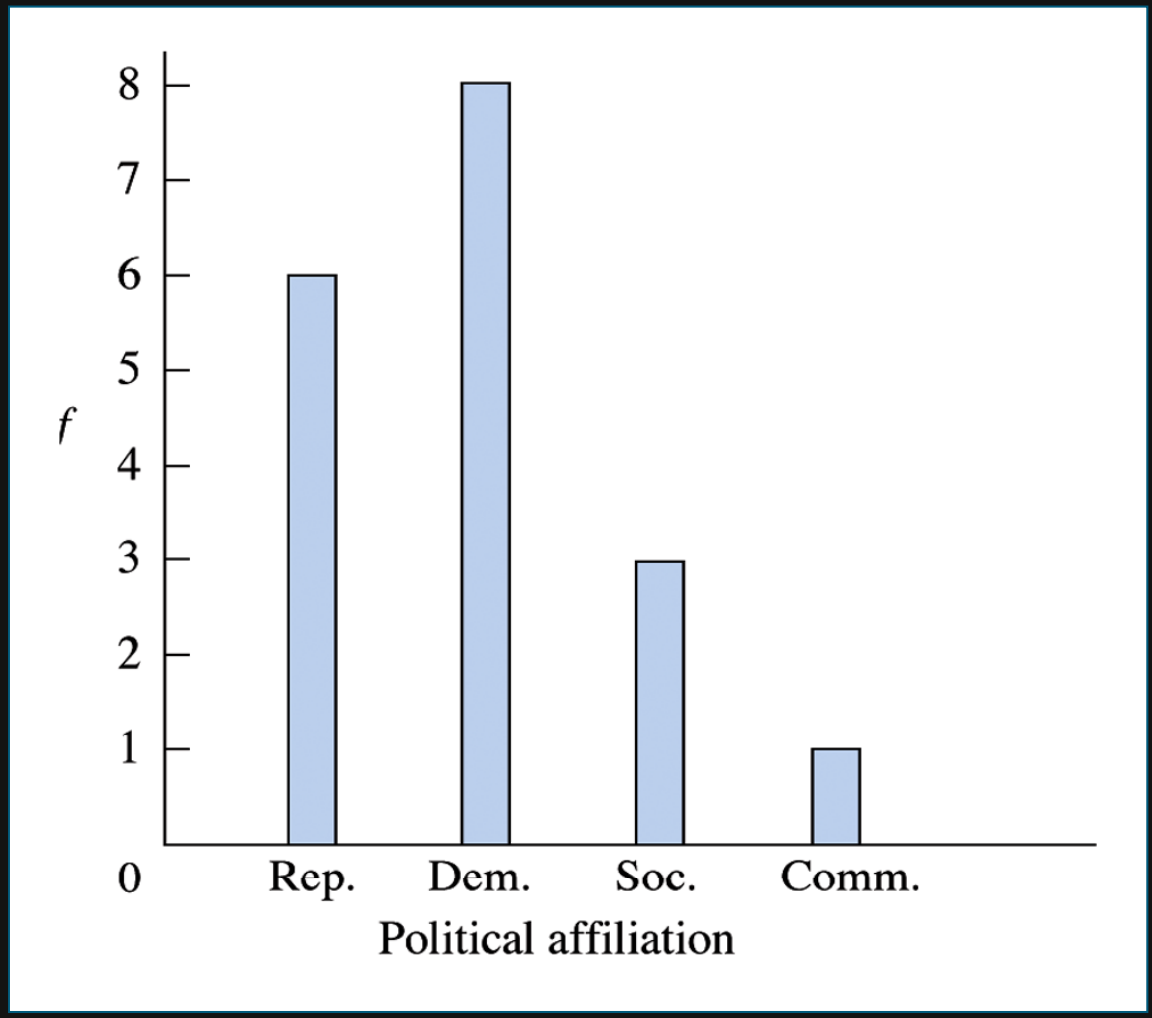
Graphing a Simple Frequency Distribution

- Scores on the X axis
- Frequencies on the Y axis
  - The type of measurement scale (nominal, ordinal, interval, or ratio) determines whether we use
  - A bar graph
  - A histogram
  - A polygon

Bar Graph

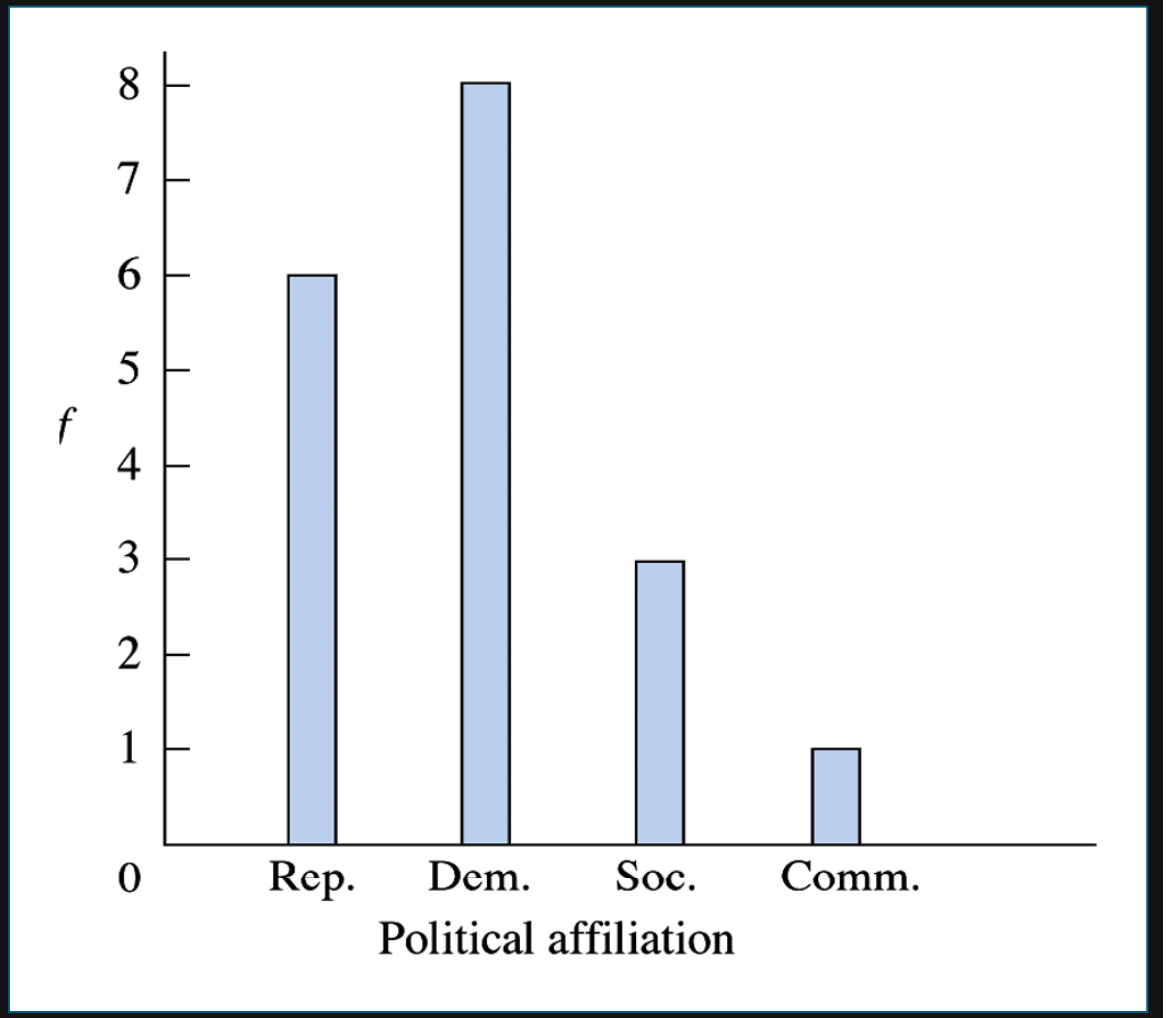
Used for:

- Nominal Data
  - Gender, marital stat.
- Ordinal Data
  - rank in class



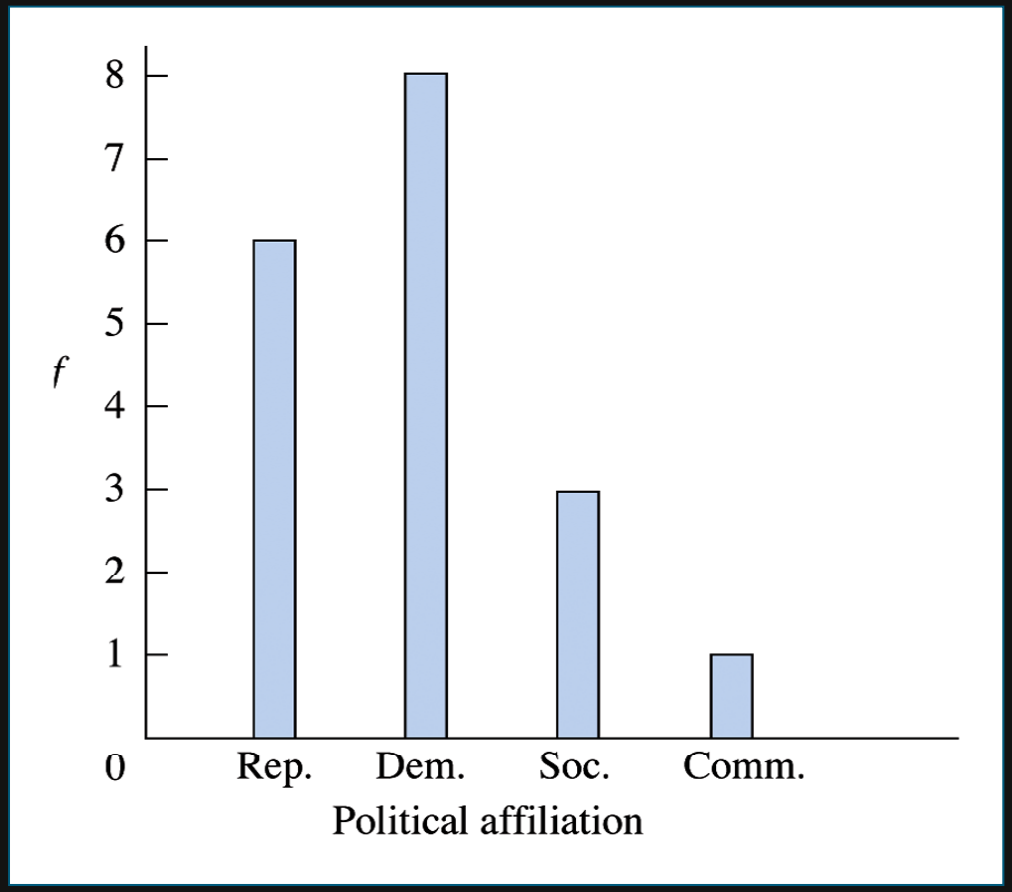
Bar Graph

- Adjacent bars do not touch
  - Why?



Bar Graph

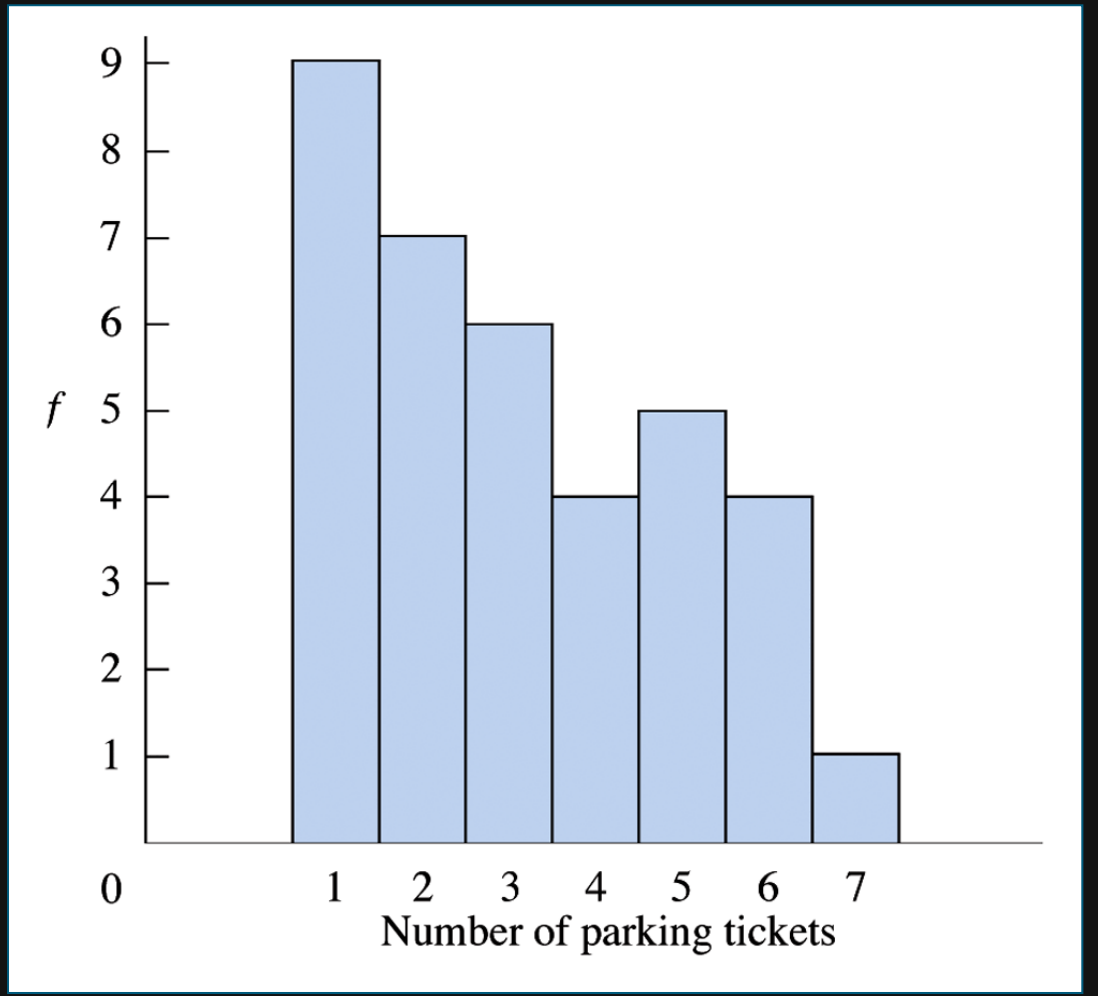
Party	*f*
Rep.	
Dem.	
Soc.	
Com.	



Histogram

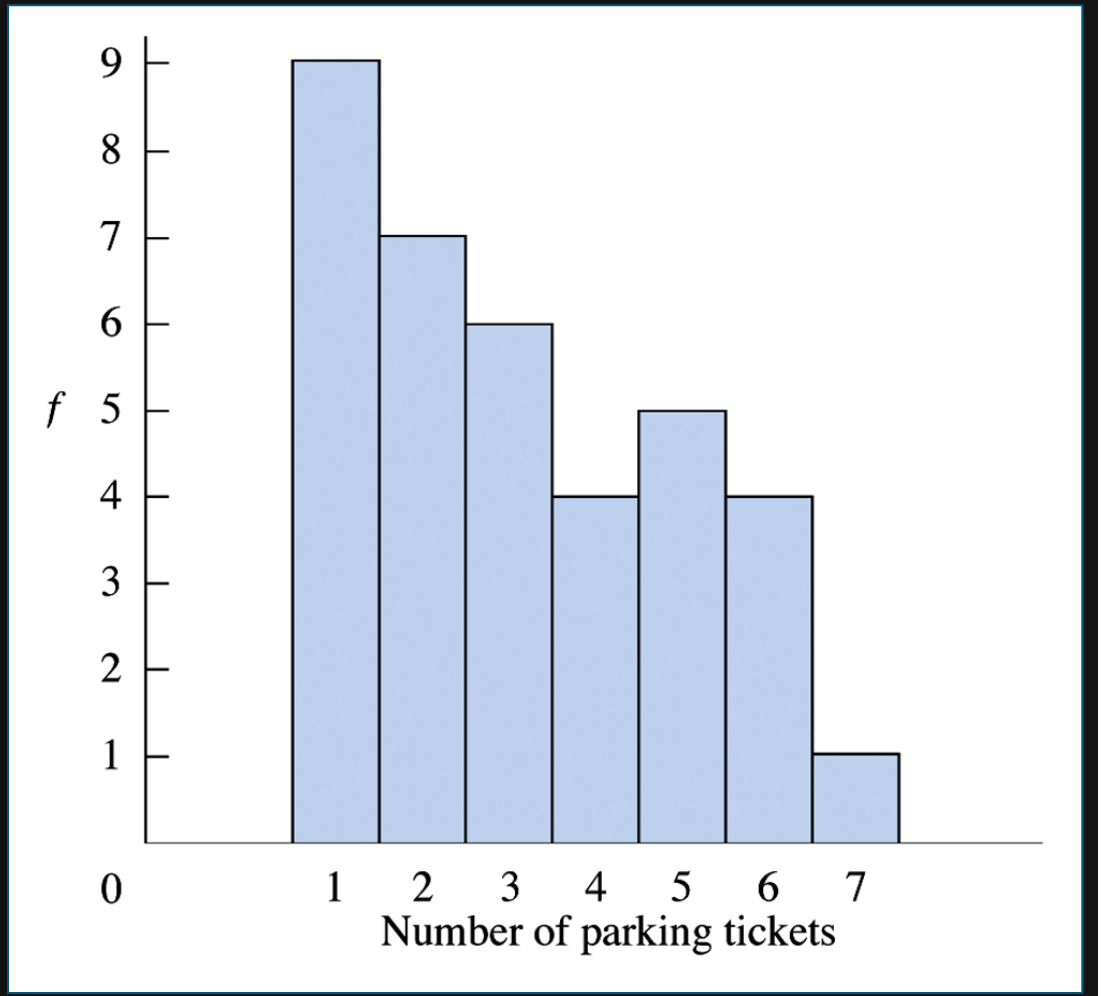
Used for:

- Interval scores
  - IQ, temperature…
- Ratio scores
  - Weight, height, time…



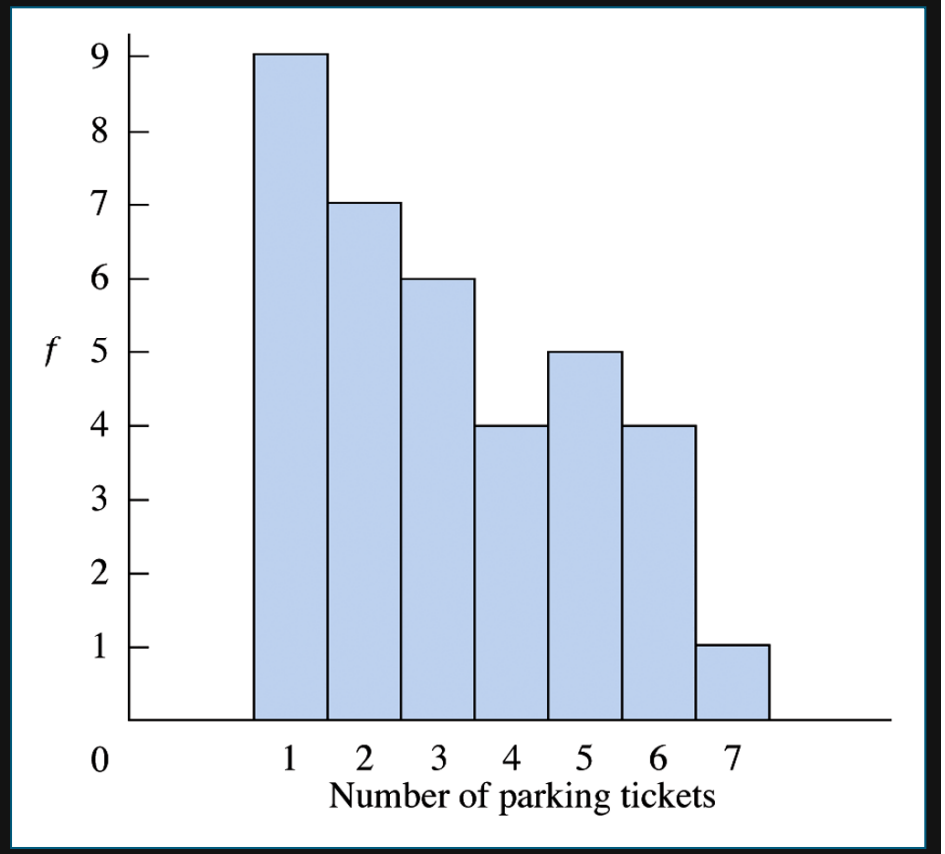
Histogram

- Adjacent bars touch
  - Why?

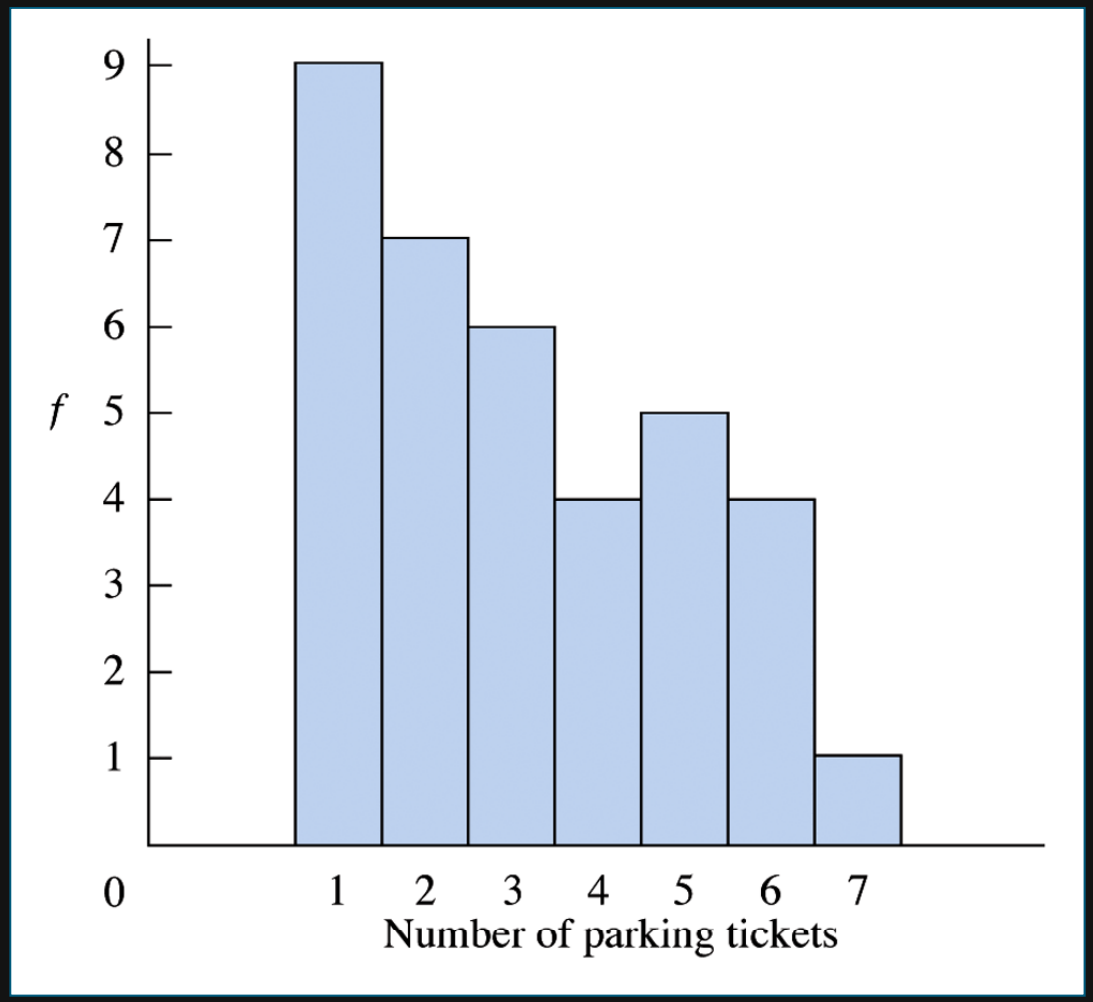


Histogram

Score	*f*
7	
6	
5	
4	
3	
2	
1	

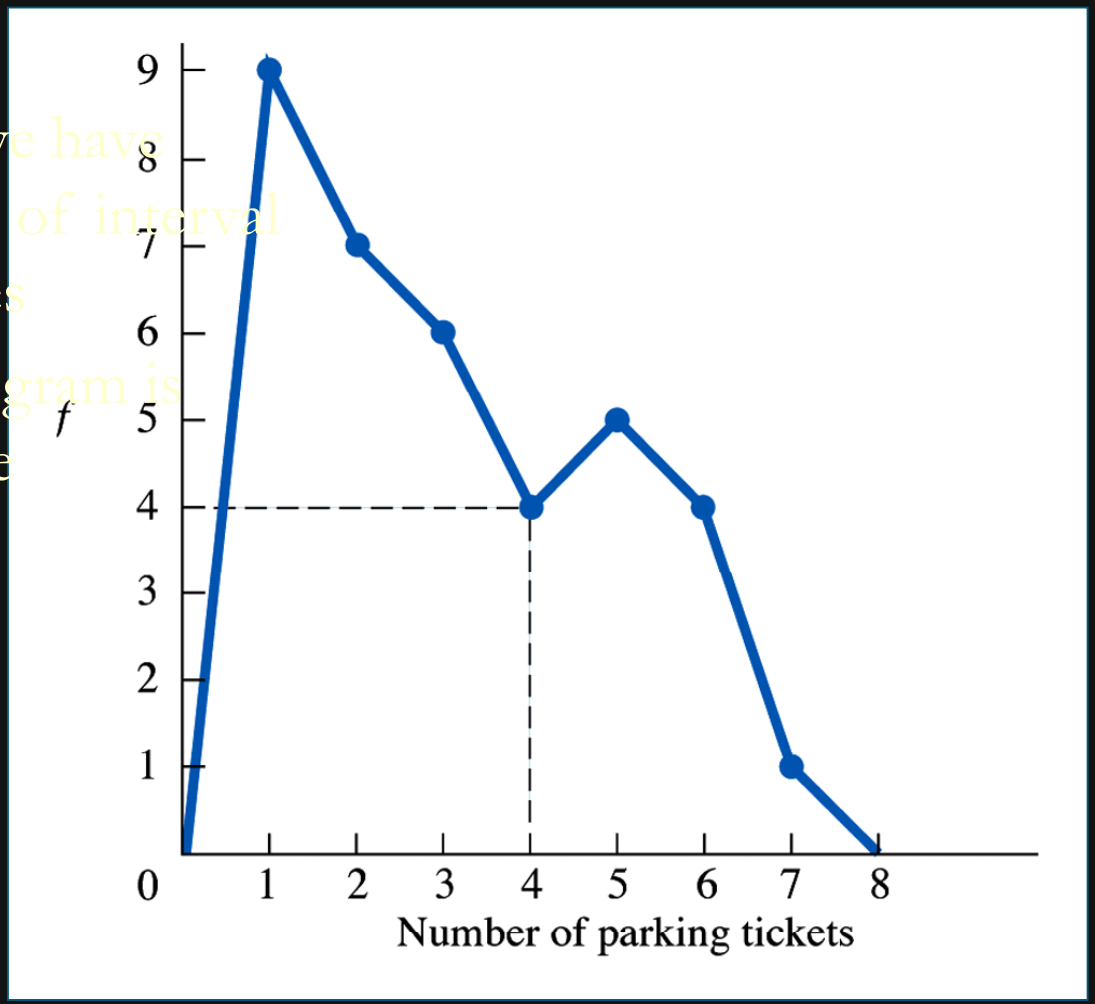


Drawing a Polygon



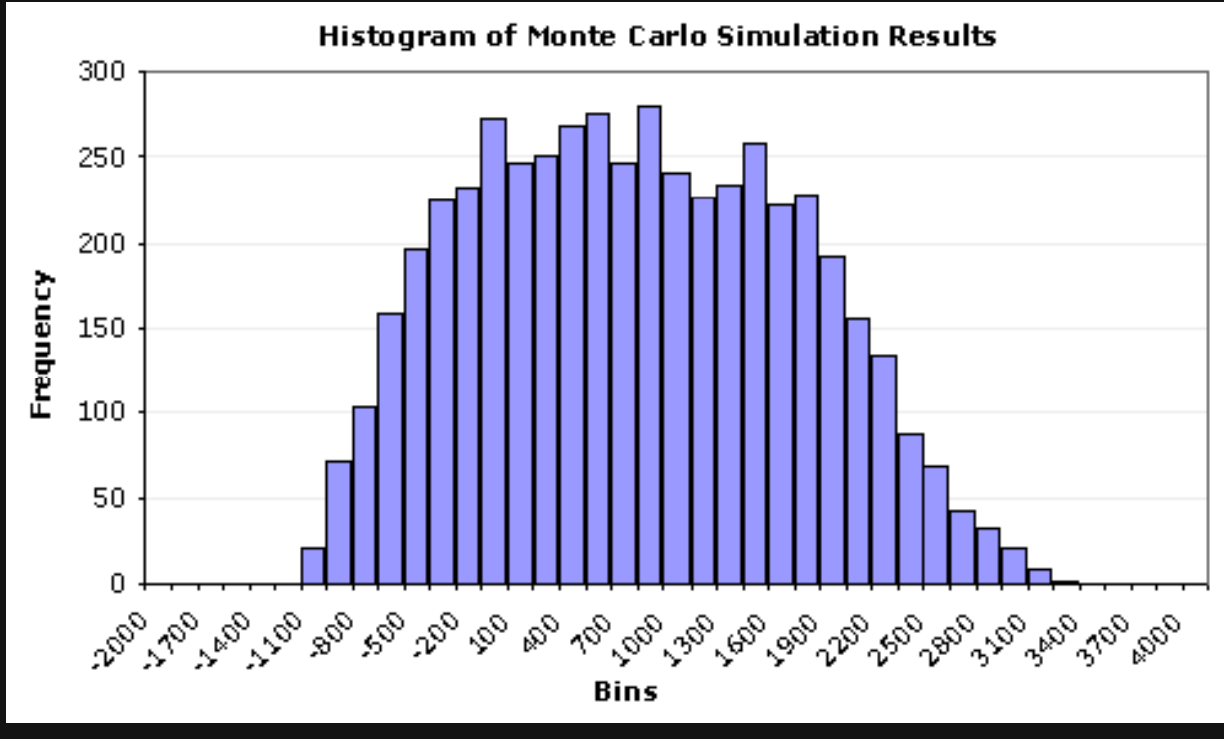
Polygon

- Used when we have large number of interval or ratio scores
- When a histogram is hard to create



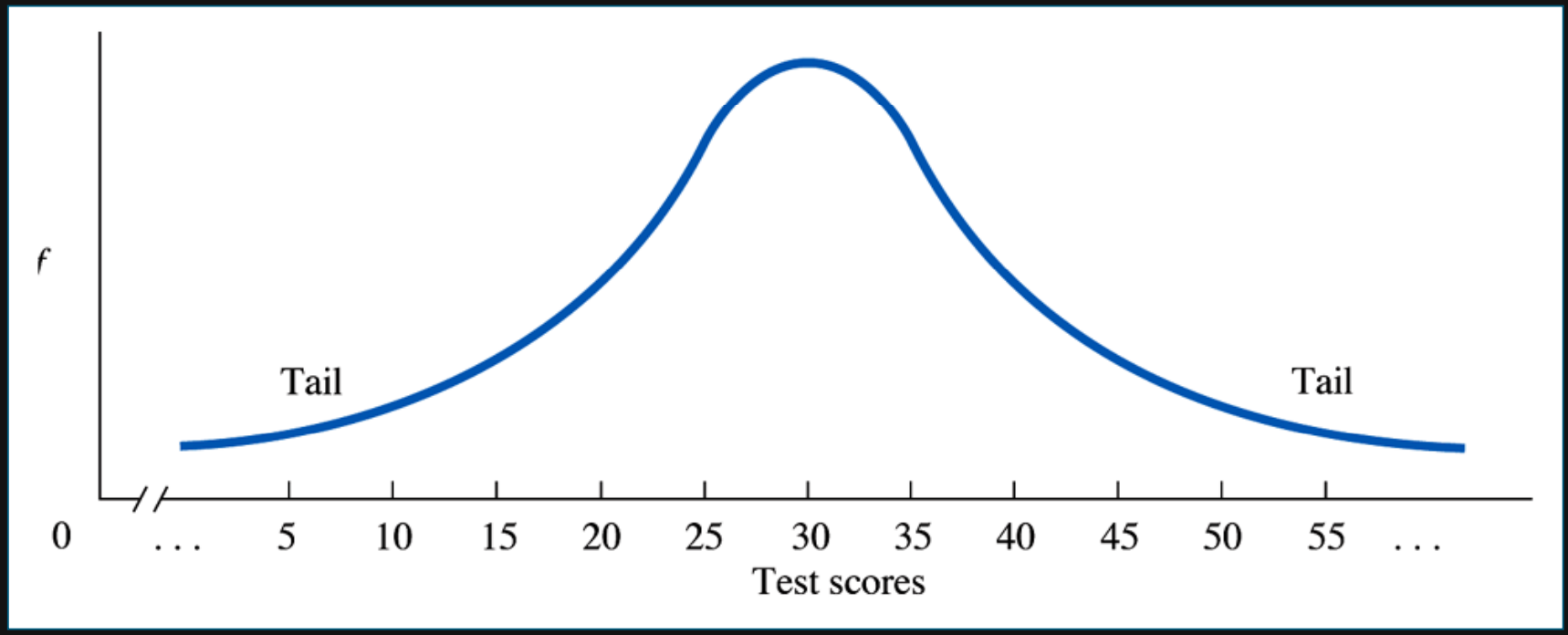
Histogram vs. Polygon

- It is easier to draw a polygon when you have a lot of scores
- Bars get thinner and thinner



Types of Simple Frequency Distributions

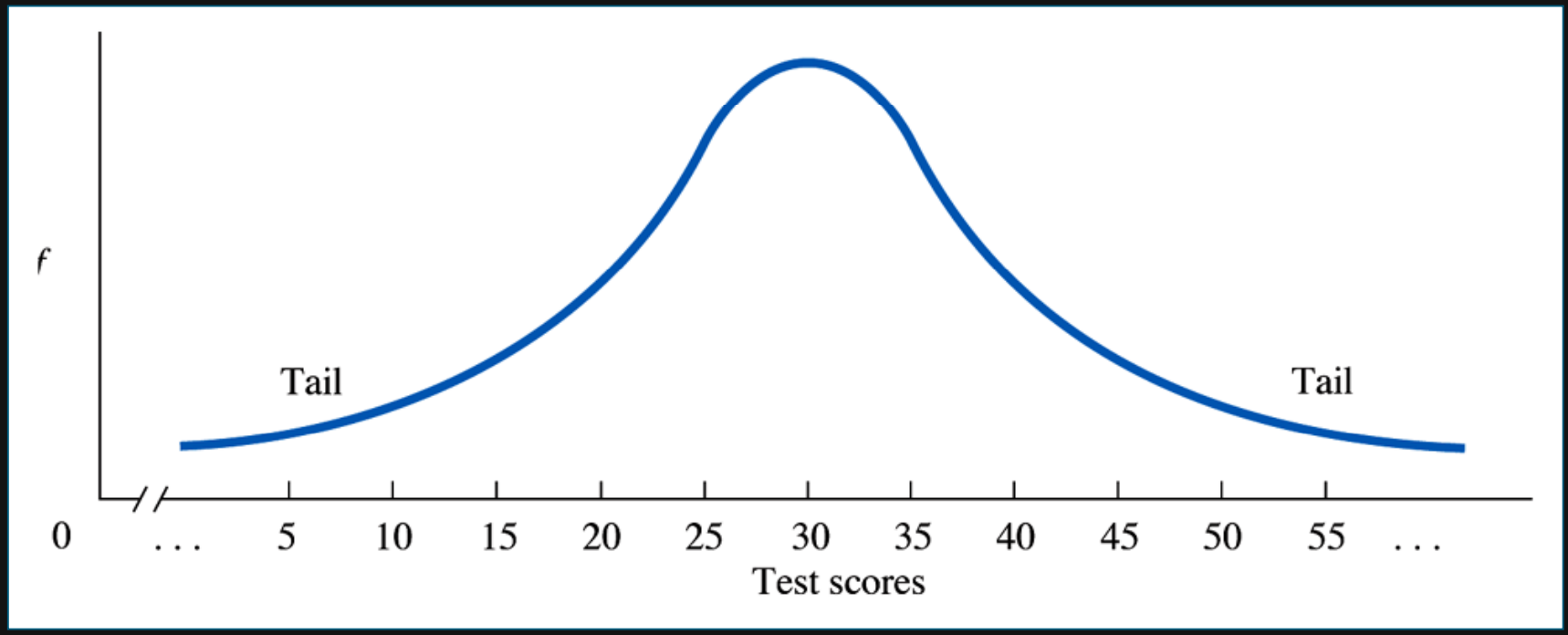
Normal Distribution



The Normal Distribution

- A bell-shaped curve
- Called the **normal curve** or a **normal distribution**
- It is *symmetrical*
- The far left and right portions containing the low-frequency extreme scores are called the *tails* of the distribution

Normal Distribution



The Normal Distribution

- Most variables are normally distributed in the population
  - IQ
  - Height of females
  - Height of males

Non-normal Distributions

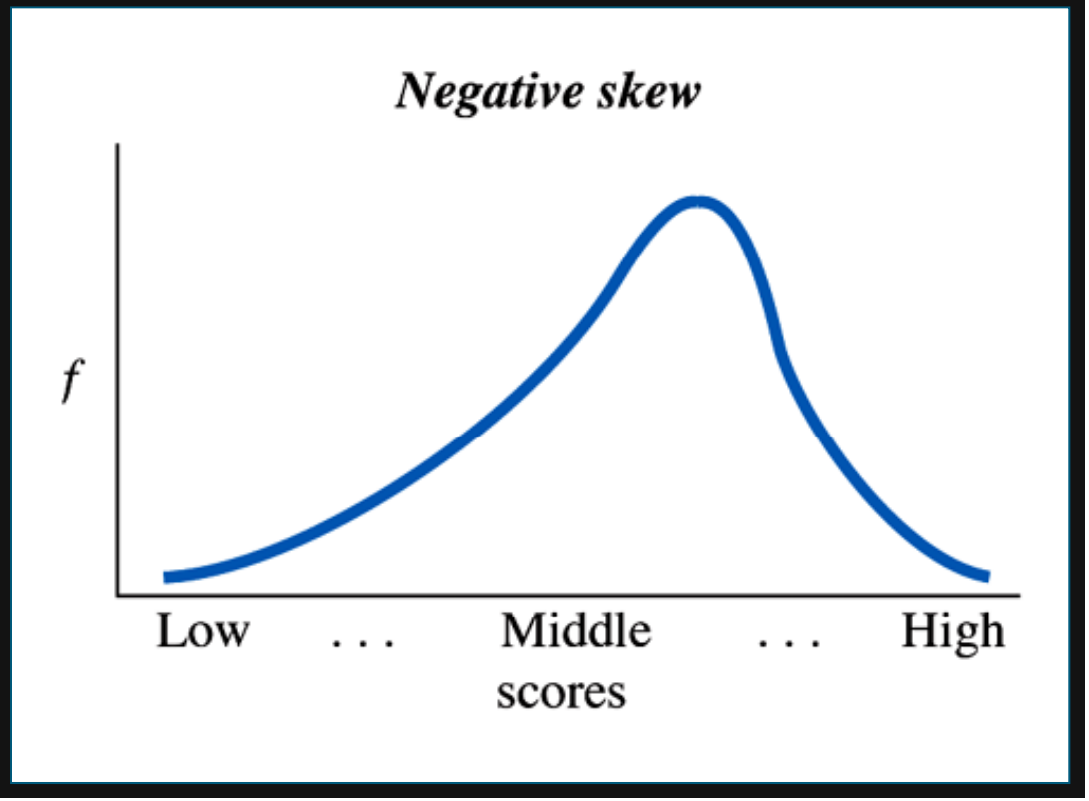
- Skewed
- Bimodal
- Rectangular

Skewed Distributions

- Not *symmetrical*
- A distribution may be either *negatively skewed* or *positively skewed*

Negatively Skewed Distribution

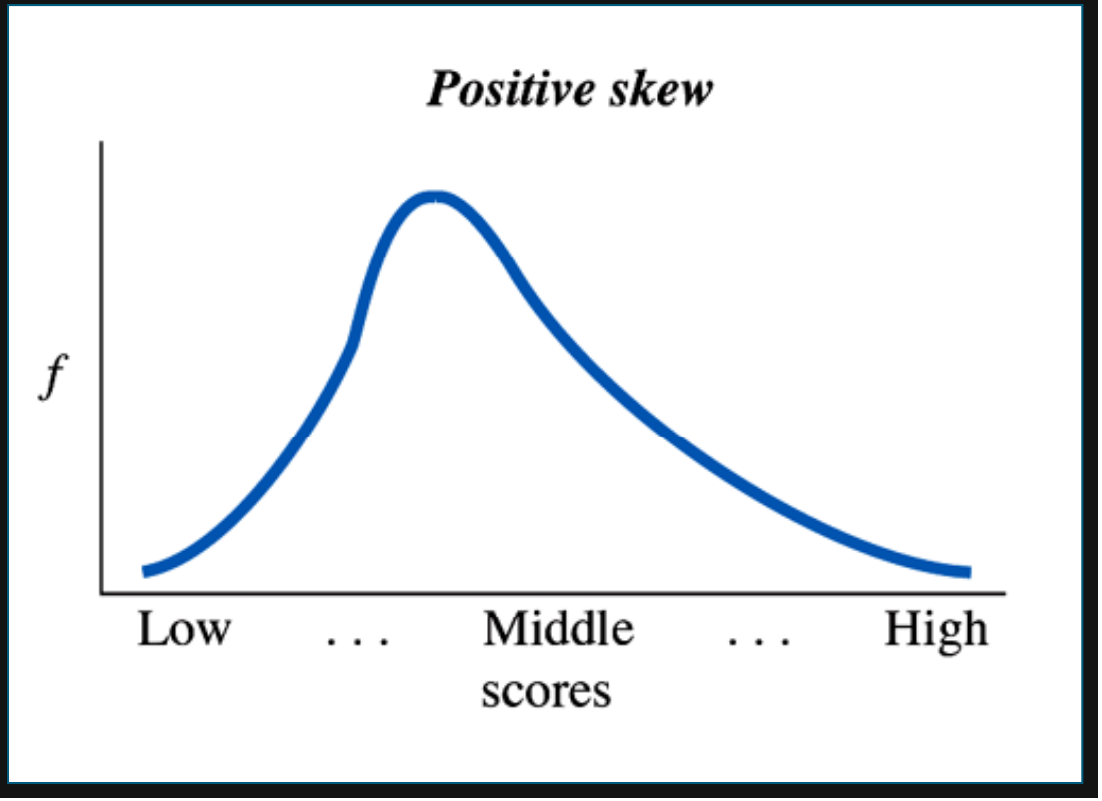
- Has extreme low scores that have a low frequency
- Does not have extreme high scores that have low frequency



Positively Skewed Distribution

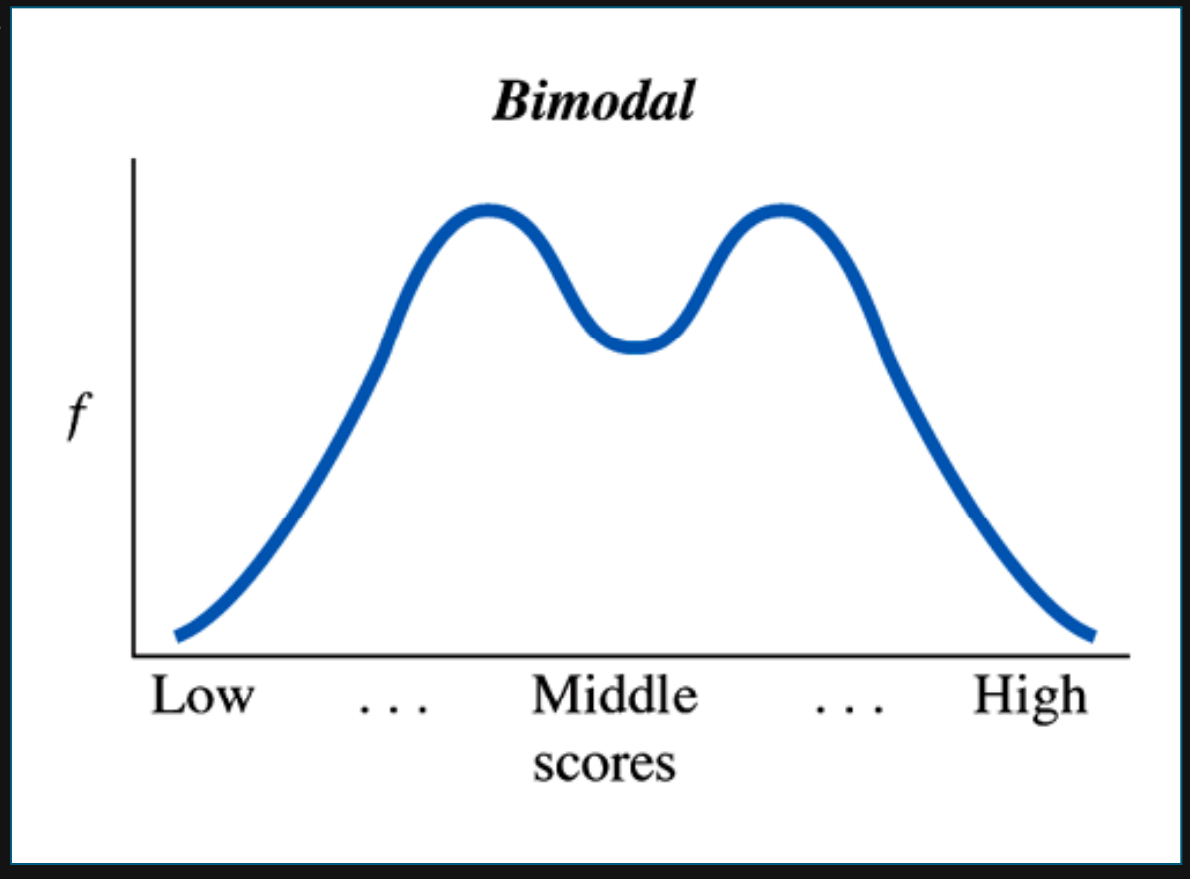
- Has extreme high scores that have a low frequency

- Does not have extreme low scores that have low frequency



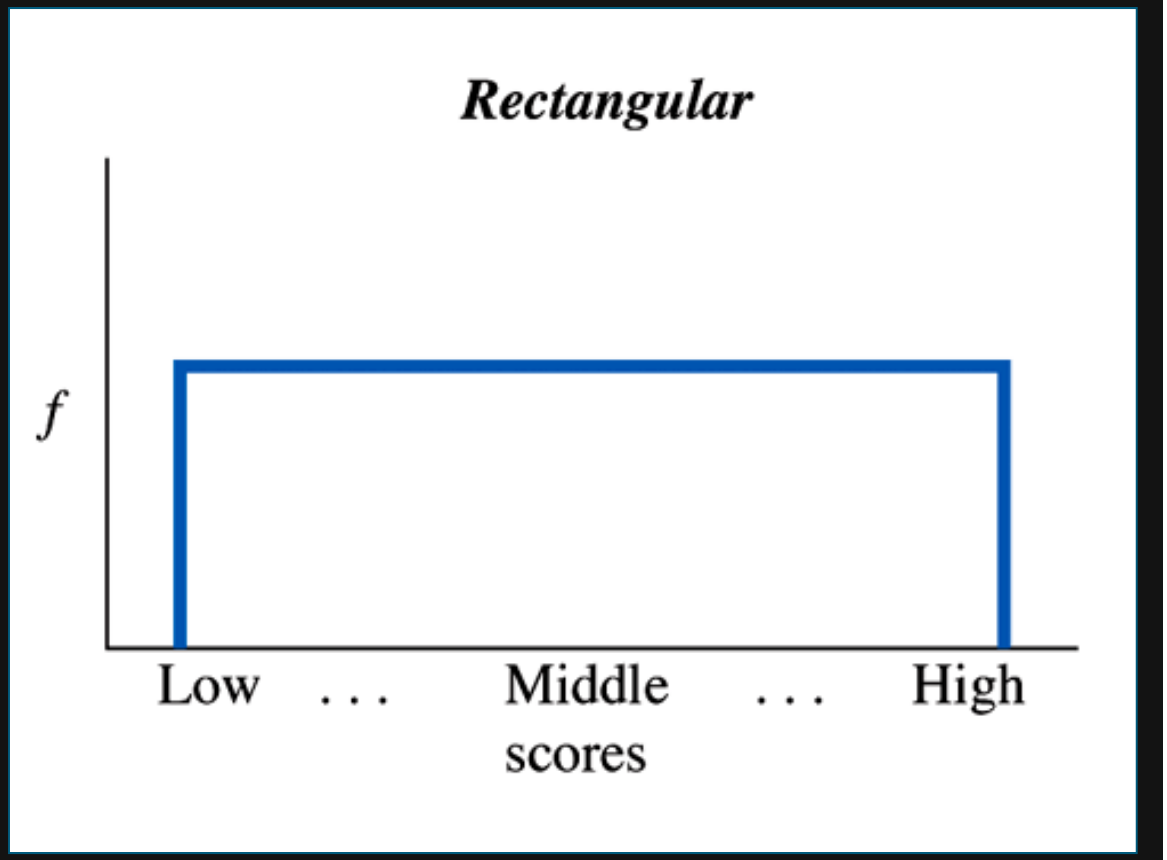
Bimodal Distribution

A symmetrical distribution containing two distinct humps

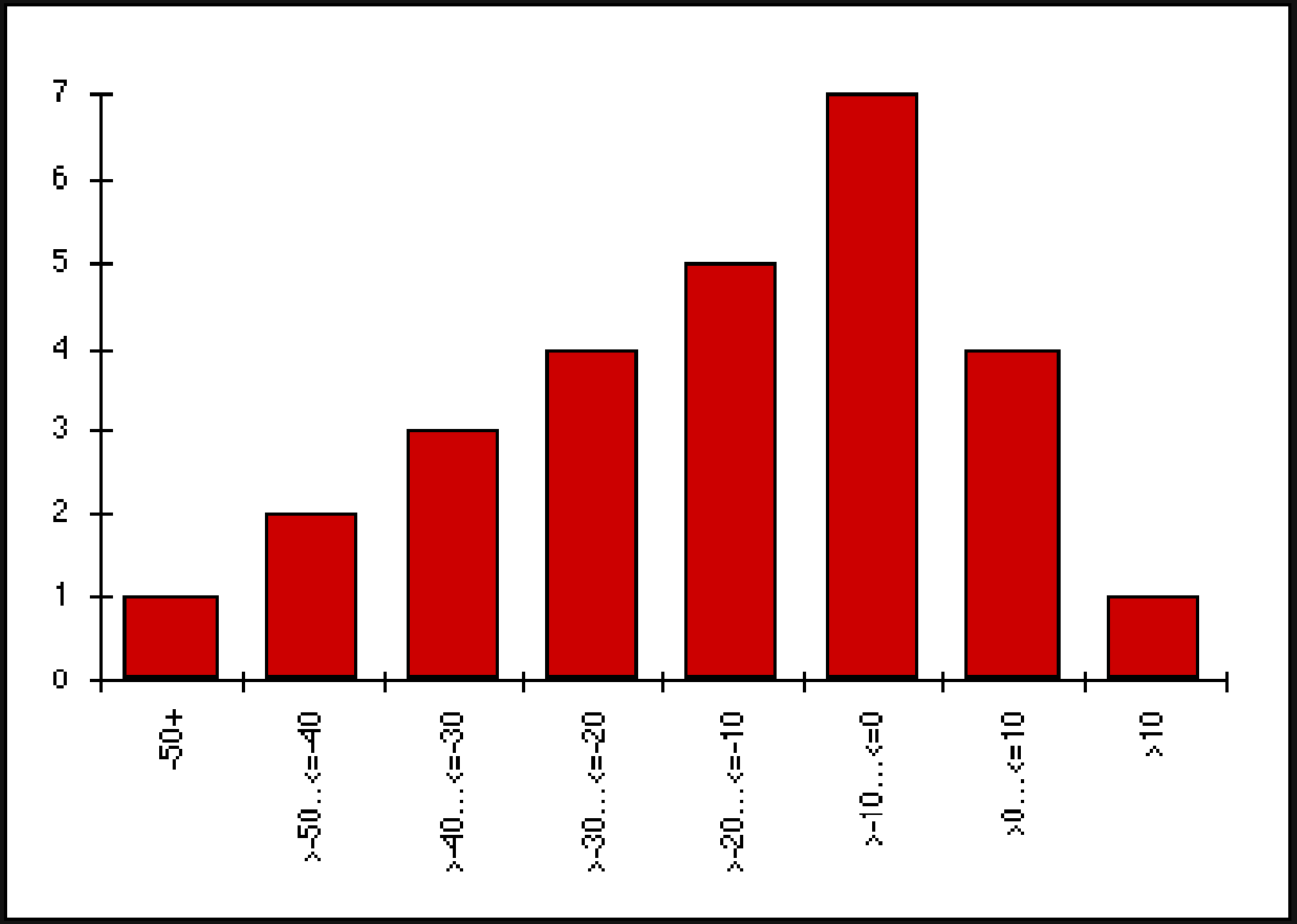


Rectangular Distribution

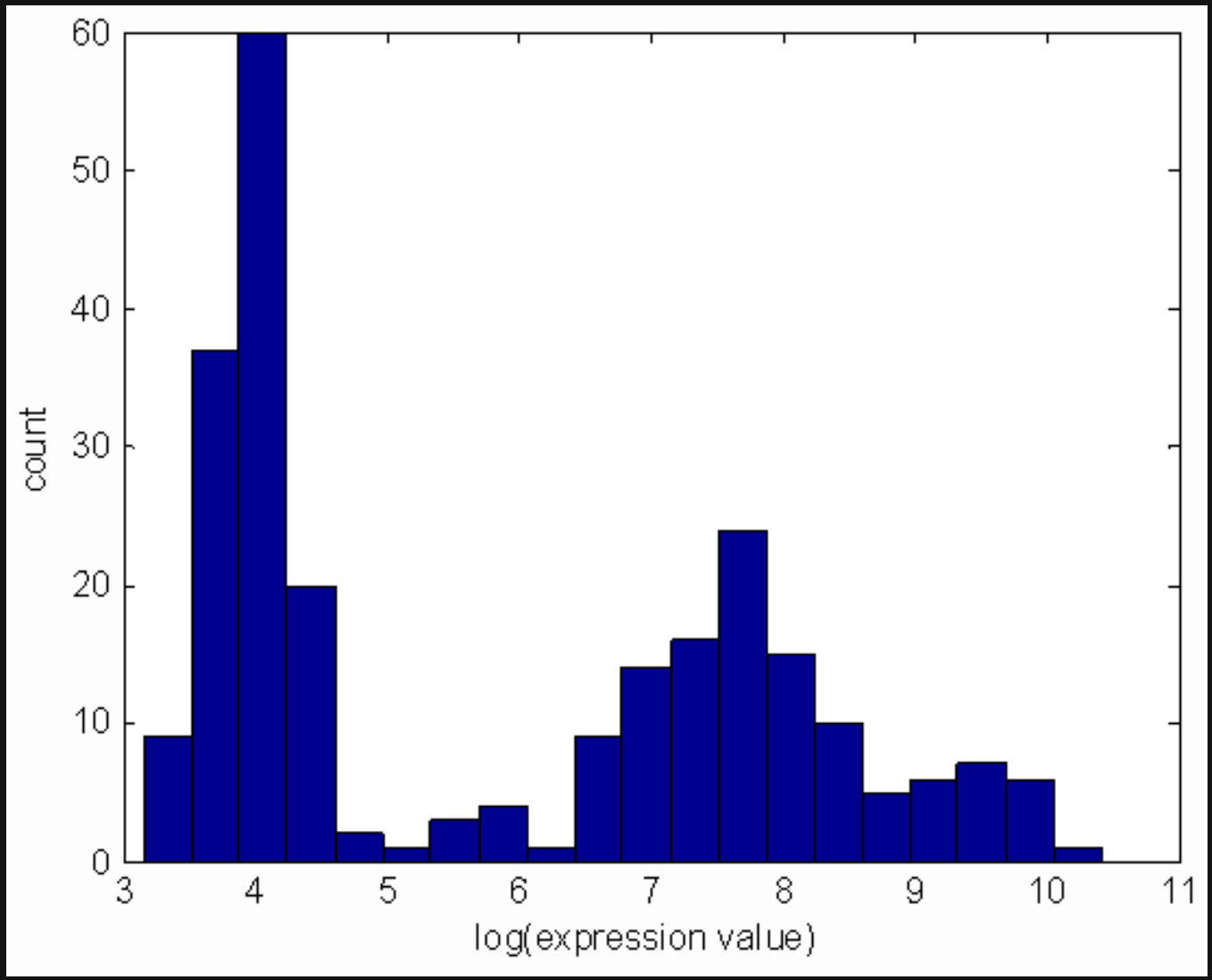
A symmetrical distribution shaped like a rectangle



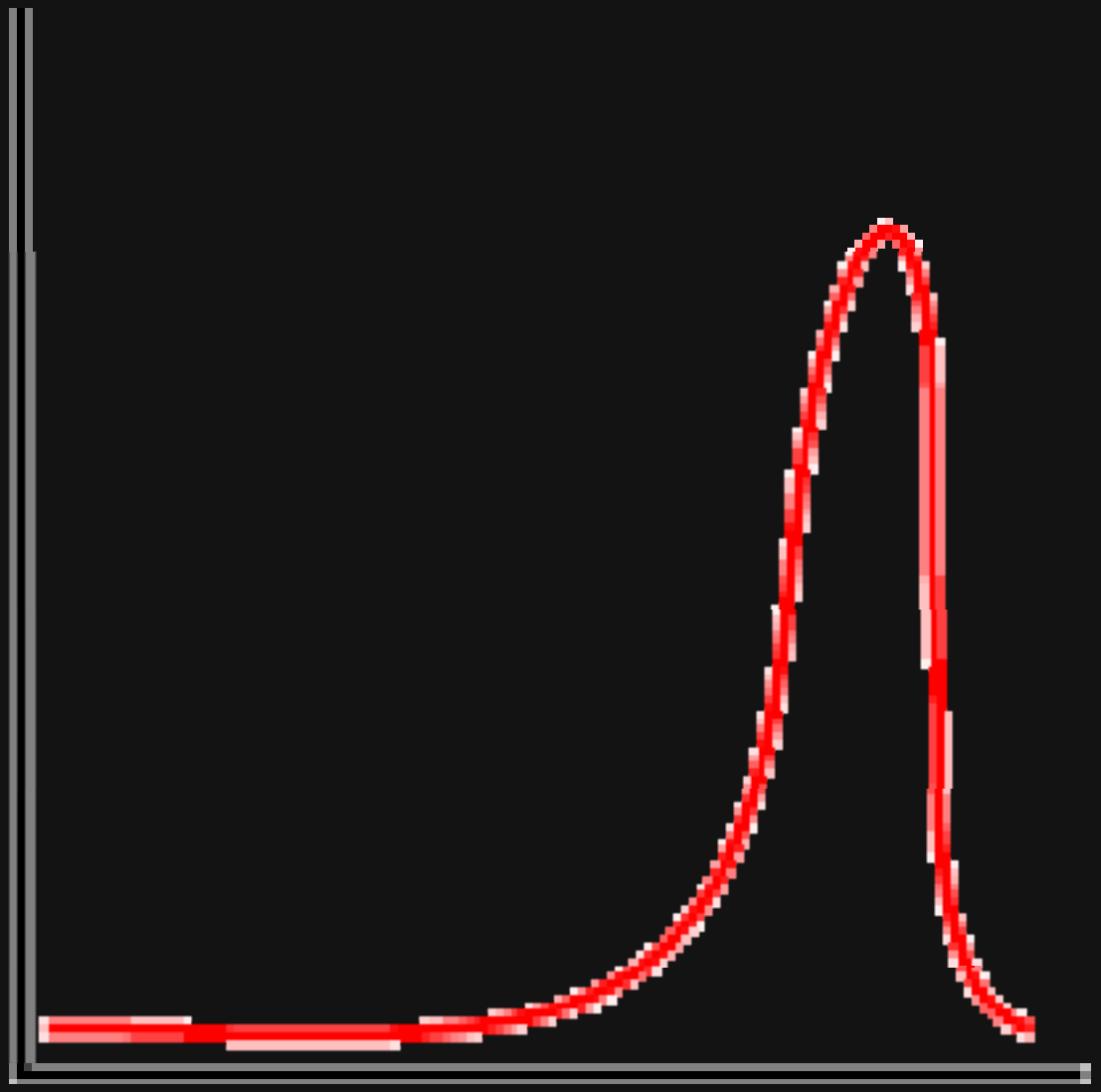
Examples



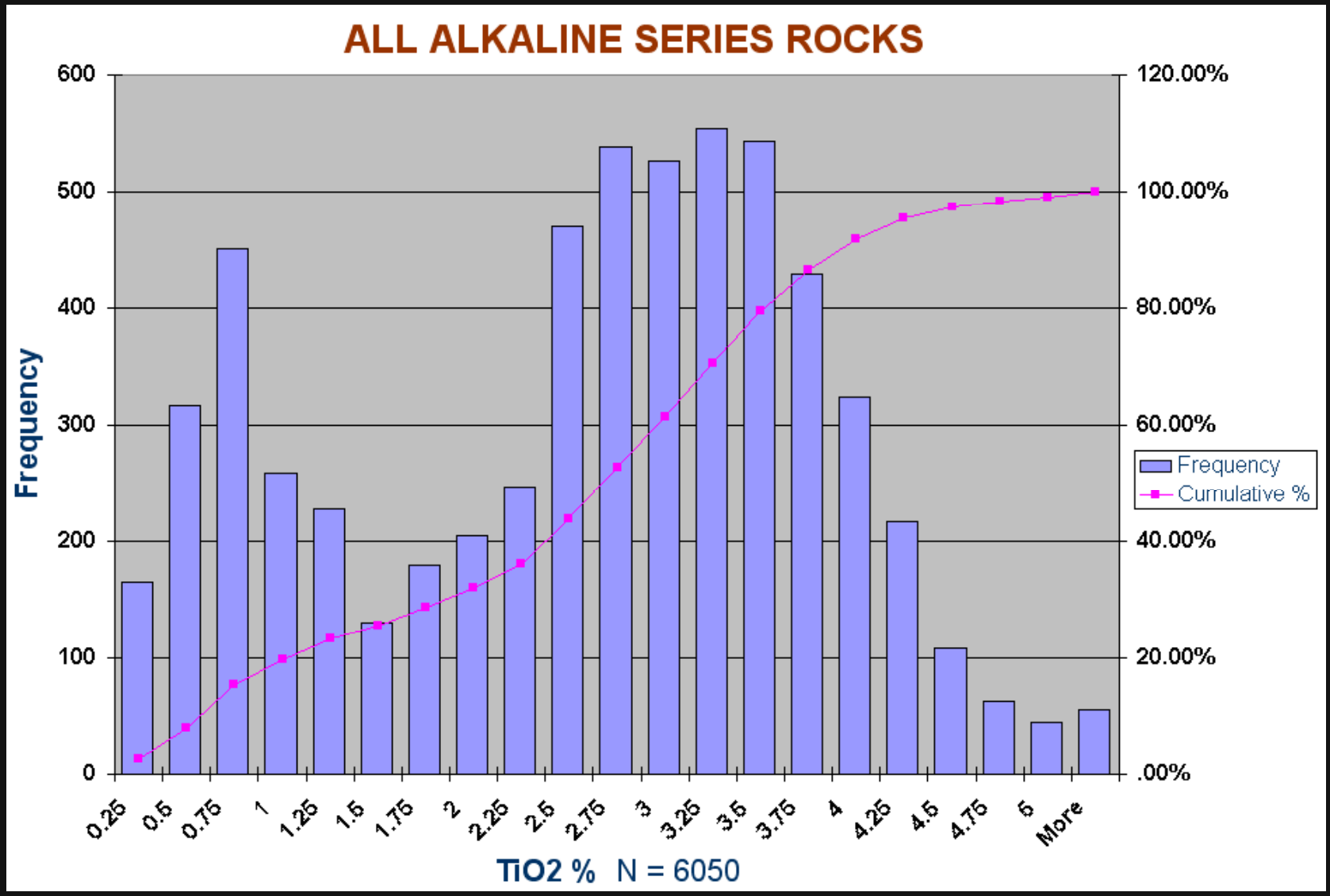
Examples



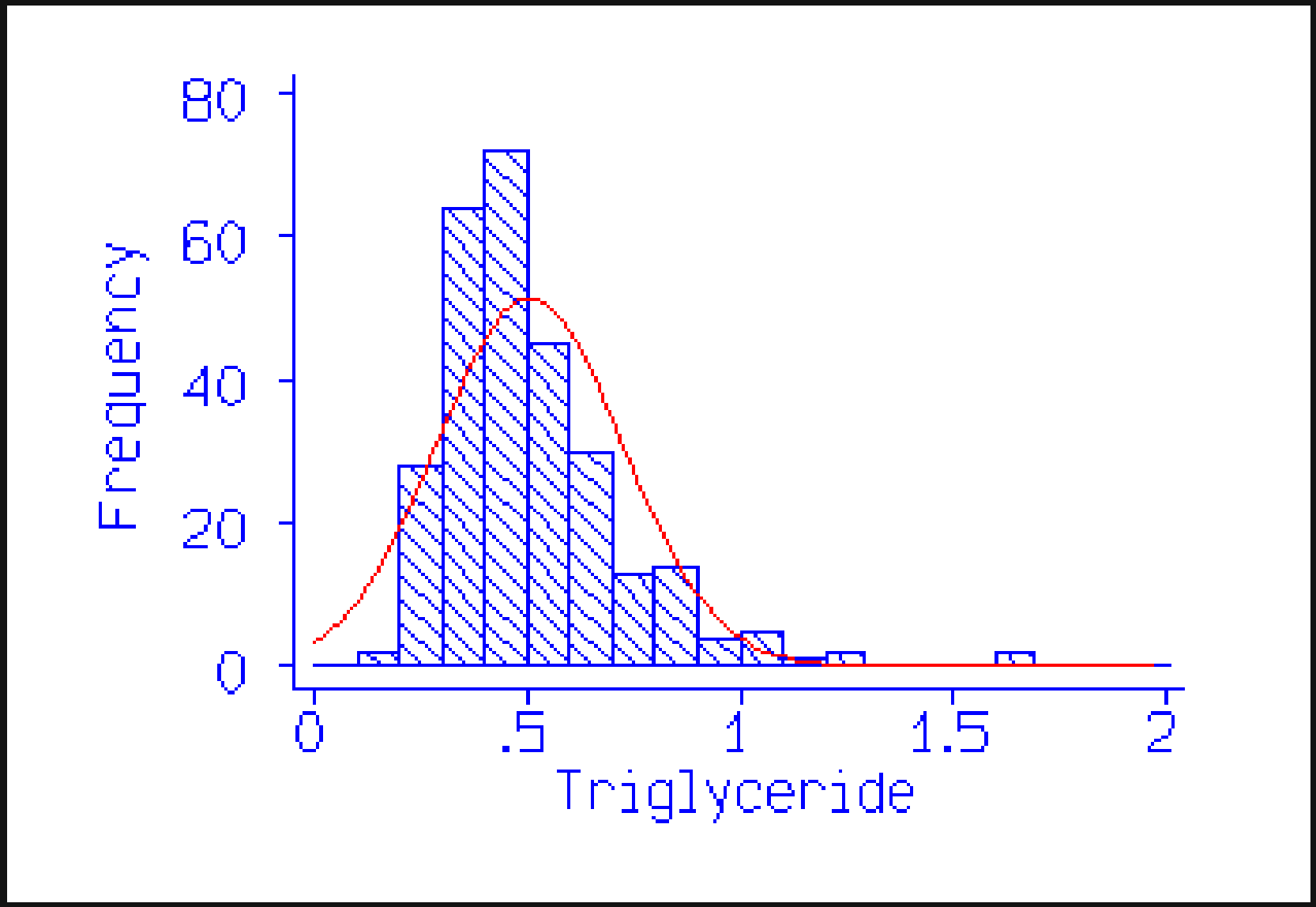
Examples



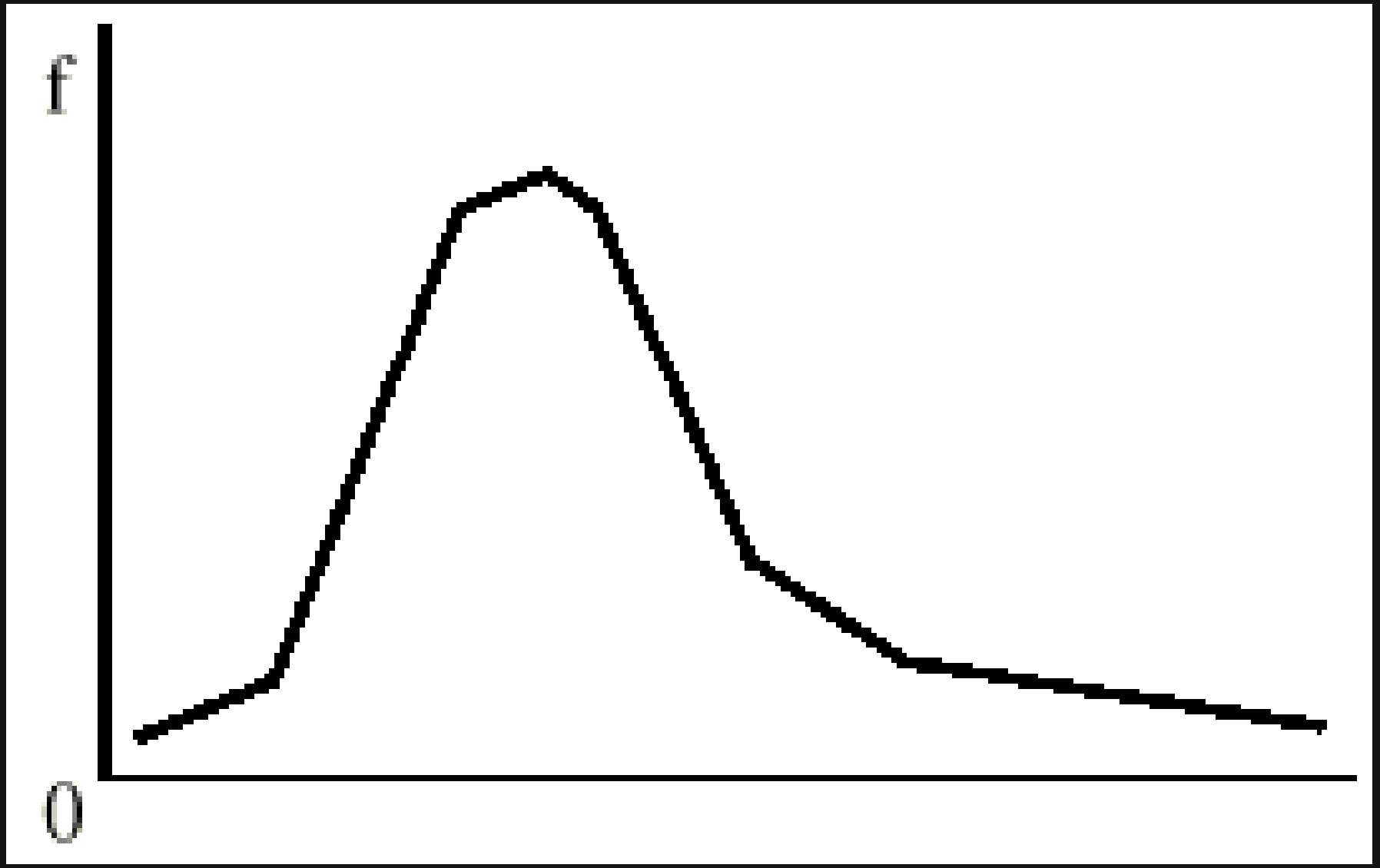
Examples


ALL ALKALINE SERIES ROCKS
Frequency
600
500
400
300
200
100
0
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
.00%
Frequency
Cumulative %
0.25
0.5
0.75
1
1.25
1.5
1.75
2
2.25
2.5
2.75
3
3.25
3.5
3.75
4
4.25
4.5
4.75
5
More
TiO2 % N = 6050

Examples



Examples



Relative Frequency and the Normal Curve

Relative Frequency

- *Relative frequency* ***(rel. f)*** *:* the proportion of time the score occurs
- $0 \leq rel.\ f \leq 1$
- The formula is:

$$rel.\ f = \frac{f}{N}$$

Relative Frequency

- Gives us a frame of reference.
  - It is easier to interpret
- **For exp:**
  - In an exam,7 students got 100.
  - Is the class successful?
    - Depends on how many students there are (N)
  - If N is 14, then *rel. f* = 7/14=0.5

Relative Frequency Table

Score	*f*	***Rel. f***
4	6	
3	8	
2	3	
1	1	

N=18

A Relative Frequency Distribution

The left-hand column identifies the scores, the middle column shows each score's frequency, and the right-hand column shows each score's relative frequency.

Score	***f***	***rel. f***
6	1	
5	0	
4	2	
3	3	
2	10	
1	4	
	Total: 20	1.00 = 100%

Example 1

Relative Frequency Tables

Score	*f*	*Rel. f*
7	1	
6	4	
5	5	
4	4	
3	6	
2	7	
1	9	

N=36

Cumulative Frequency and Percentile

Cumulative Frequency

- *Cumulative frequency* ***(cf)***: the frequency of all scores at or below a particular score
- To compute a score's cumulative frequency, we add the simple frequencies for all scores below the score with the frequency for the score

Cumulative Frequency Table

Score	*f*	*cf*
4	6	
3	8	
2	3	
1	1	

N=18

A Cumulative Frequency Distribution

The left-hand column identifies the scores, the center column contains the simple frequency of each score, and the right-hand column contains the cumulative frequency of each score.

Score	*f*	*cf*
17	1	
16	2	
15	4	
14	6	
13	4	
12	0	
11	2	
10	1	

Percentile

- ***Percentile:*** the percent of all scores in the data that are at or below the score
- Formula:

Percentile = (cf/N)*100

Percentile

Score	*f*	*cf*	*percentile*
4	6	18	
3	8	12	
2	3	4	
1	1	1	

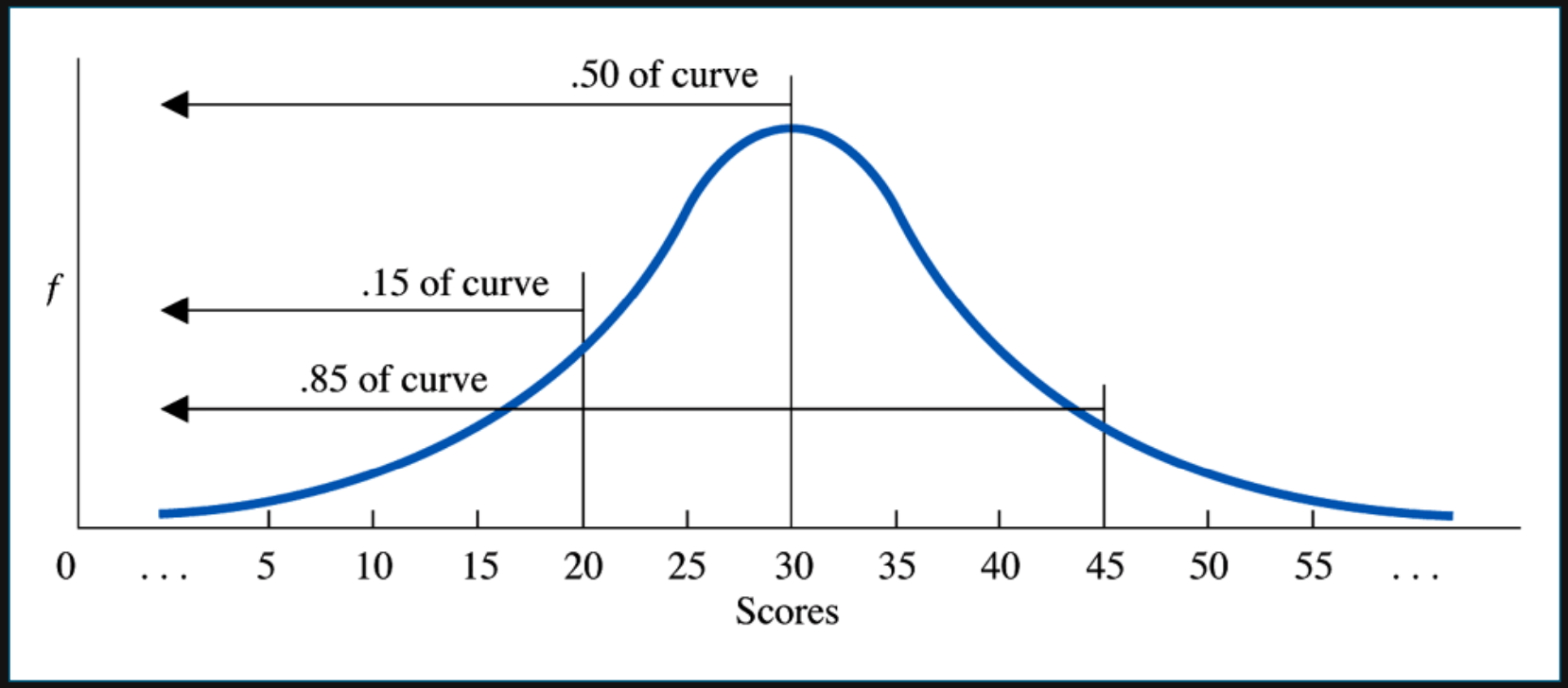
N=18

Finding Percentiles in Graphs

The percentile for a given score corresponds to the percent of the total area under the curve that is to the *left* of the score.

Percentiles

Normal distribution showing the area under the curve to the left of selected scores.



Grouped Frequency Distributions

Grouped Distributions

Grouped distribution: scores are combined to form small groups

– we report the total *f*, *rel. f*, or *cf* of each group

A Grouped Distribution

Score	*f*	*rel. f*	*cf*	*Percentile*
40–44	2	.08	25	
35–39	2	.08	23	
30–34	0	.00	21	
25–29	3	.12	21	
20–24	2	.08	18	
15–19	4	.16	16	
10–14	1	.04	12	
5–9	4	.16	11	
0–4	7	.28	7	

An Example: Grouped Frequency Distribution

- Record the limits of all class intervals, placing the interval containing the highest score value at the top.
- Count up the number of scores in each interval.

Hotel Rates	Frequency
800-899	1
700-799	4
600-699	2
500-599	0
400-499	6
300-399	8
200-299	8
100-199	4
0-99	2

Las Vegas Hotel Rates

52	205	282	325	417	732
76	250	283	373	422	749
100	257	303	384	472	750
136	264	313	384	480	791
186	264	317	400	643	891
196	280	317	402	693	

Frequency Table Guidelines

- Intervals should not overlap, so no score can belong to more than one interval.
- Make all intervals the same width.
- Make the intervals continuous throughout the distribution (even if an interval is empty).
- Place the interval with the highest score at the top.
- Choose a convenient interval width.

Hotel Rates	Frequency
800-899	1
700-799	4
600-699	2
500-599	0
400-499	6
300-399	8
200-299	8
100-199	4
0-99	2

Example 1

- Using the following data set, find the relative frequency of the score 12

14	14	13	15	11	15
13	10	12	13	14	13
14	15	17	14	14	15

Example 2

- What is the cumulative frequency for the score of 14?

Example 3

- What is the percentile for the score of 14?

Example 4

- Data set: 1,4,5,3,2,5,7,3,4,5.
- Find the mistakes below.

Score	f	cf
1	1	0
2	1	1
3	2	3
4	2	5
5	3	7
7	1	9

N=6

Example 5

- Organize the scores below in a table showing
  - Simple frequency
  - Relative frequency
  - Cumulative frequency
- Draw a simple frequency histogram
- Draw a simple frequency polygon

49 52 47 52 52 47 49 47 50

51 50 49 50 50 50 53 51 49