Experimental Frequency Distributions

1- Construction Mechanism

- Construct a frequency table
  - Divide the data into number of classes
  - List number of observations in each class (i.e. frequency of observations)
  - Relative frequency = Frequency of each class / Total number of observations
  - % Frequency = Relative frequency X 100
  - Cumulative frequency = Adding frequencies as moving down the frequency table
  - % Cumulative frequency = Adding % frequencies as moving down the frequency table.

Experimental Frequency Distributions

2- Graphing Mechanism

- Graphical representation of data
  - Histogram or bar graph
  - Frequency curve
  - Relative frequency curve
  - Cumulative frequency bar graph
  - Cumulative frequency curve
  - % Cumulative frequency curves

- The proportion (probability of occurrence) of sample values or values that are smaller than a given value can be directly read from the above mentioned curves.

Experimental Frequency Distributions

3- Example

- Consider the following data set of mineralization percentages:

8.1	24.9	32.2	36.9	41.3	47.6	54.5	74
12.8	26.8	33.4	37.3	41.5	48.5	54.6	77.3
14.3	27.1	33.6	37.4	41.7	49.4	56.1	
14.9	27.4	33.9	37.9	42.4	49.7	58.2	
15.7	28.2	34.1	38.5	42.9	50.1	59.3	
19.3	29.5	34.7	38.9	43.5	50.9	59.5	
20.3	29.2	35.1	39.1	43.6	51.5	59.8	
21.7	30.1	35.2	40.2	45.6	51.8	63.8	
22.6	31.8	35.6	40.8	46.1	52.7	64.1	
24.8	31.9	36.5	40.9	46.8	53.7	64.8	

Experimental Frequency Distributions

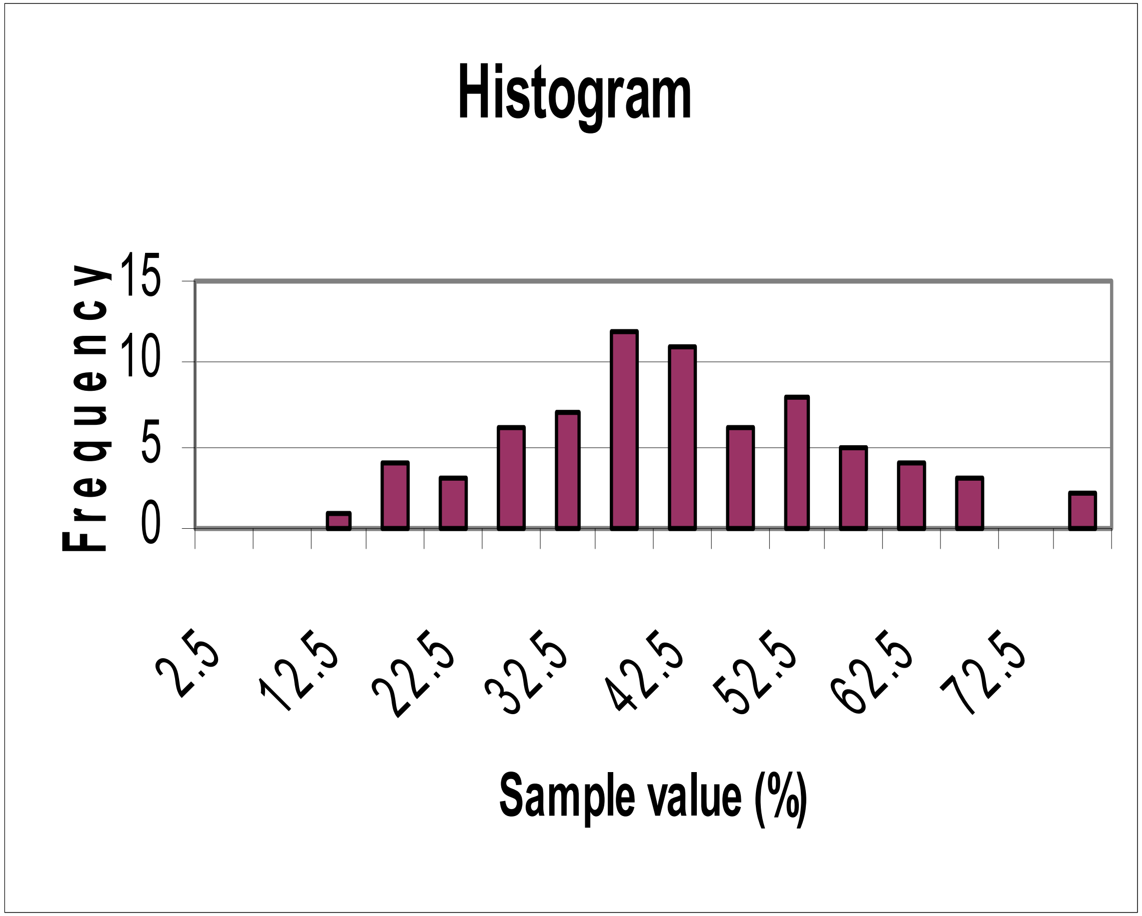
3- Example

Frequency Distribution Table

Bin(Class mid-point)	Frequency	Relative Frequency	% Frequency	CF	%CF
2.5	0	0	0	0	.00%
7.5	0	0	0	0	.00%
12.5	1	0.013888889	1.388888889	1	1.39%
17.5	4	0.055555556	5.555555556	5	6.94%
22.5	3	0.041666667	4.166666667	8	11.11%
27.5	6	0.083333333	8.333333333	14	19.44%
32.5	7	0.097222222	9.722222222	21	29.17%
37.5	12	0.166666667	16.66666667	33	45.83%
42.5	11	0.152777778	15.27777778	44	61.11%
47.5	6	0.083333333	8.333333333	50	69.44%
52.5	8	0.111111111	11.11111111	58	80.56%
57.5	5	0.069444444	6.944444444	63	87.50%
62.5	4	0.055555556	5.555555556	67	93.06%
67.5	3	0.041666667	4.166666667	70	97.22%
72.5	0	0	0	70	97.22%
77.5	2	0.027777778	2.777777778 72	100.00%	
Sum	72	1	100		

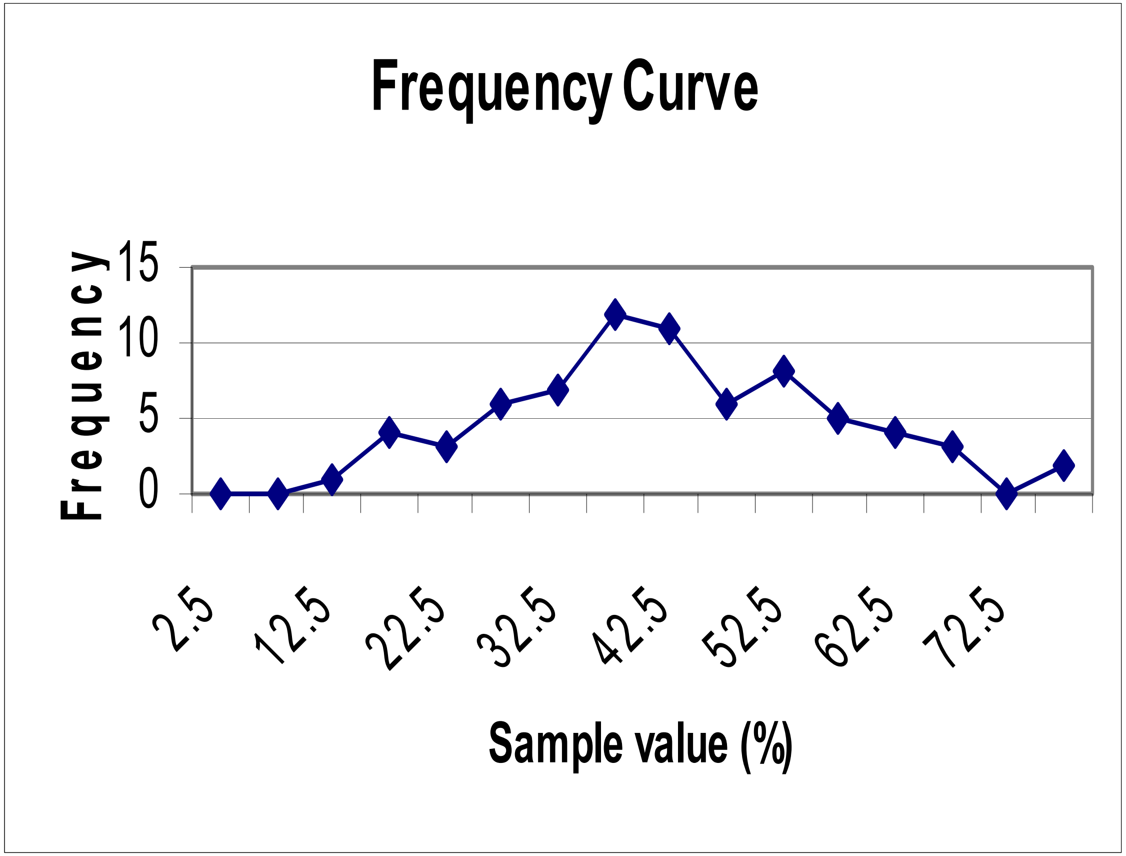
Experimental Frequency Distributions

3- Example



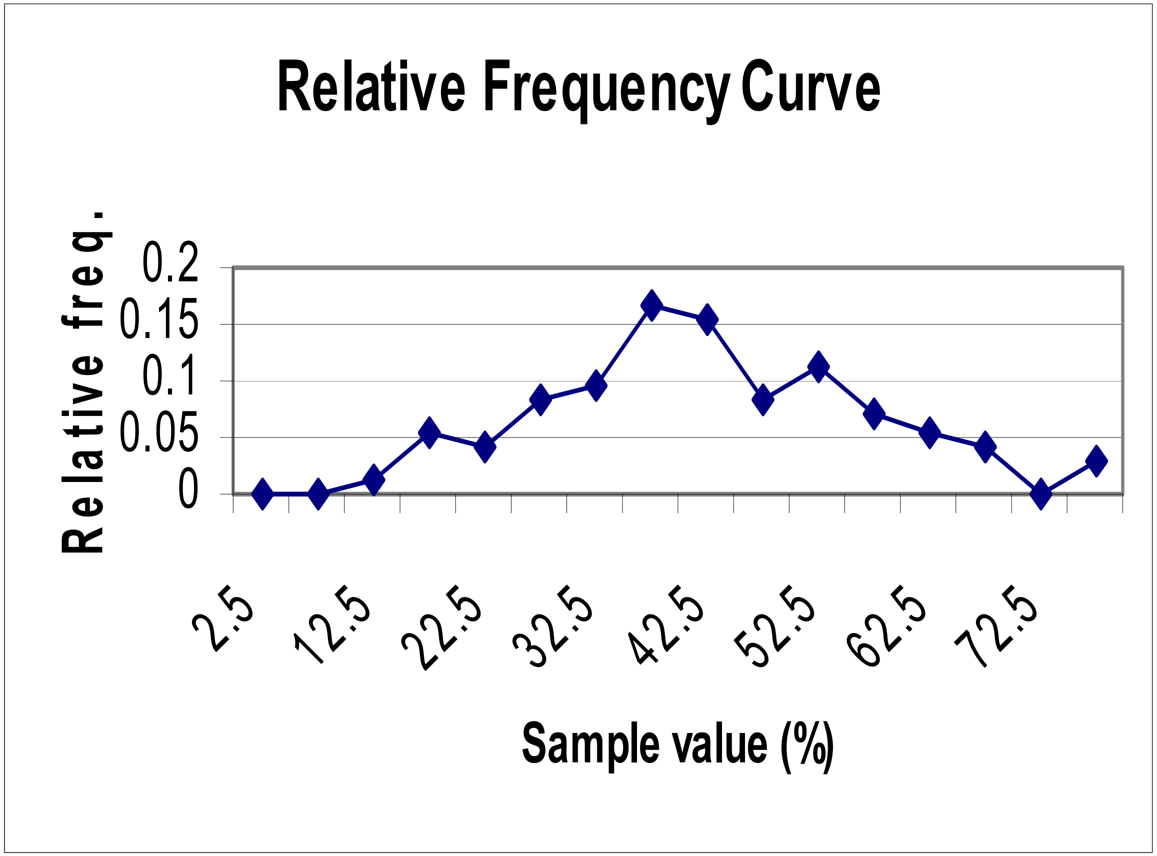
Experimental Frequency Distributions

3- Example



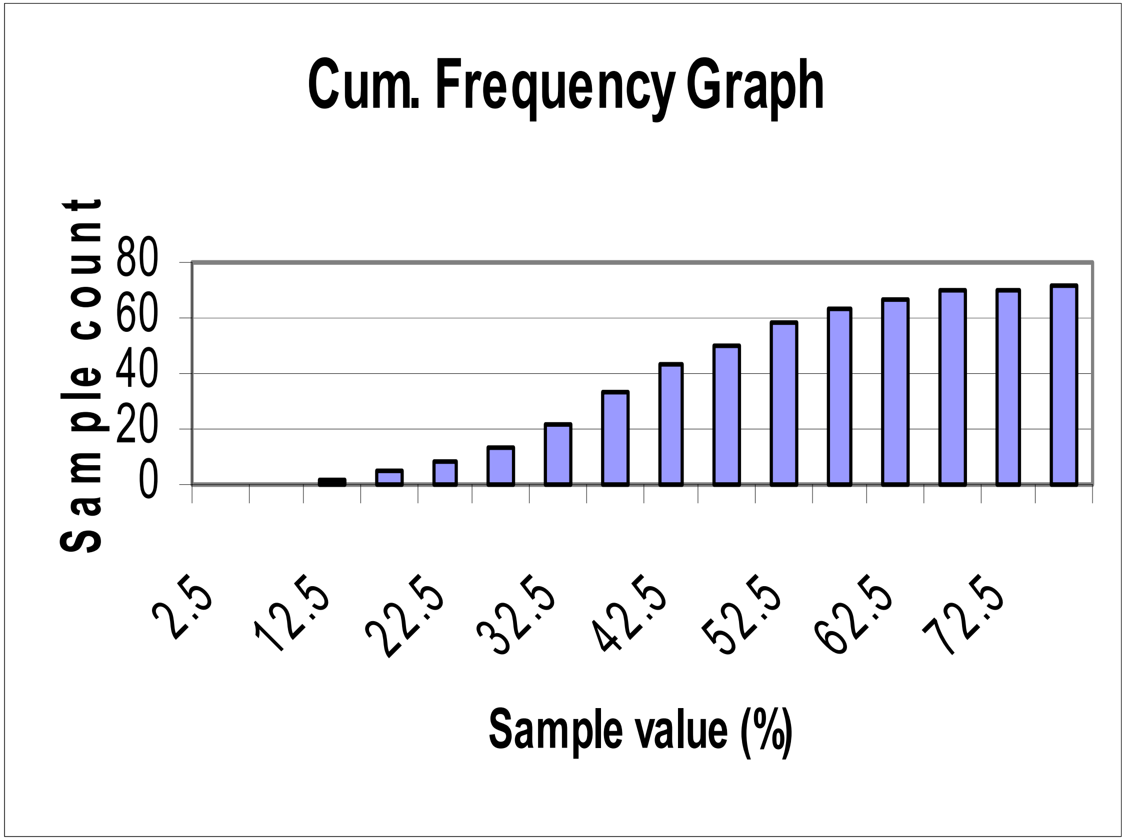
Experimental Frequency Distributions

3- Example

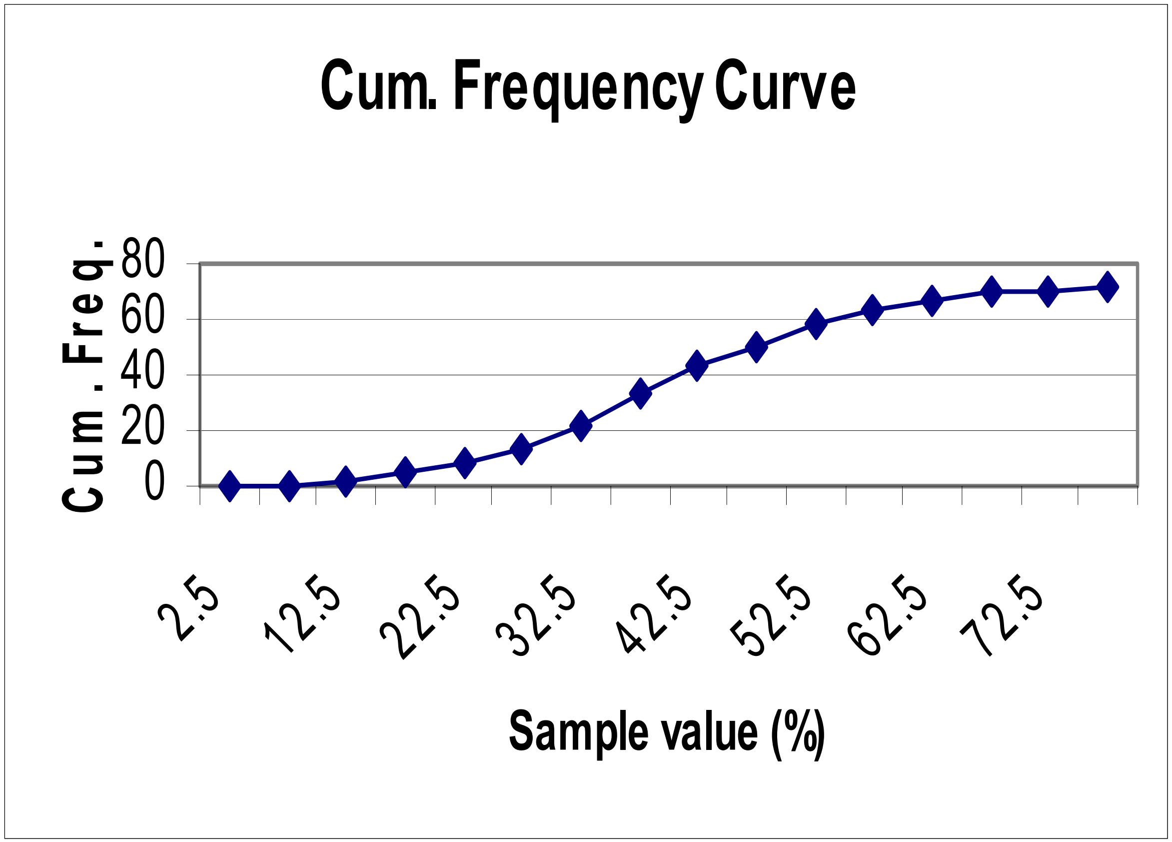


Experimental Frequency Distributions

3- Example

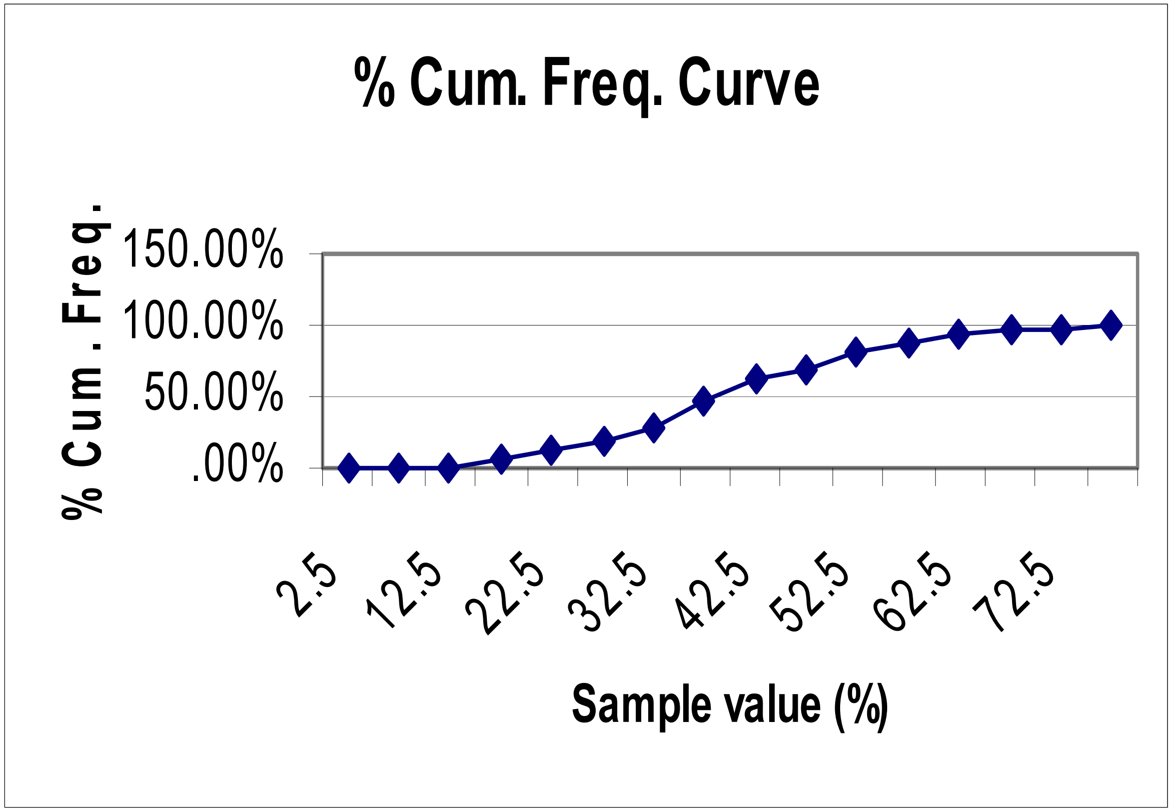


Experimental Frequency Distributions
3- Example



Experimental Frequency Distributions

3- Example



Measures of Center

Give an idea where the center of a data set distribution lies

- Mean (m)
  - Arithmetic average of a data set (sample)
  - It is sensitive to extreme values

$$m = \bar{v}_a = \frac{1}{n}\sum_{i=1}^{n} v_i$$

Measures of Center

Give an idea where the center of a data set distribution lies

- Median (M)
  - A midpoint of the observed values if they are arranged in an increasing order.
  - Half of the values are below the median and half of them are above it
  - It is not sensitive to extreme values
  - It is sensitive to gaps in the middle of a data set.

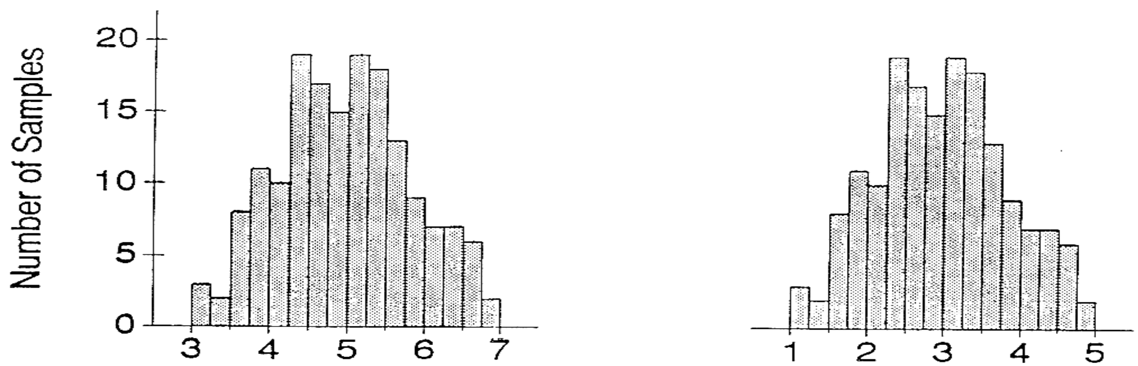
$$M_{odd} = x_{\frac{n+1}{2}} \qquad M_{even} = \left(\frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \right)$$

Measures of Center

Give an idea where the center of a data set distribution lies

- Mode
  - The observation that occurs most frequently.

SUMMARY STATISTICS—Measures of center



Measures of Location

Give an idea where the location of a specific observation in a data set distribution lies

- Minimum (*min*): the smallest value in the data set.
- Maximum (*max*): the largest value in the data set.
- Lower or First Quartile (Q_1): an observation value below which quarter of data falls.
- Upper or Third Quartile (Q_3): an observation value above which quarter of data falls.

Measures of Location

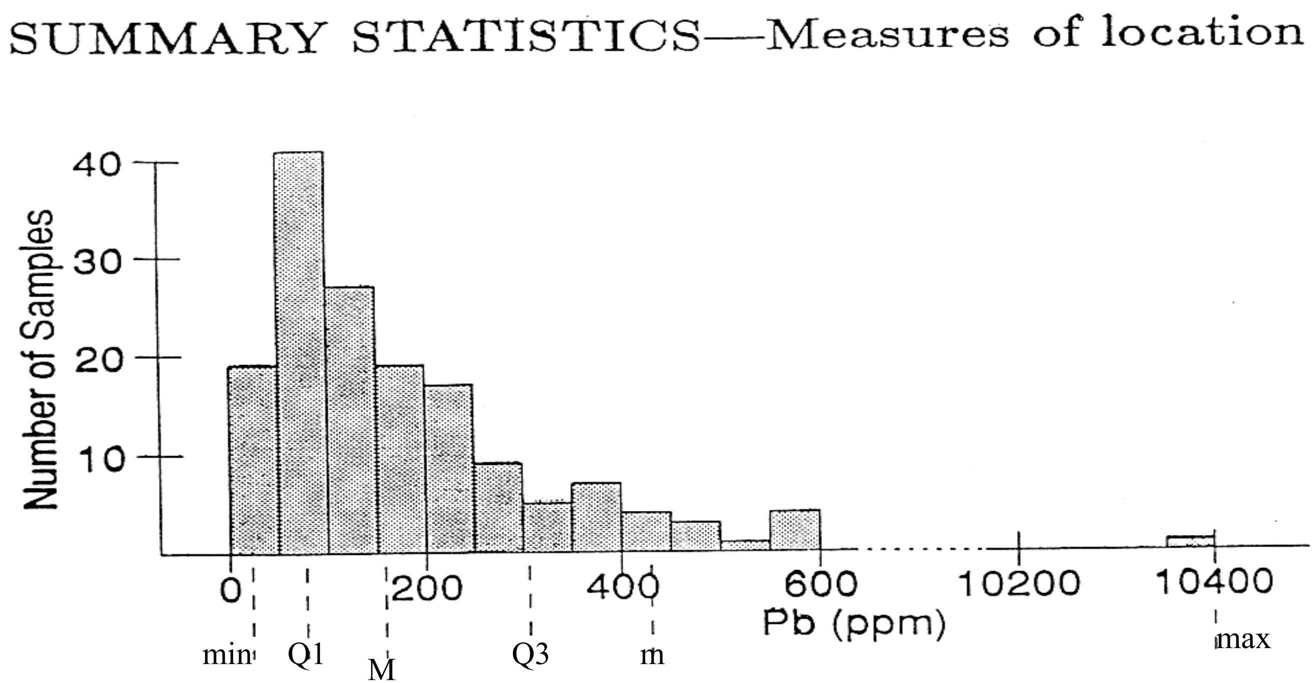
Give an idea where the location of a specific observation in a data set distribution lies

- Quantile (q_p): a general expression that describes an observation value below which a *percentage or fraction* quantity of data falls.

  - » $Min = q_{0}$
  - » $Q_1 = q_{0.25}$
  - » $M = q_{0.50}$
  - » $Q_3 = q_{0.75}$
  - » $max = q_{1}$

Measures of Location

Give an idea where the location of a specific observation in a data set distribution lies



Measures of Spread

Describe the variability of the data values

- Variance (S^2): is the average squared difference of the observed values from their mean.
  - It is sensitive to extreme value
  - **Standard Deviation (S):** is the square root of variance. It measures the uncertainty of the estimated mean value, for example. It is, also, sensitive to extreme values.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(v_i - m)^2$$

Measures of Spread

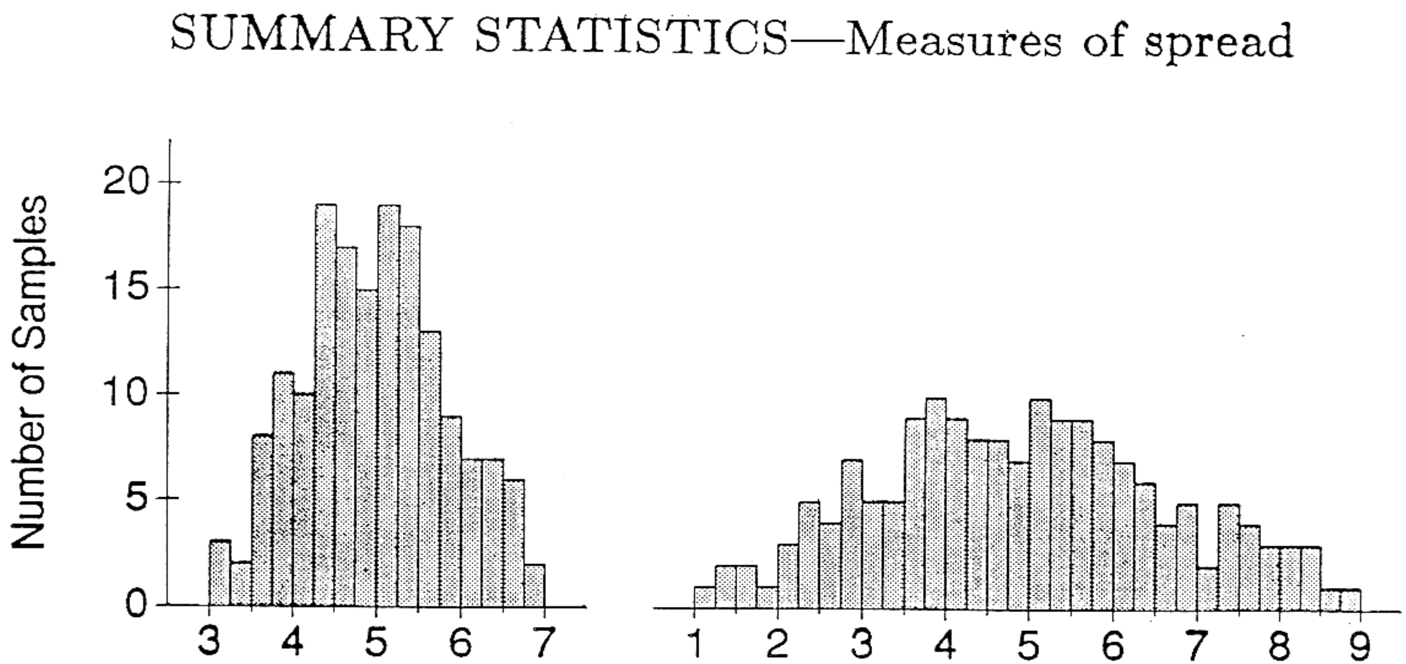
Describe the variability of the data values

- Interquartile Range (*IQR*): is the difference between upper and lower quartiles.
  - It is not sensitive to extreme value
  - It is a rough measure of spread of data values.

$$IQR = Q_3 - Q_1$$

Measures of Spread

Describe the variability of the data values



Measures of Shape

Describe the shape of distribution of the data values

- Coefficient of Skewness (*CS or g*): is the measure of symmetry of data values distribution.

$$CS = g = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}(v_i - m)^3\right)}{S^3}$$

Measures of Shape

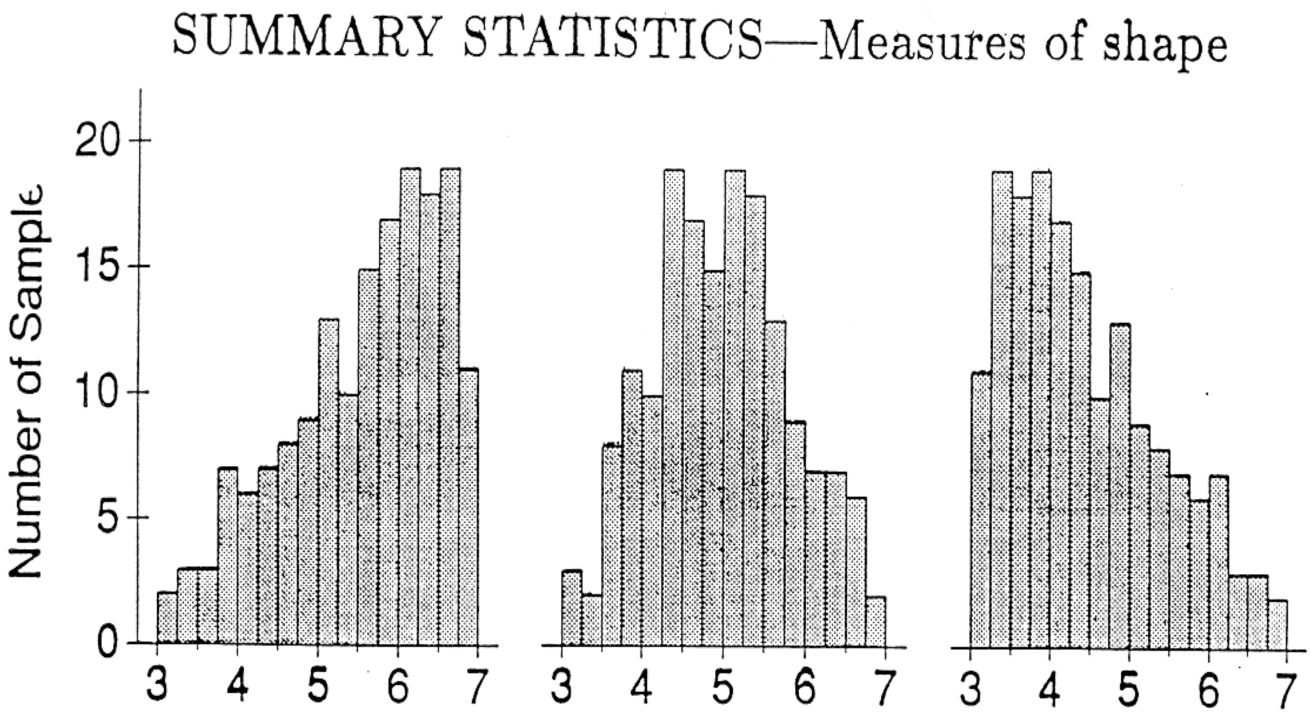
Describe the shape of distribution of the data values

- It is very sensitive to extreme value

- Its range is:
  - » SC = 0 — the data values are symmetrical around the mean value (M=m).
  - » SC = +ve — larger number of observations with low values in the data set (M<m).
  - » SC = -ve — Larger number of observations with high values in the data set (M>m).

Measures of Shape

Describe the shape of distribution of the data values



Measures of Shape

Describe the shape of distribution of the data values

- Coefficient of Variation (*CV*): is a measure of how significant is the impact of the presence of high values on the final estimates.

$$CV = \frac{S}{m}$$

  - It is a measure for deciding whether to use linear or non-linear geostatistics:
    - » CV < 1 use linear geostatistics
    - » CV 1-2 a caution is needed to deal with the problem
    - » CV > 2 use non-linear or non-parametric geostatistics

Measures of Shape

Describe the shape of distribution of the data values

- Kurtosis (κ) or (K): is a measure of peakedness of a distribution.

  - Kurtosis = 3 ==> Normal distribution with moderate peak and systematic shape (Mesokurtic)

  - Kurtosis > 3 ==> Distribution with sharply high peak (Leptokurtic)

  - Kurtosis < 3 ==> Distribution with flat top (Platykurtic)

Measures of Shape

Describe the shape of distribution of the data values