Statistical Properties & Parameters

The histogram shows important features of a data set by means of **properties** of the distribution. These histogram properties are:

1) Location: Represented by the position of a parameter along the histogram scale. Measure of center is a special case of location measurements.

2) Dispersion: Represented by the extent to which the distribution is spread out along the scale or how much values vary from some central value (i.e. average)

3) Shape: Represented by the pattern of the statistical distribution.

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.

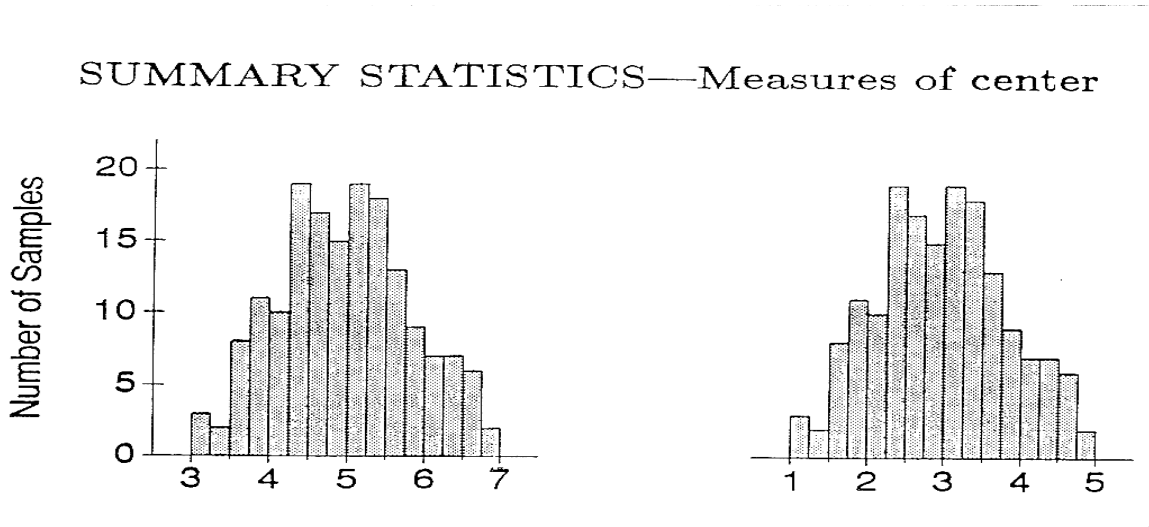
$$M_{odd} = x_{\frac{n+1}{2}}$$

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

Measures of Center

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

- **Mode**
  - The observation that occurs most frequently.



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

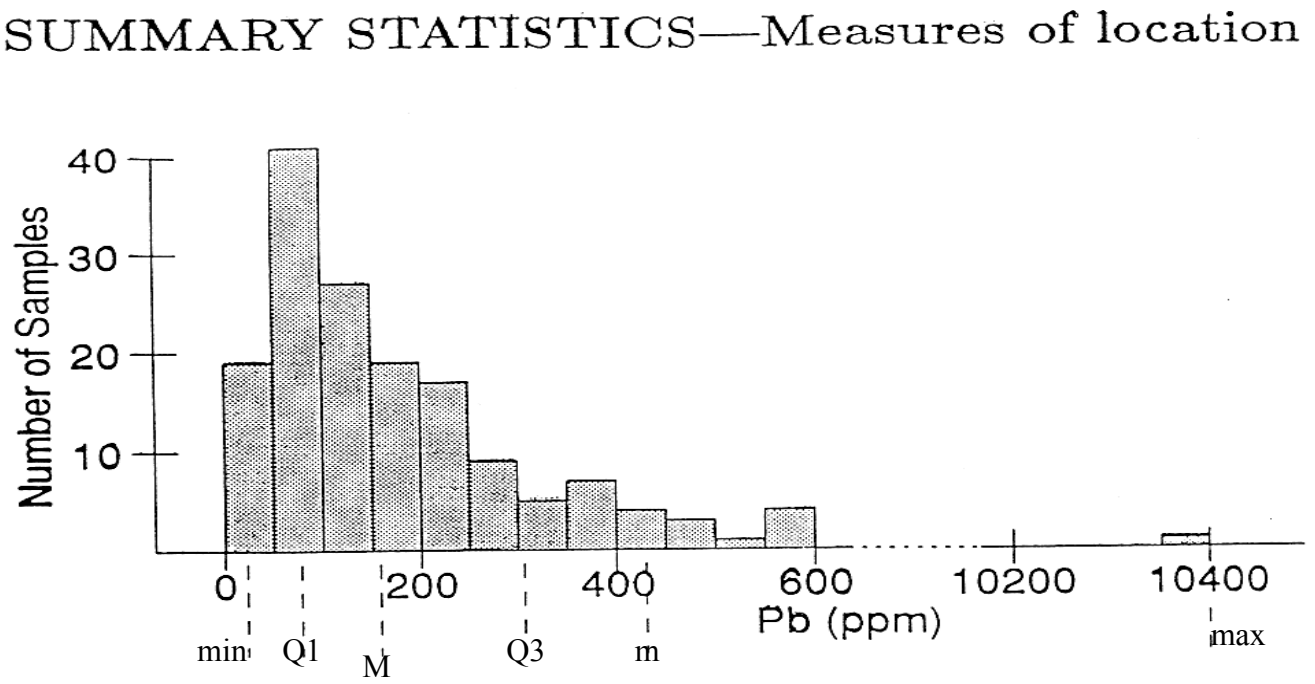
Give an idea where the location of a specific observation in a data set distribution lies

- **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.

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

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

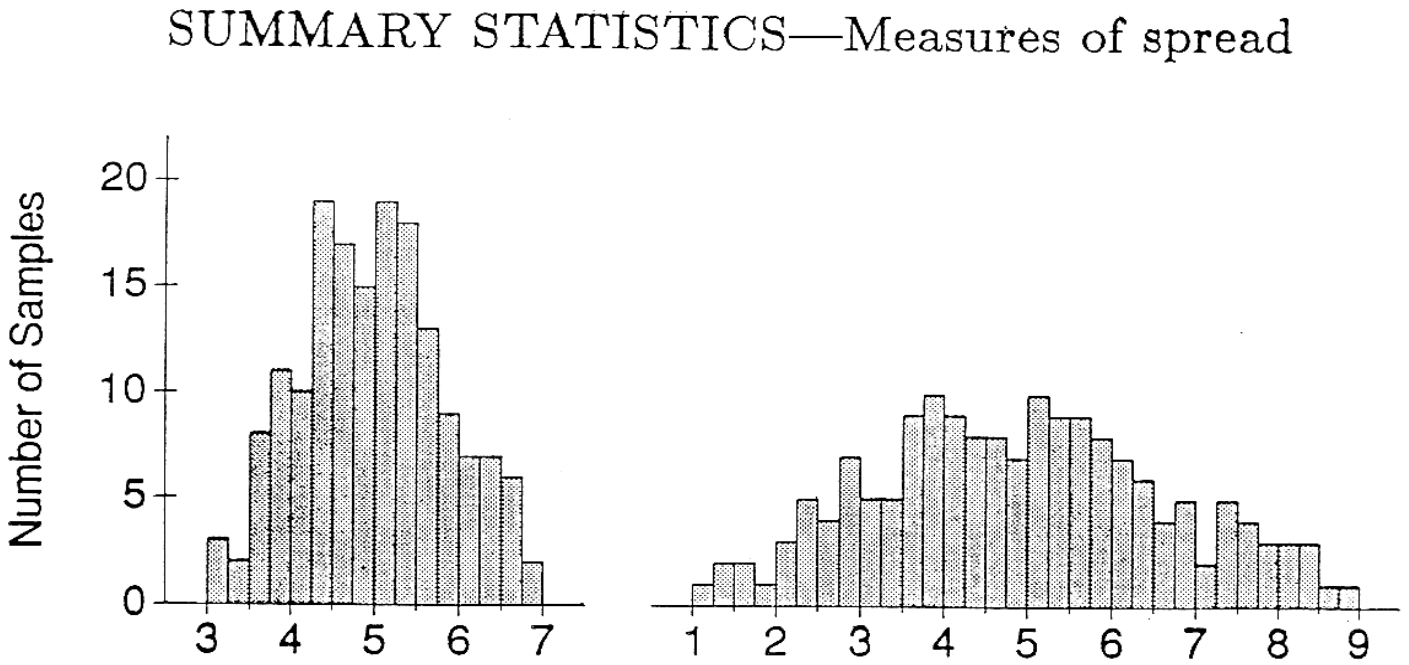
- **Coefficient of Variation (*CV*):** is a measure of how significant is the impact of the presence of high values on the final estimates.

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

  - 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 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

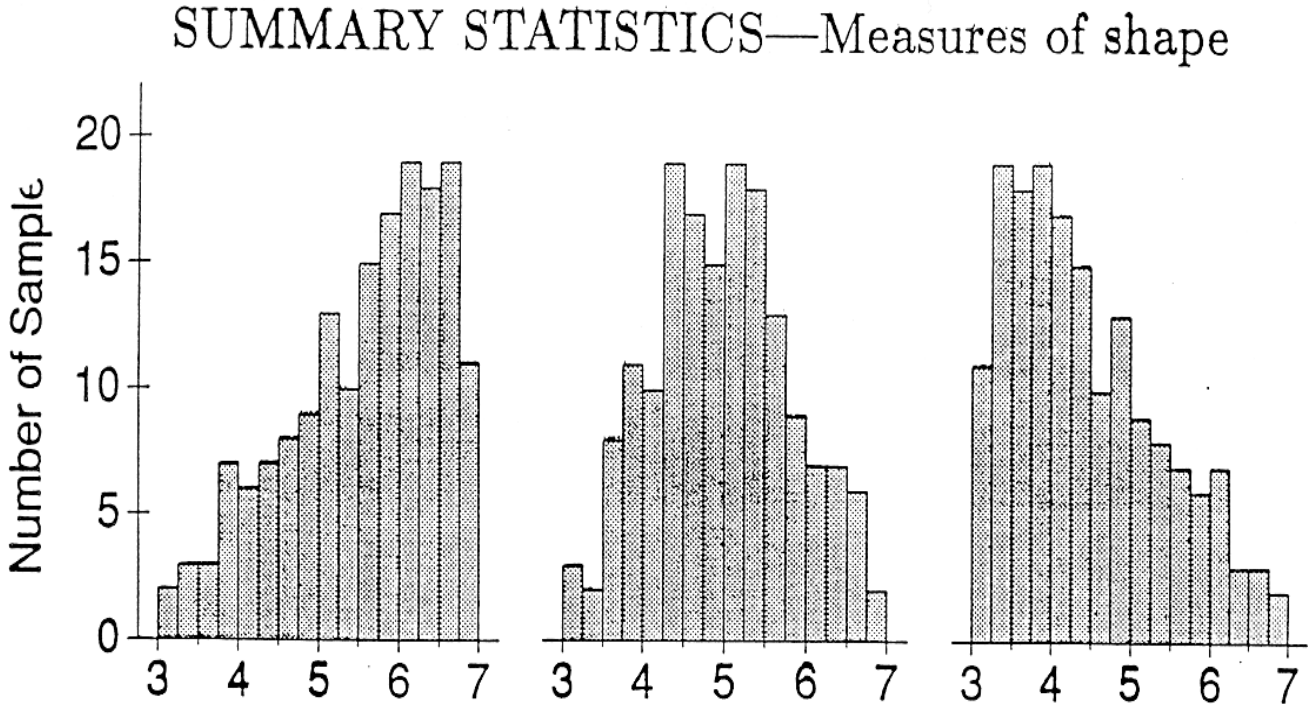
- It is very sensitive to extreme value

- 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

- **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