Chapter – 2

Literatures and Research Studies

2.1 Introduction:

The need of formulation of a theory in the field of statistics appeared in the works done under the project entitled **"Determination of the Natural Extrema of Temperature in the Context of Assam".** In this chapter, a brief discussion on formulation has been presented below.

2.2 Method of Obtaining Natural Extrema of Temperature:

Let Y be the random variable that represent the mean maximum temperature at a location. The maximum temperature at a location in a particular day in a year should remain the same in every successive year provided there is no any assignable cause of variation. But assignable cause / causes of variation rainfall, winds, clouds etc. may appear in the same day of the successive years that influence upon the temperature. So, analysis of the daily data on temperature cannot yield valid results. However, monthly data on such variable can estimate this type of causes of variation. Hence, it would be reasonable to analyses monthly data instead of daily data.

Let

Y_i = the highest maximum temperature observed at a location in the year i

$(i = 1, 2, 3, \ldots\ldots\ldots\ldots)$.

The values of Y_i should be constant if there exists no cause of variation in Y_i over years. However, chance cause (random cause) of variation exists always. Thus if no assignable cause of variation exists in Y_i over year, we have

$$Y_i = \mu + \varepsilon_i \tag{2.1}$$

where μ = the true value of the highest maximum temperature observed at a location in the year i

& ε_i = the chance error associated to Y_i.

The common assumption of the error component is that ε_i s are independently and identically distributed normal variables with zero mean and an unknown variance σ_ε^2 i.e. ε_i s are i.i.d $N(0, \sigma_\varepsilon)$ variables { *Anderson* (1952), *Fisher* (1932) et al}.

Now, since $\varepsilon_i \sim N(0, \sigma_\varepsilon)$

therefore $Y_i - \mu \sim N(0, \sigma_\varepsilon)$

which implies that

$$\text{(i)} \quad P\{(Y_i - \mu) / \sigma_\varepsilon \leq 1.96\} = 0.95, \tag{2.2}$$

$$\text{(ii)} \quad P\{(Y_i - \mu) / \sigma_\varepsilon \leq 2.58\} = 0.99 \tag{2.3}$$

$$\text{\& (iii)} \quad P\{(Y_i - \mu) / \sigma_\varepsilon \leq 3\} = 0.9973 \tag{2.4}$$

Therefore the intervals

$$\text{(i)} \quad \mu - 1.96\,\sigma_\varepsilon \leq Y_i \leq \mu + 1.96\,\sigma_\varepsilon, \tag{2.5}$$

$$\text{(ii)} \quad \mu - 2.58\,\sigma_\varepsilon \leq Y_i \leq \mu + 2.58\,\sigma_\varepsilon \tag{2.6}$$

$$\text{\& (iii)} \quad \mu - 3\,\sigma_\varepsilon \leq Y_i \leq \mu + 3\,\sigma_\varepsilon \tag{2.7}$$

are respectively 95%, 99% & 99.73% confidence intervals of Y_i.

Now, $P\{(Y_i - \mu) / \sigma_\varepsilon \leq 3\} = 0.9973$

implies that a random value of Y_i goes outside the interval

$$\mu - 3\,\sigma_\varepsilon \leq Y_i \leq \mu + 3\,\sigma_\varepsilon \tag{2.8}$$

is 0.0027 which is very small. In other words, it is near certain that Y_i falls inside the interval that is it is natural that Y_i falls inside the interval. For this reason, this interval is termed as the natural interval of Y_i {*Shewhart* (1931), *Grant* (1972)}.

Thus in order to determine the natural interval (more specifically natural maximum and natural minimum) of Y_i, it is required to estimate of μ and σ_ε.