Mathematical formulation of Kriging :

Let the symbol $Z(s)$ denotes the observed value of the studied spatial process at the point s. The principle of prediction is based on a weighted average of neighboring values $Z(s_i)$, where the weights λ_i depend on the distance and spatial relationship between observed points.
Mathematically, the prediction at the point s_0 can be described by the equation

$$Z^*(s_0) = \sum \lambda_i Z(s_i)$$

The weights sum to one to assure unbiasedness condition and they are found by minimizing the estimation variance.

The random variable Z(s) can be decomposed into a trend component m(x) and a residual component R(x).

$$Z(s) = m(s) + R(s)$$

Ordinary kriging assumes stationarity of the first moment of all random variables. i.e. it assumes constant mean $E[Z(s_i)] = \mu$ which is unknown. Nonstationary conditions are taken into account by restricting the domain of stationarity to a local neighborhood and moving it across the study area. Ordinary kriging is based on the assumption that the correlation between two random variables depends only on their spatial distance that separates them and is independent of their position. The variance of the difference of two random variables $Z(s)$ and $Z(s+h)$ depends only on their spatial distance h.

$$var[Z(s+h) - Z(s)] = 2\gamma(h)$$

Where, $2\gamma(h)$ is called variogram and $\gamma(h)$ is a semivariogram.
The residual component $R(s)$ is modeled as a stationary random variable with zero mean and under the assumption of intrinsic stationarity, its spatial dependence is given by the semivariance

$$\gamma_R(h) = \frac{1}{2}E[(R(s+h) - R(s))^2]$$

Assuming a constant mean $m(s)$ above equation is equivalent to:

$$\gamma(h) = \frac{1}{2}E[(Z(s+h) - Z(s))^2]$$

Error in prediction i.e Kriging variance associated to an Ordinary Kriging estimate is:

$$\sigma^2_{OK} = \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_0) - \mu$$

Second order stationarity and use of Universal Kriging:

If the first moment of the observed field is not stationary and a polynomial trend occurs in the data, it is appropriate to use Universal kriging.

Universal kriging considers that $m(s)$ is not constant, but that it varies smoothly within the local neighborhood, representing a local trend. The trend $m(s)$ is recalculated within each local neighborhood. This trend component is modeled as a weighted sum of known functions $f_i(s)$ and unknown coefficients a_l, $l = 0, \ldots, L$

$$m(s) = \sum_{l=0}^{L} a_l f_l(s)$$