Specification of Population and Sample Regression Functions

Dr.A.Hidhayathulla
Associate Professor of Economics
Jamal Mohamed College
Trichy

Population Regression Function

A population regression function hypothesizes a theoretical relationship between a dependent variable and a set of independent or explanatory variables, that is supposed to exist among all the values of a study population. It is a linear function. The function defines how the conditional expectation of a variable Y responds to the changes in independent variable X. $Y\text{i} = E(Y\text{i}|X\text{i}) + u\text{i}$ … (1) The function consists of a deterministic component $E(Y|X)$ and a nondeterministic or 'stochastic' component u, as depicted in equation (1).

We are concerned about examining the determinants of dependent variable (Y) conditional upon the given values of impendent variables (X).

Specification of Population Regression Function

$Y_i = \beta_0 + \beta_1 X_i + U_i$

Y_i = DV, X_i = IV, U_i = Random disturbance term, β_0 = Constant, β_1 = Intercept or Slope

Y_i = Dependent variable, response variable, endogenous variable, regressand

X_i = Independent variable, explanatory variable, exogenous variable, predictor, regressor

Sample Regression Function

We rarely have the data related to the entire population at our disposal. We only have a sample from the population. Thus, we need to use the sample to estimate the population parameters. We may not be able to find out the population regression line (PRL) because of sampling fluctuations or sampling error.

Suppose we have two samples from the given population. Using the samples separately, we obtain Sample Regression Lines (SRLs). A sample represents the population.

Specification of Sample Regression Function

$Y_i = b_0 + b_1 X_i + e_i$

If we depict the population regression line (PRL) and the sample

regression line (SRL) we can observe that the slopes of both the lines are different.

Thus, $b_0 \neq \beta_0$ and $b_1 \neq \beta 1$. Let us consider a particular value of the explanatory variable, X_1. The corresponding value of the explained variable is Y_1. On the basis of the sample regression line we obtain estimated value of the explained variable, $\hat{Y}_1$. Now let us find out the distinction between the error term (u) and the residual (e). The distance between the actual value Y_1 and the corresponding point on the population regression line is u_1. This error u_1 is not known to us, because we do not know the values of β_0 and β_1. What we know is $\hat{Y}_1$, which is estimated on the basis of b_0 and b_1. The distance between Y_1 and $\hat{Y}_1$ is the residual, e_1.

Difference between Population Regression and Sample Regression

Population regression function (PRF) is the locus of the conditional mean of variable Y (dependent variable) for the fixed variable X (independent variable). Sample regression function (SRF) shows the estimated relation between explanatory or independent variable X and dependent variable Y.