Double or Two-Phase Sampling
(Adapted from Lohr 1999)

The ratio and linear regression estimators of μ assumed the auxiliary variables $x_1, ..., x_N$ and thus μ_x (τ_x) are known. These sampling designs are used to increase the precision of the estimator. However, it may not be possible to obtain all the auxiliary measurements. For example, suppose you would like to estimate the total timber volume that has been cut in a forest by measuring the total volume in a sample of truckloads of logs. Timber volume in a truck is related to the weight of the truckload, so you would expect to gain precision by using ratio estimation with y_i = timber volume in truck i and x_i = weight of truck i. But the ratio estimator requires that the weights of all the trucks be known, and weighing every truck in the population may not be practical.

Two-Phase sampling or double sampling, which was introduced by Neyman (1938), is useful when the variable of interest (y) is relatively expensive to measure, but a correlated variable (x) can be measured fairly easily and used to improve the precision of the estimator. Suppose there the population has N units. The sample is taken in two phases:

Phase 1: The first sample of units is selected and auxiliary information (x) collected on every unit in this sample.

Phase 2: The second sample, possibly a subsample of the first, is selected and the values of the auxiliary variable (x) and the variable of interest (y) are collected on every unit in this sample.