Highlights

•The problem of unsupervised sample selection for multivariate calibration is defined as finding a subset of samples that would render a calibration model of comparable performance to the model that would result if the full available sample set were submitted to reference analyses.

•The optimal sample size for a multivariate calibration model depends on the complexity of such a model.

•It is possible to evaluate the quality of the selected samples for PLSR calibration models by comparing the covariance matrix based on the selected subset with the covariance matrix based on the full available sample set.

•By controlling the sample size and the input dimensionality, the state-of-the-art sample selection algorithms render subsets of samples with comparable quality for the calibration model