

	Bootstrap	Permutation	Cross-Validation
Use	Estimating Features of the sampling distribution of an estimator (e.g., SE,CIs, assessment of bias)	Obtaining p-values that do not depend on assumptions about the distribution of the data	Quantifying the ability of a model to produce out-of-sample predictions (predictive ability)
Implementation	<ol style="list-style-type: none"> <li>1. Generate a bootstrap sample (often the same size as our actual sample size) by sampling rows of your data with replacement</li> <li>2. Use the bootstrap sample to obtain estimates, store them.</li> <li>3. Repeat 1 &amp; 2 a large number of times.</li> <li>4. Regards the estimates you got as realizations from the sampling distribution of the estimator.</li> <li>5. Use the samples to estimate SE (i.e., the SD of estimates across bootstrap samples), CIs (e.g., using quantile()) applied to bootstrap samples, and possible assess bias by comparing the average bootstrap estimate with the estimate you obtain from the original sample.</li> </ol>	<ol style="list-style-type: none"> <li>1. Fit the model to your data, save the test-statistic that you will use for inference.</li> <li>2. Break the relationship between the response and the predictors by randomly permuting the order of the response (or one predictor, depending on the application).</li> <li>3. Apply your algorithm to the permuted data to obtain the test-statistic.</li> <li>4. Repeat 2 &amp; 3 a large number of times, each time with a different permutation of the data, save the test statistic.</li> <li>5. Estimate p-values as the proportion of times (over permutations) that you obtain a test-statistic as extreme or more extreme than the one obtained with the un-permuted data (1).</li> </ol>	<p>We discussed Training-testing partitions, replicated training-testing partitions, and cross-validation.</p> <p>Example: training testing partitions:</p> <ul style="list-style-type: none"> <li>- Randomly assign nTRN rows of your data to be a training set, and nTST rows of your data to be the testing set (N=nTRN+nTST, no overlap between the sets).</li> <li>- Fit the model to the training data.</li> <li>- Use the fitted model to predict data in the testing set.</li> <li>- Evaluate accuracy using correlation, pred. mean-squared error, R-sq. or related measures of accuracy.</li> <li>- Possibly repeat 1-3 many times, each time changing the assignment of samples to training and testing sets.</li> </ul>