Chapter 10

Glossary
This glossary tries to summarise the acronyms used in the preceding chapters. It is based mainly on the “Cambridge Dictionary of statistics in the medical sciences” by B. S. Everitt [1], adapted to the problems under discussion whatever necessary.

**APACHE:** Acute Physiology and Chronic Health Evaluation. Used also later as an acronym for Acute Physiology, Age, Chronic Health Evaluation. General outcome prediction model developed by W. Knaus et al. in 1981 and subjected to major revisions in 1985 (APACHE II) and 1991 (APACHE III).

**Assumptions:** the conditions under which statistical techniques give valid results.

**Bias:** any systematic deviation of an observation from the true clinical state.

**Bootstrap:** a data-based simulation method for statistical inference, which can be used to study the variability of estimated characteristics of the probabilities distribution of a set of observations, and provide confidence intervals for parameters, in situations where these are difficult or impossible to derive in the usual way.

The basic idea of the procedure involves sampling with replacement to produce random samples of size \( n \) from the original data, \( x_1, x_2, \ldots, x_n \); each of these is known as a bootstrap sample and each provides an estimate of the parameter of interest. Repeating the procedure a large number of times provides the required information on the variability of the estimator, and an approximate confidence interval can be derived from the quantiles of the derived value.

**C test:** see Hosmer-Lemeshow goodness-of-fit tests.

**Calibration of a model:** the degree of correspondence between the observed values of the outcome variable and the values predicted by the model. Also known as reliability of a model.

**Calibration curves:** plot of the observed versus predicted values of the outcome variable.

**Categorical variable:** a variable that gives the appropriate label of an observation after allocation to one of several possible categories.

**Confidence interval:** a range of values calculated from the sample observations that are believed, with a particular probability, to contain the true parameter value.

**Confounder:** a factor that distorts the relationship of the study variables of central interest by virtue of being related to the outcome of interest but extraneous to the study question and unequally distributed among the groups being compared.
Contingency tables: the tables arising when observations on a number of categorical variables are cross-classified. Entries in each cell are the number of individuals with the corresponding combination of variable values.

Covariates: variables that are not of primary interest in an investigation, but are measured because it is believed that they are likely to affect the response variable, and consequently need to be included in analysis and model building. Often used as an alternative name for explanatory variable.

Cross-validation: the division of data in two subsets, one of which is used to estimate the parameters of the model and the other to assess whether the model with these parameters fits adequately.

Data smoothing algorithms: procedures for extracting a pattern in a sequence of observations when this is obscured by noise.

Discrimination of a model: the ability of a model to distinguish between observations with a positive or a negative outcome.

Dummy variables: the variables resulting from recoding categorical variables with more than two categories in a series of binary variables.

Effect modification: the variation in the magnitude of an effect measure across levels of a third variable.

Efficacy: the effect of one intervention relative to a control in the ideal situation where compliance is perfect.

Effectiveness: the effect of one intervention under average conditions.

Explanatory variables: the variables used to predict or explain the response variable. Also known as independent variables.

External validation: the evaluation of a model in an external, independent population.

False negative rate: the proportion of cases in which a diagnostic test indicates disease-absent in patients who have the disease.

False positive rate: the proportion of cases in which a diagnostic test indicates disease-present in patients who are free of the disease.

Flora’s Z score: statistic that compares the number of survivors observed in the given data set with the number which would be predicted from the baseline survival curve.

Goodness-of-fit statistics: measures of the agreement between a set of observations and the corresponding values predicted by a model.

H test see Hosmer-Lemeshow goodness-of-fit tests.

Harrel’s index: non-parametric test using the rank of the magnitude of the assessment error for measuring the probability that for two randomly chosen patients, the one with the higher probabilistic prediction has the outcome of interest.
Hosmer-Lemeshow goodness-of-fit tests: chi-square statistics proposed for the formal evaluation of the calibration of outcome prediction models by Hosmer and Lemeshow.

In H test, patients are divided in 10 strata, according to their predicted probabilities: 0.0 to 0.1, 0.1 to 0.2 … 0.9 to 1.0. Then, a chi-square statistic is used to compare the actual and the expected number of deaths and the actual and expected number of survivors in each of this 10 strata.

C test is similar, with the 10 strata being formed based on deciles of the predicted probabilities.

Influential observation: an observation that as a disproportionate influence on one or more aspects of the estimate of a parameter, in particular, a regression coefficient.

Interaction: a term applied when two (or more) explanatory variables do not act independently on a response variable.

Intercept: the parameter in an equation derived from a regression analysis corresponding to the expected value of the response variable when all the explanatory variables are zero.

Intraclass correlation coefficient: a statistic used for the evaluation of the agreement between two raters in the measurement of continuous variables.

Jackknife: a procedure to reduce bias in estimation and providing approximate confidence intervals in cases where these are difficult to obtain the in the usual way.

The principle behind the method is to omit each sample member in turn from the data, thereby generating n separate samples each of size n-1. The parameter of interest can now be estimated from each of these sub-samples, giving a series of estimates, which allow the estimation of the parameter and corresponding standard error in the overall sample.

Kappa coefficient: a chance corrected index of the agreement between two raters.

Lead-times bias: the effect of previous therapy on the degree of physiologic derangement present at admission to the ICU. Usually evaluated indirectly by the location of the patient in the hospital prior to ICU admission.

Logistic regression: a form of regression analysis that is used when the response variable is binary. The method is based on the logistic transformation or logit of a proportion, namely:

\[
\text{logit} \left( p \right) = \ln \frac{p}{1 - p}
\]

where p is the probability of the event.

McNemar’s test: a test for comparing proportions in data involving paired samples.
MPM: Mortality Probability Models. General outcome prediction models developed by S. Lemeshow in 1987, based on previous research, and subjected to a major revision in 1993 (MPM II). In its actual form includes specific models for outcome prediction at admission to the ICU (MPM II₀) and at 24 (MPM II₂₄), 48 (MPM II₄₈) and 72 (MPM II₇₂) hours in the ICU.

Multi-collinearity: term used in regression analysis to indicate situations where the explanatory variables are related to a linear function, making the estimating of regression coefficients impossible.

Odds: the ratio of the probabilities of the two possible states of a binary variable.

Odds ratio: the ratio of the odds for a binary variable. It is a measure of the odds of an event in one group relative to that in the other.

Overall goodness of fit: the degree of accuracy of the predictions of a model. Usually evaluated by tests of calibration and discrimination.

Receiver operating characteristic (ROC) curve: a plot of the rate of true positives (sensitivity) of a model against the rate of false positives (1 minus specificity) over all the possible values of the model being evaluated. Its area, which corresponds directly to Harrel’s c index, can be used to evaluate the discrimination of a model. The interpretation of the value of the area under ROC curve is easy: a model with a perfect discrimination as an area of 1.0, a model no better than chance an area of 0.5.

Refinement of a model:

Reliability of a model: see calibration of a model.

SAPS: Simplified Acute Physiology Score. General outcome prediction model developed by J.-R. Le Gall et al. in 1983 and subjected to a major revision in 1993 (SAPS II).

Spread of a model: see discrimination of a model.

Standardised Mortality Ratio: the ratio of the observed number of deaths by the number of deaths predicted by the model.

Temporal validation: the test of a model developed at a specific point in time in the same setting at some future date.

Uniformity of fit: the evaluation of the appropriateness of the predictions across relevant sub-groups in order to identify in the test set the existence of subsets of observations in which the model does not perform well.

REFERENCES