

A SAS-macro package for development and validation of prognostic and diagnostic models based on the logistic regression model

Rainer Muehe¹, Christina Ring¹, Christoph Ziegler²

¹ *Institute of Biometrics, University of Ulm, Germany*

² *Biostatistics Department, F. Hoffmann-La Roche Ltd, Switzerland*

Medical diagnostic- and therapy-decisions are directly or indirectly determined by prognoses based on subjective impression and experience. Beside the subjective assessment of the doctors, mathematical models can be developed and validated for the prognosis. In practice, the outcome is frequently dichotomized so that multiple logistic regression models will be applied [1,3]. A SAS-macro package for logistic regression modelling will be characterized. The model-process can be handled in three consecutive steps: model-development, examination of the performance of prognosis and model-validation [2]. The package includes the following macros.

(1) Model-development:

Examination / description of the variables	PM_DESCRIPTOR.MAC.SAS
Multicollinearity	PM_MULTICOLLIN.MAC.SAS
Missing-Value substitution	PM_MISSING.MAC.SAS
Influential observations	PM_INFLUENCE.MAC.SAS
Univariate logistic regression / fractional polynomials	PM_UNI_LOGREG.MAC.SAS
Multiple logistic regression	PM_LOGREG.MAC.SAS
Goodness-of-fit examination	PM_GOF.MAC.SAS

(2) Performance of prognosis analyzed by ROC-analysis: PM_ROC.MAC.SAS

(3) Model-validation:

External validation	PM_EXTERNAL_VALIDATION.MAC.SAS
Data-splitting	PM_DATASPLITTING.MAC.SAS
Crossvalidation	PM_CROSSVALIDATION.MAC.SAS
Bootstrap	PM_BOOTSTRAP_VALIDATION.MAC.SAS
Shrinkage	PM_SHRINKAGE_VALIDATION.MAC.SAS

Altogether we present a tool, which can be helpful in developing and validating of a logistic prognosis-model. All macros and their statistical properties are discussed in detail in [4].

- [1] Allison, PD (1999): Logistic regression using the SAS system. SAS Institutes Books by Users, Cary NC
- [2] Harrell, FE Jr. (2001): Regression modelling strategies. Springer, New York
- [3] Hosmer DW, Lemeshow S (2000),: Applied logistic regression (2nd Eds.). J. Wiley & Sons, New York
- [4] Muehe R, Ring C, Ziegler C (2005): Entwicklung und Validierung von Prognosemodellen auf Basis der logistischen Regression. Shaker Verlag, Aachen