

PROC LOGISTIC: Warum sind die Koeffizienten nicht mit den Odds Ratios konsistent?

Braisch, Ulrike; Muche, Rainer

Universität Ulm

ulrike.braisch@uni-ulm.de

In vielen Analysen wird bei multivariablen Regressionsmodellen häufig mit kategorialen Prädiktoren gearbeitet. Das Ziel dieses Vortrages ist es, hierbei auf eine potenzielle Falle bei Anwendung einer logistischen Regression mit SAS (PROC LOGISTIC) hinzuweisen.

Eines der wichtigsten Ergebnisse der logistischen Regression ist der Odds Ratio (OR), der das Chancenverhältnis zweier Merkmalsausprägungen im Hinblick auf eine Zielvariable beschreibt. In Lehrbüchern wird der OR meist durch Exponieren des entsprechenden Regressionskoeffizienten β bestimmt ($OR = \exp(\beta)$). Diese Rechnungsweise ist in SAS allerdings nicht ohne Weiteres gültig, wenn man mit kategorialen Prädiktoren arbeitet und das CLASS-Statement einsetzt. Denn SAS hat bei der logistischen Regression die Effektkodierung als Dummy Kodierung voreingestellt, was dazu führt, dass die von SAS ausgegebenen ORs nicht mit denen übereinstimmen, die man fälschlicherweise mit obiger Formel berechnen würde. Die Gültigkeit obiger Formel kann in SAS allerdings wieder gewährleistet werden, indem `param=ref` als Option des CLASS-Statements hinzugefügt wird, wodurch SAS gezwungen wird die Dummy Kodierung gemäß der (üblichen) Referenzkodierung durchzuführen.

Im Rahmen des Beitrages wird an Hand eines Beispiels auf die zwei am häufigsten verwendeten Varianten der Dummy Kodierung eingegangen, der Effektkodierung (effect coding) und der Referenzkodierung (reference coding), und deren Einfluss auf die Odds Ratio Schätzung.

Bei der Schätzung des Hazard Ratio über die Cox Regression (PROC PHREG) hingegen ist als Dummy Kodierung die Referenzkodierung voreingestellt, so dass hier die oben beschriebene Falle gar nicht vorhanden ist, außer man ändert per Hand die Art der Dummy Kodierung. In dieser Hinsicht sollte also immer Vorsicht geboten sein, denn die Voreinstellung der Dummy Kodierung im CLASS-Statement unterscheidet sich in SAS zwischen den unterschiedlichen Regressionstypen.

Literatur

Lewis, Taylor (2007). PROC LOGISTIC: The Logistics Behind Interpreting Categorical Variable Effects <http://www.nesug.org/proceedings/nesug07/sa/sa11.pdf> (accessed September 11, 2012).

In PROC LOGISTIC why aren't the coefficients consistent with the odds ratios? UCLA: Academic Technology Services, Statistical Consulting Group. http://www.ats.ucla.edu/stat/sas/faq/proc_logistic_coding.htm (accessed September 11, 2012).