

Statistik-Übung 6 – GLMM

Beispielhafter Methoden- und Ergebnisteil

Methoden

Ziel war es zu ermitteln, welchen Einfluss das Geschlecht (männlich vs. weiblich), der Status (Mitarbeitende vs. Studierende) und das Alter auf den Fleischkonsum in zwei Hochschulmensen der ZHAW hat. Dazu standen die Informationen zu den 17900 während 12 verkauften Mensamenüs zur Verfügung. Jedes dieser Menüs wurde als fleischhaltig oder nicht-fleischhaltig klassifiziert. Das Alter wurde für diese Untersuchung in eine kategoriale Variable mit vier Levels übersetzt: AK1: 15-25 Jahre, AK2: 26-34 Jahre, AK3: 35-49 Jahre, AK4: 50-64 Jahre). Die Identität der kaufenden Person stand via die Campus-Card-ID in anonymisierter Form für jeden Menükauf zur Verfügung.

Die Analysen wurden in R, Version 4.2.2, durchgeführt (R Core Team 2024). Es wurde ein logistisches GLMM mit Geschlecht, Status und Alterskategorie als unabhängigen Variablen und der Wahrscheinlichkeit der Wahl eines Fleischgerichtes als abhängiger Variable gerechnet. Es wurden im vollen Modell neben den einfachen Variablen auch die Interaktionen Geschlecht:Alterskategorie und Geschlecht:Status berücksichtigt. Die Personen-Identität wurde als random factor spezifiziert. Das Modell wurde mit dem Befehl 'glmmTMB' aus dem Package 'glmmTMB' (Brooks et al. 2017) gefittet. Es wurde mittels des 'drop1'-Befehls (Base R) und χ^2 -Statistik schrittweise vereinfacht, bis das minimal adäquate Modell erreicht war. Es wurde mittels des Befehls 'check_overdispersion' im Package 'performance' (Lüdtke et al. 2021), v.0.10.2, auf Overdispersion getestet und es gab eine unproblematische minimale Underdispersion (dispersion ratio = 0.976). Die weiteren Modellvoraussetzungen wurden mittels Residualplots ebenfalls mit 'plotResiduals' im Package 'DHARMa' (Hartig et al. 2024) visuell inspiziert und waren nicht schwerwiegend verletzt. Mit dem 'r2'-Befehl aus dem Package 'performance' wurde schliesslich das «conditional» und das «marginal» R^2 ermittelt.

Ergebnisse

Gemäss dem minimal adäquaten Modell haben das Geschlecht ($p < 0.001$) und der Status ($p = 0.026$) einen signifikanten Einfluss auf die Wahrscheinlichkeit ein Fleischgericht in der Mensa zu wählen. Die Alterskategorie und die beiden getesteten Interaktionen waren dagegen nicht signifikant. Nach den Modellergebnissen war die Wahrscheinlichkeit, ein fleischhaltiges Gericht zu kaufen, bei Männern deutlich höher als bei Frauen und bei Studierenden etwas niedriger als bei Mitarbeitenden (Tab. 1). Das Modell hatte ein «marginal» R^2 von 0.040 und ein «conditional» R^2 von 0.338, d.h. der Einfluss von Geschlecht und Status war im Vergleich zu den Unterschieden zwischen den Individuen gering.

Abb. 1. Modellierte Wahrscheinlichkeiten, ein Fleischgericht zu wählen, in Abhängigkeit von Geschlecht und Status.

		Geschlecht	
		Frau	Mann
Status	Mitarbeitende	46.0%	68.6%
	Studierende	41.2%	64.2%

Quellen

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, A.J., Maechler, M. & Bolker, B.M. (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), 378-400. doi: 10.32614/RJ-2017-066.

Hartig, F (2024) *DHARMA: Residual diagnostics for hierarchical (multi-level / mixed) regression models. R package version 0.4.7.* <<https://CRAN.R-project.org/package=DHARMA>>.

Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P. & Makowski, D. (2021) performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software* 6(60): 3139. <https://doi.org/10.21105/joss.03139>.

R Core Team. (2024) *R: A language and environment for statistical computing. R Foundation for Statistical Computing.* <http://www.r-project.org/>.