Statistik-Übung 4 – Multiple Regression

Beispielhafter Methoden- und Ergebnisteil

Methoden

Ziel war es zu ermitteln, welchen Einfluss verschiedene Umweltfaktoren auf den Pflanzenartenreichtum auf 10 m^2 grossen Probeflächen von Steppenrasen in der Ukraine haben. Dazu wurden für 199 solche Probeflächen der Artenreichtum und 18 verschiedene Umweltvariablen (siehe Tab. 1) ermittelt.

Die Analyse wurde in R, Version 4.6.1, durchgeführt (R Core Team 2026). Zunächst wurde eine Pearson-Korrelationsmatrix zwischen allen 18 möglichen Prädiktorvariablen gerechnet. Von Paaren von Variablen, die mit $|r| \geq 0.7$ korreliert waren, wurde jeweils die Variable beibehalten, die als ökologisch aussagekräftiger angesehen wurde (Tab. 1). Somit verblieben 16 Prädiktoren für das volle Modell. Aus diesen wurde ein multiples lineares Regressionsmodell ohne quadratische Terme und Interaktionen erstellt. Da die abhängige Variable (Artenreichtum) zwar eine Zählvariable ist, ihr Mittelwert (40.2) aber weit von 0 entfernt war, wurde angenommen, dass die Verletzung der Modellvoraussetzungen nur minimal sind.

Das minimal adäquate Modell wurde mittels Multimodel-Inferenz (Burnham & Anderson 2002) ermittelt. Dazu wurden mit der «dredge»-Funktion im Package MuMIn (Bartón 2024) analysiert alle möglichen im vollen Modell enthaltenen Teilmodelle erzeugt und mittels AICc verglichen. Da sich das beste Modell um weniger als Δ AICc = 2 vom nächstbesten unterschied, wurde ein gemitteltes Modell erstellt und die Bedeutung der einzelnen Prädiktoren mittels kumulierter Akaike weights (*Variable importance*; Befehl «sw» in MuMIn) ermittelt. Die Modellvoraussetzungen (Varianzhomogenität, Normalität der Residuen, Linearität) wurden optisch mit Residualplots sowohl für das volle Modell wie auch das Modell mit den Variablen mit *Variable importance* ≥ 0.5 getestet und keinerlei nennenswerte Abweichungen gefunden.

Ergebnisse

Von den 16 Prädiktoren des vollen Modells, war Heat_index mit einer *Variable importance* von nahezu 1.00 der bedeutsamste und weitere vier hatten eine *Variable importance* ≥ 0.50 (Tab. 1). Dabei hatten Heat_index, Litter und CN_ratio einen negativen Einfluss auf die Artenzahl, CaCO3 und grazing_intensity dagegen einen positiven (Tab. 1). Das Modell mit den fünf Variablen mit einer *Variable importance* ≥ 0.50 erzielte eine erklärte Varianz ($R^2_{adj.}$) von 0.19.

Tab. 1. Das aufgrund von Multimodel-Inferenz gemittelte Modell für die Erklärung des Artenreichtums ukrainischer Steppenrasen auf 10 m^2. Die Prädiktoren sind nach absteigender *Variable importance* aufgeführt. Die angegebenen Schätzwerte basieren auf dem *Full average*. Ferner sind jene Variablen gelistet, die im vollen Modell wegen hoher paarweiser Korrelation nicht berücksichtigt wurden, incl. korrelierter Variable und Richtung der Korrelation.

Prädiktor	***Variable importance***	**Schätzwert**	**Korrelierte Variable(n)**
(Achsenabschnitt)		16.353	
Heat_index	1.00	-12.964	
Litter	0.91	-0.099	
CaCO3	0.81	0.181	
CN_ratio	0.77	-0.535	
Grazing_intensity	0.63	0.772	
Temperature	0.49	0.203	Altitude (-)
Precipitation	0.43	0.031	
Stones_and_rocks	0.40	-0.038	
Conductivity	0.33	0.005	
Microrelief	0.32	0.017	
Fine_soil	0.30	-0.020	
Gravel	0.30	-0.021	
pH	0.28	-0.178	
C_org	0.28	0.058	N_tot (+)
Temperature range	0.28	-0.015	
Inclination	0.26	0.003	

Quellen

Bartoń K (2024) *MuMIn: Multi-Model Inference*. R package version 1.48.4, https://CRAN.R-project.org/package=MuMIn.

Burnham, K.P., Anderson, D.R. (2002) *Model selection and multimodel inference – a practical information-theoretic approach*. 2nd ed. Springer, New York, US: 488 pp.

R Core Team. (2026). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://doi.org/10.32614/R.manuals