Onlinebeispiel 8.5. Wachstumsrate einer Erdkrötenpopulation an einem Amphibienschutzzaun

Kapitel 8.3.2 aus Henle, K., A. Grimm-Seyfarth & B. Gruber: Erfassung und Analyse von Tierpopulationen. Ulmer Verlag

Annegret Grimm-Seyfarth

22.04.2025

In diesem Beispiel zeigen wir eine Regressionsanalyse in R. Für klassische Regressionsanalysen mittels *lm* und *glm* sind keine weiteren Pakete nötig. Als Datensatz nutzen wir das Beispiel aus 8.2. Wir laden hier die Daten noch einmal neu ein, mit leichten Anpassungen.

Daten einladen

Am Wasserwerk Hedem, Preußisch-Oldendorf, wurden vom Kreis Minden von 1996 bis 2012 Erdkröten an den Amphibienschutzzäunen erfasst. Die Bestandsentwicklung ist in untenstehender Tabelle dargestellt (Seyring et al. 2024). Zusätzlich kreieren wir die Zeitspalte, die die Datenlücke von 2008 berücksichtigt.

```r
kroeten <- data.frame(Jahr = c(1996:2007,2009:2012),
                      t = c(0:11,13:16),
    Anzahl = c(311,564,257,645,797,589,412,177,398,138,235,265,315,138,112,37))
kroeten
```

```
##    Jahr  t Anzahl
## 1  1996  0    311
## 2  1997  1    564
## 3  1998  2    257
## 4  1999  3    645
## 5  2000  4    797
## 6  2001  5    589
## 7  2002  6    412
## 8  2003  7    177
## 9  2004  8    398
## 10 2005  9    138
## 11 2006 10    235
## 12 2007 11    265
## 13 2009 13    315
## 14 2010 14    138
## 15 2011 15    112
## 16 2012 16     37
```

Alternativ hätten wir einen Datensatz kreieren können, in dem das fehlende Jahr 2008 als NA dargestellt wird. Dies macht für die Analyse keinen Unterschied.

```r
kroeten2 <- data.frame(Jahr = c(1996:2012),
                       t = 0:16,
    Anzahl = c(311,564,257,645,797,589,412,177,398,138,235,265,NA,315,138,112,37))
kroeten2
```

```
##    Jahr  t Anzahl
## 1  1996  0    311
## 2  1997  1    564
## 3  1998  2    257
## 4  1999  3    645
## 5  2000  4    797
## 6  2001  5    589
## 7  2002  6    412
## 8  2003  7    177
## 9  2004  8    398
## 10 2005  9    138
## 11 2006 10    235
## 12 2007 11    265
## 13 2008 12     NA
## 14 2009 13    315
## 15 2010 14    138
## 16 2011 15    112
## 17 2012 16     37
```

Logarithmisches Wachstumsmodell

Wir logarithmieren die Populationsgröße und lassen sie von der Zeit abhängen. Dabei ist wichtig, dass das fehlende Jahr berücksichtigt wird (ein Zeitschritt in `t` wird übersprungen).

```r
mod1 <- lm(log(Anzahl)~t, data=kroeten)
# Zusammenfassung des Modells
summary(mod1)
```

```
##
## Call:
## lm(formula = log(Anzahl) ~ t, data = kroeten)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04639 -0.49609  0.08778  0.37963  0.75902
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.45064    0.25953   24.86 5.55e-13 ***
## t           -0.11208    0.02823   -3.97   0.0014 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5583 on 14 degrees of freedom
## Multiple R-squared:  0.5296, Adjusted R-squared:  0.496
## F-statistic: 15.76 on 1 and 14 DF,  p-value: 0.001396
```

Das Modell ist signifikant (gesamt-p-Wert: 0,001, in diesem Fall kann der p-Wert aus der summary-Funktion genutzt werden) und zeigt einen Wachstumsfaktor von -0,11. Eine Nicht-Beachtung des fehlenden Jahres würde jedoch einen stärkeren Trend ergeben.

Granger-Kausalitätstest (Granger-Causality-Test)

Eine weitere parametrische Analysemethode, wenn man mehrere Zeitreihen vorliegen hat, ist der Granger-Kausalitätstest oder Granger-Causality-Test (Granger 1969). Er stammt eigentlich aus ökonomischen Analysen, kann aber auch für ökologische Analysen angewendet werden.

Wir benötigen hierzu das R-Paket lmtest (Zeileis & Hothorn 2002). Außerdem nutzen wir das R-Paket dplyr (Wickham et al. 2023) zur Datenaufbereitung.

```r
# check.packages function: install and load multiple R packages.
# Function from: https://gist.github.com/smithdanielle/9913897
check.packages <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg, dependencies = TRUE)
  sapply(pkg, require, character.only = TRUE)
}

# benoetigte R pakete
pakete <- c("lmtest","dplyr")

# Pruefe und installiere
check.packages(pakete)
```

```
## lmtest  dplyr
##   TRUE   TRUE
```

Zur Demonstation erstellen wir eine Zählspalte in unserem Krötendatensatz, die dem gleichen Wachstumsmodell folgt. Dazu benötigen wir zunächst die Koeffizienten des Wachstumsmodells, sowie den Zufallseffekt (geschätzte Fehlerstreuung):

```r
alpha <- coef(mod1)[1]
beta  <- coef(mod1)[2]
sigma <- summary(mod1)$sigma  # Residuen-Std auf Log-Skala
```

Das Modell hieße $log(Anzahl) = a + b \cdot t$, folglich berechnet sich die Anzahl mit:

$Anzahl = exp(a + b \cdot t)$

Als Zeitpunkte nutzen wir t aus dem Datensatz `kroeten2`.

```r
t_new <- kroeten2$t
n <- length(t_new)
```

Stochastische Simulation der Anzahl:

```r
set.seed(123)

log_Anzahl_sim <- rnorm(n, mean = alpha + beta * t_new, sd = sigma)
Anzahl_sim <- exp(log_Anzahl_sim)

hypodata_sim <- data.frame(t = t_new, Anzahl_sim = round(Anzahl_sim,0))
head(hypodata_sim)
```

```
##   t Anzahl_sim
## 1 0        463
## 2 1        498
## 3 2       1208
## 4 3        470
## 5 4        435
## 6 5        942
```

Bringen wir die Daten zusammen, fügen eine einjährige Verzögerung ein (`lag = 1`) und streichen Zeilen mit NA-Werten heraus.

```r
kroeten3 <- kroeten2 %>%
  mutate(Anzahl_sim = hypodata_sim$Anzahl_sim,
         Anzahl_sim_lag1 = lag(Anzahl_sim, n = 1),
         Anzahl_lag1 = lag(Anzahl, n = 1)) %>%
  filter(!is.na(Anzahl_sim_lag1) & !is.na(Anzahl_lag1)  & !is.na(Anzahl))

kroeten3
```

```
##    Jahr  t Anzahl Anzahl_sim Anzahl_sim_lag1 Anzahl_lag1
## 1  1997  1    564        498             463         311
## 2  1998  2    257       1208             498         564
## 3  1999  3    645        470            1208         257
## 4  2000  4    797        435             470         645
## 5  2001  5    589        942             435         797
## 6  2002  6    412        418             942         589
## 7  2003  7    177        143             418         412
## 8  2004  8    398        176             143         177
## 9  2005  9    138        180             176         398
## 10 2006 10    235        409             180         138
## 11 2007 11    265        226             409         235
## 12 2010 14    138         97             157         315
## 13 2011 15    112        320              97         138
## 14 2012 16     37        139             320         112
```

Der Granger-Kausalitätstest hat folgende Hypothesen:

Nullhypothese (H0): Die Zeitreihe x ist keine Granger-Ursache für die Zeitreihe y

Alternativhypothese (HA): Die Zeitreihe x ist eine Granger-Ursache für die Zeitreihe y

Der Begriff „Granger-Kausalität“ bedeutet, dass die Kenntnis des Wertes der Zeitreihe x zu einem bestimmten Zeitpunkt nützlich ist, um den Wert der Zeitreihe y zu einem späteren Zeitpunkt vorherzusagen. Um einen Granger-Kausalitätstest in R durchzuführen, können wir die Funktion `grangertest()` aus dem R-Paket lmtest verwenden, die folgende Syntax verwendet:

`grangertest(x, y, order = 1)`

wobei:

`x`: Die erste Zeitreihe

`y`: Die zweite Zeitreihe

`order`: Die Anzahl der Zeitverschiebungen (Lags), die in der ersten Zeitreihe verwendet werden sollen. Der Standardwert ist 1 (also Einfluss aus dem Jahr t-1).

Der Test wird folglich gerichtet aufgerufen, die Reihenfolge bestimmt also die Kausalität. Wir wollen hier wissen, ob die Erdkrötenzahl die Anzahl hypothetischer Kröten vorhersagen kann:

```r
grangertest(Anzahl_sim ~ Anzahl, order = 1, data = kroeten3)
```

```
## Granger causality test
##
## Model 1: Anzahl_sim ~ Lags(Anzahl_sim, 1:1) + Lags(Anzahl, 1:1)
## Model 2: Anzahl_sim ~ Lags(Anzahl_sim, 1:1)
##   Res.Df Df      F  Pr(>F)
## 1     10
## 2     11 -1 5.4055 0.04241 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wir bekommen hier einen signifikanten Zusammenhang der beiden Zeitreihen mit der Verzögerung (lag) von einem Jahr. Achtung: Der Granger-Kausalitätstest kann durch vorhandene Trends verunschärft werden (was hier auch ein Grund dafür sein könnte, dass keine starke Signifikanz vorliegt). Wir können hier daher also auch zwei Zeitreihen vergleichen, wenn keinerlei Trends vorliegen.

Literaturverzeichnis

Granger, C.W.J. (1969): Investigating causal relations by econometric models and cross-spectral methods. — Econometrica 37(3): 424–438.

Seyring, M., Henle, K., Barth, B. et al. (2024): Empfehlungen für ein bundeseinheitliches Vorgehen bei der Erfassung von Amphibien-Schutzzaun-Daten zur Unterstützung von Bestandstrendanalysen. — S. 114-–133 in: Henle, K., Pogoda, P., Podloucky, R. et al. (Hrsg.): Neue Methoden der Feldherpetologie. — Chimaira, Frankfurt/M. (Mertensiella 32).

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. (2023): dplyr: A grammar of data manipulation. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr.

Zeileis, A., Hothorn, T. (2002): Diagnostic checking in regression relationships. — R News 2(3): 7.-10.