Fitting mixed models and selecting among them

QERM 514 - Homework 6

8 May 2020

Background

This week's homework assignment focuses on fitting and evaluating linear mixed models. In particular, you will consider different forms for a stock-recruit relationship that describes the density-dependent relationship between fish spawning biomass in "brood year" t (S_t) and the biomass of fish arising from that brood year that subsequently "recruit" to the fishery (R_t).

Ricker model

The Ricker model (Ricker 1954) is one of the classical forms for describing the stock-recruit relationship. The deterministic form of the model is given by

$$R_t = S_t \exp\left[r\left(1 - \frac{S_t}{k}\right)\right]$$

where r is the intrinsic growth rate and k is the carrying capacity of the environment. In fisheries science, the model is often rewritten as

$$R_t = aS_t \exp\left(-bS_t\right)$$

where $a = \exp r$ and $b = r/k$. We can make the model stochastic by including a multiplicative error term $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$, such that

$$R_t = aS_t \exp\left(-bS_t\right) \exp(\epsilon_t)$$

This model is clearly non-linear, but we can use a log-transform to linearize it. Specifically, we have

$$\begin{aligned}
\log R_t &= \log a + \log S_t - bS_t + \epsilon_t \\
&\Downarrow \\
\log R_t - \log S_t &= \log a - bS_t + \epsilon_t \\
&\Downarrow \\
\log(R_t/S_t) &= \log a - bS_t + \epsilon_t \\
&\Downarrow \\
y_t &= \alpha - \beta S_t + \epsilon_t
\end{aligned}$$

where $y_t = \log(R_t/S_t)$, $\alpha = \log a$, and $\beta = b$.

Data

The data for this assignment come from 21 populations of Chinook salmon (*Oncorhynchus tshawytscha*) in Puget Sound. The original data come from the NOAA Fisheries Salmon Population Summary (SPS) database, which was subsequently cleaned and summarized for use in a recent paper by Bal et al. (2019). The data are contained in the accompanying file `ps_chinook.csv`, which contains the following columns:

- `pop`: name of the population
- `pop_n`: integer ID for population (1-21)
- `year`: year of spawning
- `spawners`: total number of spawning adults (1000s)
- `recruits`: total number of surviving offspring that "recruit" to the fishery (1000s)
- `prop_wild`: the proportion of spawners that are of wild origin (ie, not hatchery fish)

Problems

As you work through the following problems, be sure to show all of the code necessary to produce your answers. (Hint: You will need to define a new response variable before you can do any model fitting.)

a) Plot the number of recruits by population (y) against the number of spawners by population (x), and add a line indicating the replacement level where recruits = spawners. Describe what you see.

b) Fit the following model and report your estimates for α and β. Also report your estimate of σ_ϵ^2. Based on the R^2 value, does this seem like a promising model?

$$\log(R_{i,t}/S_{i,t}) = \alpha - \beta S_{i,t} + \epsilon_{i,t}$$
$$\epsilon_{i,t} \sim \mathrm{N}(0, \sigma_\epsilon^2)$$

c) Fit the following model and report your estimates for α, each of the δ_i, and β. Also report your estimate of σ_ϵ^2 and σ_δ^2. Based on the R^2 value, how does this model compare to that from part (b)?

$$\log(R_{i,t}/S_{i,t}) = (\alpha + \delta_i) - \beta S_{i,t} + \epsilon_{i,t}$$
$$\delta_i \sim \mathrm{N}(0, \sigma_\delta^2)$$
$$\epsilon_{i,t} \sim \mathrm{N}(0, \sigma_\epsilon^2)$$

d) Fit the following model and report your estimates for α, each of the η_i, and β. Also report your estimate of σ_ϵ^2 and σ_η^2. Based on the R^2 value, how does this model compare to that from part (c)?

$$\log(R_{i,t}/S_{i,t}) = \alpha - (\beta + \eta_i)S_{i,t} + \epsilon_{i,t}$$
$$\eta_i \sim \mathrm{N}(0, \sigma_\eta^2)$$
$$\epsilon_{i,t} \sim \mathrm{N}(0, \sigma_\epsilon^2)$$

e) Fit the following model and report your estimates for α, each of the δ_i, β, and each of the η_i. Also report your estimate of σ_ϵ^2, σ_δ^2, and σ_η^2. Based on the R^2 value, how does this model compare to that from part (d)? (Hint: Refer back to the beginning of Lab 6 for how to fit uncorrelated random effects for both intercept and slope.)

$$\log(R_{i,t}/S_{i,t}) = (\alpha + \delta_i) - (\beta + \eta_i)S_{i,t} + \epsilon_{i,t}$$
$$\delta_i \sim \mathrm{N}(0, \sigma_\delta^2)$$
$$\eta_i \sim \mathrm{N}(0, \sigma_\eta^2)$$
$$\epsilon_{i,t} \sim \mathrm{N}(0, \sigma_\epsilon^2)$$

f) Based on the 3 models you fit in parts (c - e), test whether or not there is data support for including a random effect for population-level intercepts. Also test whether or not there is data support for including a random effect for population-level slopes. Make sure to specify your null hypothesis for both of the tests.

g) Now fit the following model and report your estimates for α, each of the δ_i, β, each of the η_i, and each of the γ_t. Also report your estimate of σ_ϵ^2, σ_δ^2, σ_γ^2, and σ_η^2. Based on the R^2 value, how does this model compare to that from part (d)?

$$\log(R_{i,t}/S_{i,t}) = (\alpha + \delta_i + \gamma_t) - (\beta + \eta_i)S_{i,t} + \epsilon_{i,t}$$
$$\delta_i \sim \mathrm{N}(0, \sigma_\delta^2)$$
$$\gamma_t \sim \mathrm{N}(0, \sigma_\gamma^2)$$
$$\eta_i \sim \mathrm{N}(0, \sigma_\eta^2)$$
$$\epsilon_{i,t} \sim \mathrm{N}(0, \sigma_\epsilon^2)$$

h) Conduct a diagnostic check of the model you fit in (g) to evaluate the adequacy of the model assumptions. Do you see any cause for concern?