Analysis of Environmental Data – Reading questions week 9

Olivia Dinkelacker

Q1 (1 pt.): Briefly (1 - 2 short paragraphs) describe at least two tradeoffs between the customized ML methods and the canned methods.

The canned methods use optimization algorithms that are faster and less likely to encounter numerical problems. Sometimes these methods even choose appropriate starting parameters for us.

If a standard method is used, it is enough to generally state in the Method section that linear regressions were used. A non-standard method needs more explanation which can quickly repel the reader because it seems very complicated – although it actually is the more appropriate and simpler than a standard model.

Q2 (1 pt.): Briefly (1 - 2 sentences) describe each of the four key assumptions of the general linear modeling approach.

We need **independent observations**, so we are able to multiply the observations.
We need **constant variance**, where the stochastic part (noise) is constant around the deterministic average > no megaphone shape > wouldn't be constant.
We need a **fixed x,** meaning that there are no measurement errors in our predictor variables.
Lastly, we need a **normal distribution** of the residuals (the difference between predicted and observed value).

Q3 (1 pt.): Explain how the normality assumption can be met in a general linear model, even if the response variable is not normally-distributed. (1 - 2 paragraphs)

The normal distribution does not refer to the values for each variable, but to the variance around the expected values, the residuals. The differences between the predicted and observed values have to be normally distributed.