

STATISTICAL RETHINKING 2023

WEEK 9

The data in `data(Achehunting)` are 14,364 individual hunting trips by 147 men among the indigenous Aché of Paraguay. Each trip has recorded the hunter's identity (anonymized id), hunter's age at the time of the trip, the duration in hours of the trip, and the kilograms of meat returned. Many of the trips have missing values for duration.

1. In this first problem, estimate the influence of age on the probability of trip success. Define “success” as a trip that returns any non-zero amount of meat. Ignore individual hunter identities for now. You can use any functional relationship for age and success that you think is sensible, but be sure to justify it and check it against posterior predictions.

2. Now incorporate individual hunter varying effects into your analysis from the previous problem. Allow the influence of age on success to vary by each individual. How much variation in success is explained by individuals and how much by age?

3. Now include trip duration in the model from the previous problem. The focus is still the causal relationship between age and success, but duration is a potential competing cause or a mediator, so it can be useful to include it for either increased efficiency (competing cause) or to study how variation in trip durations mediate differences in hunting returns. Since there are many missing values for trip duration, compare (1) a complete case analysis, which drops all the trips with missing durations, to (2) an analysis which uses all 14,364 trips and imputes missing duration values where necessary. Use any functional relationship between duration and success that you think is sensible, but be sure to justify it and check it against posterior predictions.

4-OPTIONAL CHALLENGE. Repeat the analysis in problem 3 above for the measured kilograms of meat returned on each trip, not just a 0/1 indicator of success. Is the relationship between age and kilograms returned different from the relationship between age and success? Why or why not? Hint: The tricky part of this problem is modeling the outcome variable.