STATISTICAL RETHINKING 2023
WEEK 7

This homework set reuses the code from the lecture, but it will be challenging because it requires thinking about how to stratify by weird variables like number of children and age. Go slow and know that if you finish any of it, you are my hero. As hard as these problems are, real research is even harder.

1. The data in data(bangladesh) are 1934 women from the 1989 Bangladesh Fertility Survey. For each woman, we know which district she lived in, her number of living.children, her age.centered, whether she lived in an urban center, and finally whether or not she used contraception (use.contraception).

In the lecture, we estimated the total effect of urban residence on contraceptive use. Using the DAG from lecture, or your own version if you prefer, now estimate only the direct effect of urban residence on contraceptive use.

2. Using the same DAG as before, now estimate the effect of number of surviving children on contraceptive use. Obviously contraceptive use can also influence the number of surviving children. But leave that issue aside for the moment (it will return in the optional challenge further down).

3. Now let the causal effect of children vary by district. Incorporate this new district feature into the same multivariate prior that contains the urban/rural features. How much do districts vary in how surviving children are associated with contraceptive use?

4-OPTIONAL CHALLENGE. Write a synthetic data simulation for this problem set, using the relationships in the DAG. The simulation should include district, urban, children, age, and (reported) contraceptive use. The best, and hardest, option is to make a dynamic population simulation, in which individuals age and produce children. In such a model, reciprocal causality between children and contraceptive use could exist. Whatever you decide, use a cross-section of the synthetic population to test your statistical models. When can they recover true effects?