

Using Code to Model What Statistics Could Not Capture

Format: Why CS Common App Short Answer | Level: College Application | ~250 words

In my AP Statistics class junior year, we spent three weeks on linear regression. The model I was building was supposed to predict student performance based on study hours, but every time I added a behavioral variable, sleep quality, anxiety level, phone use, the model's predictive power dropped rather than improved. The residuals were not random; they were patterned in a way that suggested the relationship I was trying to model was not linear, and linear regression had no way to tell me that.

I looked up alternatives and found simulation. I built a simple agent-based model in Python where each simulated student had a set of attributes and a decision rule for how to allocate time between studying and sleeping under deadline pressure. The simulation produced behavior patterns that matched the actual data far better than the regression had, not because it was more mathematically sophisticated but because it modeled the process rather than the correlation.

That experience changed how I think about the relationship between a question and the tool used to answer it. Statistics tells you what happened. Simulation tells you why it might have happened, and lets you test whether your explanation produces the same outcome. I want to study computer science because it is the discipline that builds the tools for the second kind of question: not just measuring the world as it is, but constructing models of the processes that produced it.

