AMS 5

BUILDING A STATISTICAL MODEL

What we have seen so far

- Data Gathering : how to collect information to reduced your uncertainty about something on interest
- Descriptive Methods: summarize factuals, i.e. the data you have.
- Probability: the mathematical language to quantify uncertainty.

Factuals vs Counterfactuals

- Factuals: Data you have
- Counterfactuals: Data you wish to had (e.g. what would have happened to that patient if he had received the drug instead of the placebo?)

Factual dataset + counterfactual dataset = no – uncertainty dataset.

IDEA: Use the information in the factual dataset to guess at the no – uncertainty data set (reasoning from the part to the whole ⟶ INDUCTION)

HOW: By relating the factual to the counterfactual.

A Statistical model is a mathematical story that relates the data you have to the data you wish you had.

Counterfactuals

1. Parametric Inference

There is a population you'd like to summarize, for instance the mean education level of a bunch of people (unknown parameter) and because you cannot afford to do a complete census you take a sample.

Example: I am interested in the quality of hospital care of about 300,000 patients with a diagnosis of a heart attack. I cannot afford to assess all patients, and therefore I have taken a sample of 4000 of them (factual dataset). The remaining dataset of 296,000 patients is the counterfactual dataset.

Counterfactuals

2. Causal Inference

You'd like to measure the effect caused by some treatment of interest to you, so you get some people and assign some of them to a treatment group and some to a control group.

Example: I have a drug for lowering the blood pressure of hypertensive patients. To assess its effectiveness I find 120 hypertensive patients and I randomly assign 60 of them to a treatment group (who get the new drug) and 60 to a control group (who get a placebo). I measure their blood pressure before and after 12 weeks on the drug or placebo, and compute the change for each person. Lets suppose that we have somebody in the treatment group. The factual for this person is how much his/her blood pressure changed in the 12 weeks, while the counterfactual for this person is how much his blood pressure changed in the 12 weeks if we had given him the placebo.

Counterfactuals

3. Prediction

You have some data on how some interest to you has come out in the past, and you'd like to forecast (predict) how it will come out in the future, or you'd like to guess at the unobserved value of some variable for an individual, on the basis of that person's observed value on another variable.

Example: I am trying to make a business decision that depends on the wholesale price of gasoline in LA next year. I have weekly data on gasoline prices in the past 30 years (factual data), and my task is to use these past data and any other information that I have to forecast next year's weekly gasoline prices (counterfactual data).