# Causal Models

# Causal Modeling

- How can we create a causal model?
  - This is a huge topic
  - After this course, you can go on to learn about identification strategies, simultaneous equation modeling, do calculus, etc.
- A lot of methods stem from counterfactual theories of Neyman, Rubin, etc.
  - This is a human-centric approach
  - It’s sufficient for most data science purposes
- Causality is challenging to reason about.
  - I’ll try to give you a little intuition to get you started.

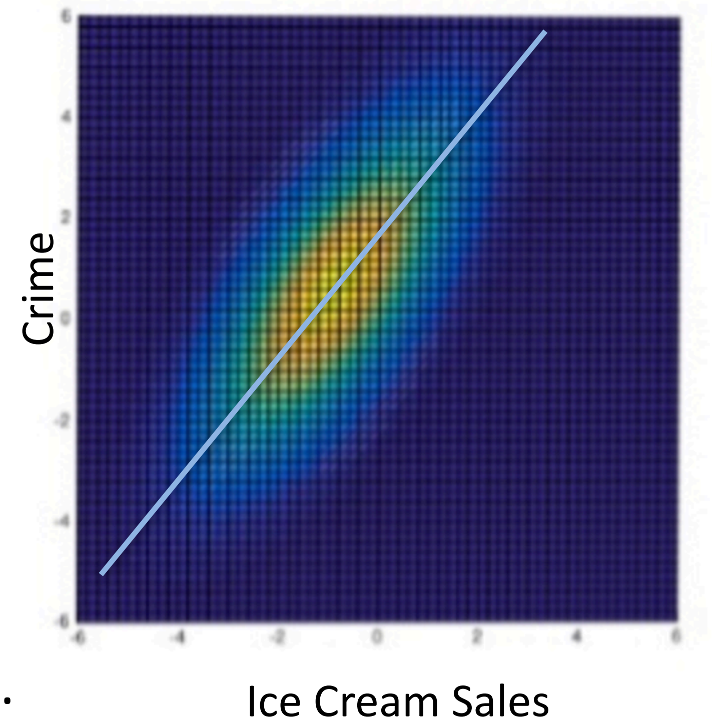
# Assuming Causality

- First, remember that causality is an extra assumption on top of our population model.
- Say our model is $y = \beta_0 + \beta_1 x + u$.
  - This is a way of describing a joint distribution between x and y.
- We can then introduce a manipulation, a change in x.
  - For example, a differential change in x, dx.
- Taking the partial, we have $\frac{\partial y}{\partial x} = \beta_1$
- As long as $\frac{\partial u}{\partial x} = 0$
- 
- We have a causal interpretation as long as the error term doesn't change as we manipulate x.
- The joint distribution doesn't tell us anything about this
  - It's just about relative occurrences of x and y in a static sense.

- How can we assert that u doesn’t change?
- It comes down to omitted variables.
- A causal modeler believes that all the causes are out there, even if we can’t measure them.
- Imagine that you start writing down all the variables that could affect your outcome.
- For example, remember the example of crime predicted by ice cream sales.

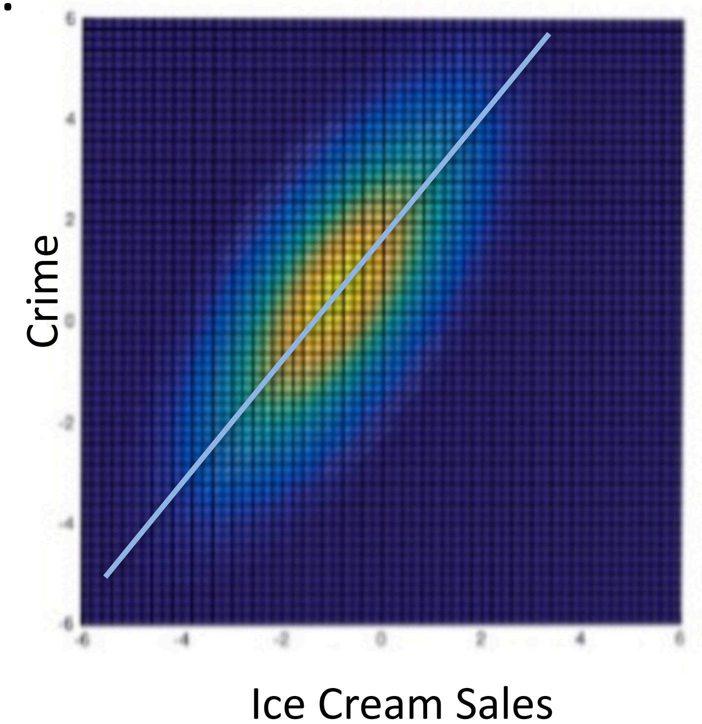
# Causes of Crime

- We look at the joint distribution and we see a strong positive relationship
- But we know that there are omitted variables.
- Temperature is one important omitted variable (we might also call it a confounding variable).
- As we move towards higher ice cream sales on the right of the plot, we're also looking at hotter days
  - And there is just more crime on hotter days.
- So let's include temperature in our model:
- crime = $\beta_0 + \beta_1$ sales + $\beta_2$ temperature + u.



# Causes of Crime

- Let's keep going. Think of more variables.
  - crime = $\beta_0$ + $\beta_1$ sales + $\beta_2$ temperature + $\beta_3$ daylight_hours + $\beta_4$ police_per_capita + $\beta_5$ mean_income + … + u.
- If you could write down all the factors that affect crime, eventually there wouldn't be any more error
  - Or at least the error would truly be entirely random
- Also, by including these variables in the model, we can hold them constant (ceteris paribus)
  - Now $\beta_1$ is the effect of sales, holding temperature and these other variables constant.
  - We believe that we're really modeling the causal effect.



# The Causal Perspective

- The central problem of causal modeling:
- True causal model:
  - crime = $\beta_0 + \beta_1$ sales + $\beta_2$ temperature + $\beta_3$ daylight_hours + $\beta_4$ police_per_capita + $\beta_5$ mean_income + … + u.
- But we can't measure all the variables.
- This means that all those other factors become part of our error.
  - crime = $\beta_0 + \beta_1$ sales + v
- Where
  - v = $\beta_2$ temperature + $\beta_3$ daylight_hours + $\beta_4$ police_per_capita + $\beta_5$ mean_income + … + u.
- But now cov(sales, v) = $\beta_2$ cov(sales, temperature) + …
- Which is probably bigger than zero, so OLS will not be consistent
- Our estimate for $\beta_1$ will be too high.
- This is called endogeneity bias.
- Next, we will derive an expression for endogeneity bias and we'll show you how to reason about it.