Race Causality Discrimination – Causation as Manipulation

Marcello Di Bello - ASU - Fall 2023 - Week #2

This week we will read a classic article by Paul Holland[1] that outlines, in simple and succinct language, the Rubin's model for studying causality.[2] This is a popular framework that describes how many statisticians thinks about causality.

Holland's article is technical and it is important to be precise about the formal details. So we'll start with a list of questions for testing your understanding.

[1] Holland (1986), Statistics and Causal Inference, *Journal of the American Statistical Society*, 81(396).

[2] Donald Rubin is an Emeritus Professor of Statistics at Harvard University, known for the Rubin Causal Model.

Associational Inference (Sec. 2)

Basic notation and terminology

Universe U and unit u. Give examples.

Response variable Y and attribute A. What is the difference?

What do the expressions $Y(u)$ and $A(u)$ mean? Give examples.

What do the expressions $Y(u) = y$ and $A(u) = a$ mean?

Probability and expectation

What is a probability, in symbol, Pr?

How do we define the expected value of a variable, e.g. $E(Y)$?

What is conditional expected valued, e.g., $E(Y|A)$?

What is a joint distribution, e.g., $Pr(Y = y, A = a)$?

What is a conditional distribution, e.g., $Pr(Y = y|A = a)$?

Rubin's Model (Sec. 3)

What does $Y_t(u)$ mean (or equivalently $Y(u, t)$)?[3]

What does $Y_c(u)$ mean (or equivalently $Y(y, c)$)?

What does the expression $Y_t(u) - Y_c(u)$ mean?

What does the expression $E(Y_t - Y_c)$ mean, call it T?

What does the expression $E(Y_t) - E(Y_c)$ mean?[4]

What does $E(Y_t|S = t) - E(Y_c|S = c)$ mean, call it T_{PF}?

What is Y_S?

What is the average causal effect?

What is the prima facie causal effect T_{PF}?

Why are $E(Y_t)$ and $E(Y_t|S = t)$ not the same? What is the difference?

[3] It is crucial here to understand the difference between the associational model and the causal model. The expression $Y_t(u)$ makes no sense in the associational model because it requires a counterfactual reading.

[4] Are $E(Y_t - Y_c)$ and $E(Y_t) - E(Y_c)$ equivalent? If so, prove it.

Key idea: what is a cause?

For Holland, a cause operates relative to another, competing cause and its effect on an individual unit is measured relative to a response variable of interest. So, we say that a treatment t has a causal effect on unit u, where this causal effect is measured by a response variable Y and is relative to the control c.

In symbols, the causal effect of t (relative to c) on unit u (measured by Y) is expressed by the following difference (p. 947):

$$Y_t(u) - Y_c(u)$$

This difference can hardly be directly observed because the same unit u cannot be both subject to treatement and control (at the same time).[5]. So, under certain conditions, it is more common to study the average causal effect of a treatment t across units in a population U, that is, the difference, call it T:

[5] See the fundamental problem of causal inference, p. 947, also discussed later.

$$E(Y_t - Y_c)$$

To be sure, all that can be observed is just the *prima facie* causal effect (call it T_{PF}):

$$E(Y_t|s = t) - E(Y_c|S = c)$$

Causal inference, at the individual level, can be studied provided certain assumptions hold that allow to derive the causal effect of treatment t on unit u from the *prima facie* causal effect T_{PF}.

Special cases of causal inference (Sec. 4)

The fundamental problem of causal inference arises because

$$Y_t(u) - Y_c(u)$$

cannot be observed. So causal inference, strictly speaking, is impossible. There are workarounds, though: the scientific solution and statistical solution. To test your understanding, answer the following:

What is the scientific solution to the fundemental problem?

What is unit homogeneity assumption? State it formally.

What is the statistical solution to the fundamental problem?

What is the independence assumption? State it formally.

What is the constant effect assumption?

In addition, you should be able to write up a few formal proofs:

Show that unity homogeneity solves the fundamental problem of causal inference.

Show that the independence assumption and constant effect assumption, toghether, solve the fundamental problem of causal inference? [Hint: show that under these assumptions $T = T_{PF}$.]

Show that the constant effect assumption, alone, is not enough to guaranteert that $T = T_{PF}$

Show that, if unity homogeneity holds, then $T = T_{PF}$.

Suppes' probabilistic theory of causality (Sec. 5.3)

The philosopher Patrick Suppes proposed the following probabilistic account of causality:[6]

[6] See Suppes (1970), *A Probabilistic Theory of Causality*, North-Holland Publishing Company

One event C_r is the cause of another E_s if C_r raises the probability of E_s, that is,

$$Pr(E_s|C_r) > Pr(E_s)$$

and, in addition, there is no third event that explains away the probability raising, that is, there is no event D_q such that:

$$Pr(E_s|C_r, D_q) = Pr(E_s|D_q)$$

and

$$Pr(E_s|C_r, D_q) \geq Pr(E_s|C_r).$$

Holland comments this proposal as follows:

> At bottom, Suppes's notion of a genuine cause is simply a correlation between a cause and effect that will not go away by "partialling out" legitimate competing causes. In a sense then for Suppes all genuine causes are only temporarily so as they await the cleverness of the analyst to identify the proper conditioning event that will render null their association with the effect. Although this may, in- deed, describe much informal scientific practice, it does not appear to me to get to the heart of the notion of causation, which, I believe, Rubin's model does (p. 952)[7]

[7] Compare Suppes's model of causation and Rubin's model. Which one (if any) captures what economists Ronald Fryer was attempting to do in his study of racial bias in police shooting? See handout from last week.

What other stisticians said (sec. 6)

Some of the ideas in Rubin's model of causality seem to have been around for a while, though it was not until Rubin that the discipline of statistics reached greater clarity on the question of causality.

Among other things, Rubin's model can help to illustrate a disagreement between Fisher and Neyman:

Explain the disagreement between Fisher and Neyman about the choice of the null hypothesis in experiments with randomized bloack design? Use Rubin's model to explain the difference between $E(Y_t - Y_c) = 0$ and $Y_t(u) - Y_c(u) = 0$, for all $u \in U$.

What can(not) be a cause (Sec. 7)

Holland is clear that not anything can be a cause. Not any event or facts can be a cause. The causal model it presents study causality at the level of individual units u, such as people, households, plots of land. In this setting, the word 'cause' and 'treatment' are used interchangeably. It is crucial that individual units can be **exposed to causes**. Anything individual units cannot be exposed to will not be regarded as causes in any meaningful sense.

A consequence of this view is that stable attributes of individuals, such and race or gender, or even income, cannot be regarded as causes. Individuals cannot be exposed to race or gender, and it is unclear how individual can be exposed to income. Instead, a cure or a school program are clearly things individuals can be exposed to and thus can be regarded as causes.

Interestingly, Holland notes:

> One may view Fisher's (1957) attack on those who used the association between smoking and lung cancer as evi- dence of a "causal link" between them as an example of the difficulty in deciding whether or not smoking is an attribute or a cause (p. 955).

Summary (Sec. 9)

Holland concludes his article (p. 959) by emphasizing three key ideas of the statistical account of causality:

1. The analysis of causation should begin with studying the effects of causes rather than the traditional approach of trying to define what the cause of a given effect is.
2. Effects of causes are always relative to other causes (i.e., it takes two causes to define an effect).
3. Not everything can be a cause; in particular, attributes of units are never causes.[8]

In addition, Holland makes two further points:

1. The difference between the model $(S, Y_t, Y,)$ and the process of observation (S, Y_s).[9]
2. The Fundamental Problem of Causal Inference–only Y_t, or Y_c but not both can be observed on any unit u.

[8] The third idea is summarized in the slogan **no causation without manipulation**.

[9] What does Holland mean when he says that 'it is a great mistake to confuse Y_t or Y_c with Y_S, and yet it is done all the time' (p. 959)?