

Statistics 514: Design of Experiments

Topic 5

Topic Overview

This topic will cover

- Causation
- Randomization Inference

Central Issue

“How do I defend my answers to scientific questions of interest in this situation?”

- How do we avoid what can seem like a numerological ritual?
- Without calculus and measure-theoretic proof, whence intuition?

Sir Francis Bacon (1561-1626)

- 1620 *Novum Organum* response to Aristotle’s reliance on deduction
 - *Aristotle*: “The premises must be the causes of the conclusion, better known than it, and prior to it; its causes, since we possess scientific knowledge of a thing only when we know its cause; prior, in order to be causes; antecedently known, this antecedent knowledge being not our mere understanding of the meaning, but knowledge of the fact as well.”
 - “But chance also and spontaneity are reckoned among the causes: many things are said both to be and to come to be as a result of chance and spontaneity.”
 - *Frosini on Aristotle*: “Even the probability calculus was viewed just as a surrogate, or an approximation, of deterministic laws not yet discovered; the use of probabilities was deemed as an avowal of one’s ignorance.”
- stressed observation along with deduction
- Manipulation replaced “noting and naming”
- *Ideal*:
 - Experimenters remove from their thinking “time-honored illusions and fallacies” or “traditional beliefs and dogmas”
 - make themselves objective with no loyalties to any hypothesis that might blind them to any portion of evidence.

- Correct conclusions would emerge automatically, as dictated by Nature without ambiguity (“course of Nature”)
- *Critiques:*
 - Personal judgment is inextricable from the scientific process – in deciding what to study and deciding what variables are relevant.
 - Ratzch (2000) – “There is no rigorous logical procedure which accounts for the birth of theories or of the novel concepts and connections which new theories often involve. There is no *logic of discovery*.”
 - “Even if we hit on the correct theory [without data], we cannot prove its correctness conclusively [with data].”
 - No room for getting things *approximately* right.

Scale of Evidence

- Typically rely on p -values to quantify much evidence there is to conclude factor has an effect. (“ego” of statistical reasoning)
 - “Superego” of hypothesis testing – if $\alpha = 5\%$, then 0.049 and 0.003 are both significant p -values.
 - “Id” of statistical reasoning seems to point to Bayesian analysis.
- Scientists often take inference one step further and make conclusions in terms of *causation*.
 - Third assumption after uniformity and understandableness of nature.
 - “Every natural event has a cause, and if [just] that cause is duplicated, the event is duplicated.” (Assumption of Finite Causation)
- Physics and economics represent one extreme (relying heavily on internal validity) and belief in the “realism” of their theories.
- Corollary of assumption of finite causation is bias toward simplicity.

Causation

- Typically rely on intuitive definition and leave details to logicians and philosophers
 - “A causal process forms a net whose knots represent the interaction between processes.”
 - “True beliefs lead to successful behavior.”
 - “It is dangerous to muddle up metaphysics with methodology.”

- However, we should be aware that scientists often speak of causation when talking about results of randomized experiments.

– *Postmodern answer*: Causation is whatever an experiment is seeking to find.

- Typical critique: “*Association* does not imply *causation*.”
- Is it a question of what constitutes evidence?

Having it both ways is an irritating feature of “The Blank Slate: [The Modern Denial of Human Nature by Steven Pinker].” Pinker can write, in refutation of the scarecrow theory of violent behavior, “The sad fact is that despite the repeated assurances that ‘we know the conditions that breed violence,’ we barely have a clue,” and then, a few pages later, “It is not surprising, then, that when African-American teenagers are taken out of underclass neighborhoods, they are no more violent or delinquent than white teenagers.” Well, that should give us *one* clue.

Louis Menand, “What Comes Naturally” *The New Yorker*, 11-25-2002

How do Statisticians Treat Causation?

- Mostly negative in tone
- Often includes acknowledgment that notion of causation is central to scientists’ understanding of the world
- Statisticians do not often think in terms of *experiments* to verify assumptions.
 - Experimental design – devising “controlled” conditions for which theory of interest would make a different prediction than would a rival hypothesis.
- **Syllogism of Falsification:**
 - If theory T is true, the data will follow pattern P.
 - The data do not follow predicted pattern P
 - Therefore, theory T is false.
- Problems
 - No derived models (“problem of realism”)
 - Interpretation of small p -values (What does a p -value of 10^{-24} mean?)
 - What about deciding “this needs more study”?
- R. A. Fisher (on step from association to causation): “Make your theories elaborate.”

[Probably thought he was making a joke.]

Bradford Hill (1937, 1965)

Effect has causal interpretation if effect

1. is large (practically significant)
2. is reproduced in independent studies (repeatable)
3. shows a *monotone* relation with “dose” (!?)
4. corresponds to “natural experiment” (generalizable?)
5. behaves appropriately when potential cause is applied, removed, then reinstated (case for crossover design)
6. is consistent with subject-matter knowledge (reduce role of chance)
7. is, for example, predicted by reasonably well-established theory (experiment as verification of what’s already known)

Mosteller and Tukey (1977)

Need at least two of following 3 concepts to support causal relationship

- *Consistency*: all other things being equal, relationship between two variables is consistent across populations in direction and maybe in amount
- *Responsiveness*: changing causal variable leads to change in response variable
- *Mechanism*: there is a step-by-step mechanism leading from cause to effect

Experiments can demonstrate first two. How important is the third?

Causality via Association

Rule

C causes E
implies
 C must happen if E is observed

- Necessary and sufficient condition
- Not appropriate if there are multiple causes
- Not practically useful

Levels of Causality (Cox and Wermuth 2004)

- **Zero-level causality:** statistical association “with clearly established ordering from cause to response.”
- **First-level causality:** looks for explicit connections between causes and effects
- **Second-level causality:** research aiming at disclosing the underlying generating process (looking for *how* and possibly *why*).

Another Notion

Causation is statistical association that cannot be explained as a dependence on other features.

C is a candidate cause of E if

$$P(E|C) > P(E|\text{not } C)$$

but there is no B such that

$$P(E|C \text{ and } B) = P(E|\text{not } C, \text{ and } B)$$

Qualifications

- Assume all possible candidates B observable
- There is a verifiable asymmetry between cause and effect

C occurs before E
 C closer [spatially] than B to E

- Subject-matter knowledge establishes causal ordering (or mechanism)

Often requires non-trivial assumptions

In Practice

“Screen off irrelevant factors until a homogeneous reference class is obtained.”

Compromise?

- Most of the time, “should aim for models that are at least potentially causable.” (“no causes in, no causes out”)
- Interpretability can play a part.
- Easier: “‘Intention-to-treat’ analysis assesses the effect of one treatment compared with another, where the imposition of a treatment carries with it any other changes that the experimental set-up allows.”

Valid inference (and accurate measurement) of intervention may be enough.
Common issue in clinical trials: need reliable regimen compliance.

Causation and Randomization Inference (Rubin 2005)

Elements of Standard Problem

- N objects (Ex: plots of land, individual people, one person at different times)
- Exposed or not to treatment
 - Active treatment (Ex: taking aspirin)
 - Control treatment
- Covariates X
 - usually known before treatment assignment
 - not affected by treatment
- Potential Outcomes Y (Ex: headache pains)
 - $Y(1)$ active treatment
 - $Y(0)$ control
- “Unit-level causal effects” – comparison between $Y_i(1)$ and $Y_i(0)$
 - difference: $Y_i(1) - Y_i(0)$
 - ratio: $Y_i(1)/Y_i(0)$

Units	Covariates X	Potential outcomes		Unit-level Causal effects	Summary Causal effects
		Treatment $Y(1)$	Control $Y(0)$		
1	X_1	$Y_1(1)$	$Y_1(0)$	$Y_1(1)$ vs. $Y_1(0)$	Comparison of $Y_1(1)$ vs. $Y_1(0)$ for a common set of units
\vdots	\vdots	\vdots	\vdots	\vdots	
i	X_i	$Y_i(1)$	$Y_i(0)$	$Y_i(1)$ vs. $Y_i(0)$	
\vdots	\vdots	\vdots	\vdots	\vdots	
N	X_N	$Y_N(1)$	$Y_N(0)$	$Y_N(1)$ vs. $Y_N(0)$	

“Fundamental Problem of Causal Inference”

- Can’t observe both $Y_i(1)$ and $Y_i(0)$
- Can summarize effects measured by comparing distribution of $Y_i(1)$ and $Y_i(0)$
 - **Important:** need common set of units (Ex: females)

Assumptions

- SUTVA assumption (Stable Unit Treatment Value Assumption)
 - No interference between units: $Y_i(1)$ and $Y_i(0)$ not affected by what other unit received
 - No hidden treatments: if i receives treatment 1, will see $Y_i(1)$
- Scientific assumption – treated outcome not affected by how (or if) we attempt to learn about it
 - Science doesn't change if it's done at Purdue, Indiana U., etc.
- These assumptions often made without thought.
 - “Nothing is wrong with making assumptions . . . It is the scientific quality of those assumptions, not their existence, that is critical.”
 - Hopefully, assumptions are reported so they can be scrutinized and perhaps improved.
- Analysis complicated (but not impossible) if not true.

Inference

- Neyman (1923)
 - $\bar{y}_1 - \bar{y}_0$ (the difference of sample averages) is an “unbiased” estimator of $\bar{Y}(1) - \bar{Y}(0)$ (average causal effect) . . .
 - if assignment is randomized
- Fisher's (1926) null hypothesis

$$Y_i(1) = Y_i(0) \text{ for all } i$$

- leads to randomization inference,
- since it means we know all possible values of $\bar{y}_1 - \bar{y}_0$.

Why are randomized experiments so special?

- The assignment mechanism is the only thing that is random.
- “No model of science may be needed beyond SUTVA.”

$$p\text{-value (Fisher)} = Pr(\bar{y}_1 - \bar{y}_0 \geq \bar{y}_{1,obs} - \bar{y}_{0,obs} | X, Y(1) \equiv Y(0), Y(0) \equiv Y(1))$$

- $\bar{y}_1 - \bar{y}_0$ random depending on assignment
- $\bar{y}_{1,obs} - \bar{y}_{0,obs}$ actual observed value of $\bar{y}_1 - \bar{y}_0$

Contrast: observational studies

- Observed Y related to assignment and X .
- “Mixes up science – the potential outcomes, and what we do to try to learn about science – and assignment mechanism.”

What’s at Stake (Louis Menand again)

“The new sciences of human nature.” Well, why not? The old sciences of human nature didn’t have such a fabulous track record. They gave us segregated drinking fountains, “invented spelling,” and the glass ceiling – all consequences of scientific theories about the way human beings really are. Possibly, there is a lesson here, which is that the sciences tend to validate the practices and preferences of whatever regime happens to be sponsoring them. In totalitarian regimes, dissidence is treated as a mental illness. In apartheid regimes, interracial contact is treated as unnatural. In free-market regimes, self-interest is treated as hard-wired. Maybe this is unfair to the *new* sciences, though. It could be that the problem with the old sciences was simply that they weren’t scientific enough – that they were mostly wishful thinking projected onto dubious data about skull size and the effects of estrogen on the ability to balance a checkbook. Today’s scientists might have the capacity to get right down there among the chromosomes and neurotransmitters, and to send reports, undistorted by fear, favor, or the prospect of funding, about what’s going on. Maybe the new sciences are really scientific.