

Inferring Causality

Psychology (Statistics) 484

Statistics, Ethics, and the Social and Behavioral Sciences

June 18, 2013

Beginning Quotations

Inferring
Causality

Psychology
(Statistics)
484

If once a man indulges himself in murder, very soon he comes to think little of robbing; and from robbing he next comes to drinking and Sabbath-breaking, and from that to incivility and procrastination.

– Thomas De Quincey (1788–1859)

The First Law of Geography: Everything is related to everything else, but near things are more related than distant things.

– Waldo Tobler

Causal interpretation of the results of regression analysis of observational data is a risky business. The responsibility rests entirely on the shoulders of the researcher, because the shoulders of the statistical technique cannot carry such strong inferences.

Jan de Leeuw

A medical maxim—when you hear hoofbeats, think of horses before zebras.

– Harley Smith

Week 10: Inferring Causality

Inferring
Causality

Psychology
(Statistics)
484

— inferring causality; Bradford–Hill criteria; historic medical conceptions of disease causality; medical error as the causative factor

Required Reading:

SGEP (307–331) —

Causuistry

The Bradford–Hill Criteria for Determining a Causal Connection

Some Historical Health and Medical Conceptions of Disease

Causality

Medical Error as a (the) Causative Factor

Popular Articles —

The Truth Wears Off, Jonah Lehrer (*New Yorker*, December 13, 2010)

Lies, Damned Lies, and Medical Science, David H. Freedman (*The Atlantic*, November 2010)

Film:

Smoking and Cigarette Film Collection (102 minutes)

Introduction

Inferring
Causality

Psychology
(Statistics)
484

The aim of any well-designed experimental study is to make a causal claim, such as “the difference observed between two groups is *caused* by the different treatments administered.”

To make such a claim we need to know the counterfactual:

what would have happened if this group had not received the treatment?

This counterfactual is answered most credibly when subjects are assigned to the treatment and control groups at random.

In this instance, there is no reason to believe that the group receiving the treatment condition would have reacted any differently (than the control condition) had it received the control condition.

If there is no differential experimental mortality to obscure this initial randomness, one can even justify the analyses used by how the groups were formed (for example, by randomization tests, or their approximations defined by the usual analysis methods based on normal theory assumptions).

As noted by R. A. Fisher, “the actual and physical conduct of an experiment must govern the statistical procedure of its interpretation.”

Quasi-experimentation

Inferring
Causality

Psychology
(Statistics)
484

When the gold standard of inferring causality is not met, however, we are in the realm of quasi-experimentation, where causality must be approached differently.

When data from observational/quasi-experimental studies are to be interpreted, the question of causality often gets rephrased as one of whether the experiment is “internally valid.”

Internal validity addresses the question of what manifestly appears to distinguish the groups of respondents, and whether this could also be the reason for the observed differences.

Such a reinterpretation has been a part of the graduate education of psychologists for over forty years, primarily through Campbell and Stanley (1966) (*Experimental and Quasi-Experimental Designs for Research*), and later through all of its many successors (e.g., Shadish, Cook, & Campbell, 2002).

Basically, one tries to argue for causality by eliminating various threats to the internal validity of the quasi-experiment.

We are reminded of the quotation from Sir Arthur Conan Doyle (and Sherlock Holmes) (*The Sign of Four*): “Eliminate all other factors, and the one which remains must be the truth.”

We list eight threats to internal validity reviewed by Winch and Campbell (1969):

Threats to Internal Validity

Inferring
Causality

Psychology
(Statistics)
484

- (1) *History*: Events, other than the experimental treatment occurring between pre-test and post-test and thus providing alternate explanations of effects.
- (2) *Maturation*: Processes within the respondents or observed social units producing changes as a function of the passage of time per se, such as growth, fatigue, secular trends, etc.
- (3) *Testing*: The effect of taking a test on the scores of a second testing. The effect of publication of a social indicator on subsequent readings of that indicator.
- (4) *Instrumentation*: Changes in the calibration of a measuring instrument or changes in the observers or scores used producing changes in the obtained measurements.

- (5) *Statistical regression artifacts*: Pseudo-shifts occurring when persons or treatment units have been selected upon the basis of their extreme scores.
- (6) *Selection*: Biases resulting from differential recruitment of comparison groups, producing different mean levels on the measure of effects.
- (7) *Experimental mortality*: The differential loss of respondents from comparison groups.
- (8) *Selection-maturation interaction*: Selection biases resulting in differential rates of “maturation” or autonomous change.

In addition to eliminating threats to internal validity in arguing for plausible causal explanations when true experiments are not possible (for example, as in establishing the links between smoking and lung cancer in humans), it is important to verify the tell-tale signs of a potentially causal connection:

the association needs to be consistent;

higher “doses” should be connected to stronger responses;

alleged causes should be plausible and temporally appropriate by preceding the effect in question.

Lord's Paradox

Inferring
Causality

Psychology
(Statistics)
484

A potential statistical solution to the equating of intact groups is sometimes offered in the form of analysis of covariance, where adjustments in treatment effects are made based on existing group differences for a collection of covariates.

Unfortunately, and contrary to the dreams of a quasi-experimenter, this is just not possible.

We quote from Fred Lord's (1967) classic article, "A Paradox in the Interpretation of Group Comparisons":

[T]here simply is no logical or statistical procedure that can be counted on to make proper allowances for uncontrolled preexisting differences between groups. The researcher wants to know how the groups would have compared if there had been no preexisting uncontrolled differences. The usual research study of this type is attempting to answer a question that simply cannot be answered in any rigorous way on the basis of available data.

In contexts where group differences on other variables may influence the response variable of interest, it is best to block on these variables and randomly assign subjects within blocks to the treatments.

Given the horrendous difficulties encountered in treatment-covariate confounding in ANCOVA, this type of statistical control by blocking is just not legitimately done after the fact.

The analysis of covariance cannot be relied on to adjust properly for or to control the differences between naturally occurring groups.

Faulty Causal Reasoning

Inferring
Causality

Psychology
(Statistics)
484

Recalling from an earlier chapter the traditional adage that “correlation does not imply causation,” we might add a corollary:

merely identifying a plausible causal mechanism doesn't change the import of this caveat.

The plausible mechanism may result from a third confounding or “lurking” variable just as the original correlation is spurious and due to the influence of a third variable.

A salient illustration of this faulty reasoning was the initial observation that women with herpes are more likely than other women to develop cervical cancer.

Because of general beliefs about biomedical activity, some investigators concluded that the relationship was causal and cancer must result from the herpes.

Later research, however, showed that the primary cause of cervical cancer was human papilloma virus (HPV), and herpes was just a marker for sexual activity.

Women with multiple sexual partners were more likely to be exposed not just to herpes but also to HPV.

The apparent association between herpes and cervical cancer was the result of other variables.

Causuistry

The legitimate word *casuistry* has several meanings.

It can refer to specious reasoning intended to mislead, or to the evaluation of right and wrong by analyzing specific cases that illustrate general ethical rules.

In his 2010 book, *Proofiness*, Charles Seife adds another “u” to coin a new word, “causuistry,” the ascribing of cause to what are just accidental associations.

As Seife points out, sometimes things just happen for no apparent reason.

In medicine, for example, the term *idiopathic* is commonly applied when some condition appears to arise spontaneously or for an unknown reason.

Randumbness

Inferring
Causality

Psychology
(Statistics)
484

People generally have a difficult time accepting that an event or situation is of unknown or idiopathic origin (think of Job: “Why me, Lord?”).

Seife coins another word that helps explain this human failing, “randumbness,” the insistence on order in the presence of what is more often just chaos, or in creating pattern when this is none to see.

In a *ScienceNews* article by Bruce Bower (February 12, 2011), “In the Zone,” an argument is reviewed (though it is completely unprovable) that evolutionary pressures may be partially responsible for our minds seeing streaks, such as the “hot hand” in basketball.

One particularly troubling result of the need to have causal explanations is when we observe some grouping of unusual medical cases, such as that for cancer, birth defects, or similar bad things.

Even though such clusters most often happen just randomly, we still search for cause, and perhaps, for someone or something to blame.

A good and insightful review of this kind of specious “culling for clusters” phenomenon, and the subsequent “search for causes” is given in a *New Yorker* article by Atul Gawande, “The Cancer-Cluster Myth” (February 8, 1999).

Specious Causation

Inferring
Causality

Psychology
(Statistics)
484

Assertions of specious causation never seem to disappear for events that are merely identified (i.e., “cherry-picked”) as a result of their salient extremeness.

Because of the implicit search carried out over a large database, what we try to explain is likely a result of mere chance and the operation of various random processes.

Currently, we have the earnest news stories about master teachers who supply “added value” to their students, and are identified by an amazing change seen in test scores over time.

Such a teacher's behavior is then scrutinized, hoping to add to the collection of "best practices" that could be passed on to others with less or negative "value added."

Most (or all) of this is just chasing moonbeams.

It is best to keep in mind that subjective belief is insufficient as evidence of causation, as is inferring causality not from a controlled study but from a few selected anecdotal observations.

Anti-Vaccine Movement

Inferring
Causality

Psychology
(Statistics)
484

One insidious public health instance of such search for causation and a good exemplar for the fallacy of *post hoc ergo propter hoc* (after this, therefore because of this) is the anti-vaccine movement.

This campaign is fueled by a false belief that a causal link exists between the administration of childhood vaccines and illnesses such as autism.

As a direct result of not immunizing children for the various childhood illnesses that were once fairly common (for example, polio, MMR [measles, mumps, and rubella], DPT [diphtheria, pertussis (whooping cough), and tetanus]), the protective “herd immunity” levels are dropping, possibly allowing a resurgence for these diseases.

Two books by Paul Offit lay out the issues well for this false ascription of causality between the administering of vaccines and the subsequent onset of an illness:

Autism's False Prophets: Bad Science, Risky Medicine, and the Search for a Cure (2008)

Deadly Choices: How the Anti-Vaccine Movement Threatens Us All (2011).

Several causation fallacies are important to avoid.

One has to do with the effect of small numbers and when random variability produces wide shifts that are then interpreted causally.

If states, for example, award special recognition to schools showing marked improvement in test scores, they should not be overly surprised when only small schools are the ones garnering attention.

Bigger changes and smaller sample sizes are intimately tied.

A second trap is to be seduced by the lure of a coherent narrative.

Thus, we might argue that heroin addiction is primed by the use of a “gateway” drug such as tobacco or marijuana; here, the word “gateway” lends credence to a causal assertion even though it isn’t true.

Just the naming or postulating of some mechanism doesn’t provide proof that a causal link is present.

A well-publicized fiasco to keep in mind that is discussed elsewhere is on the “heart protective” value of hormone replacement therapy, or the now debunked presence of a “timing hypothesis” as to when hormone replacement therapy should begin for it to be protective.

The Bradford–Hill Criteria for Determining a Causal Connection

Inferring
Causality

Psychology
(Statistics)
484

Sir Austin Bradford Hill (Fellow of the Royal Statistical Society; 1897–1991) is considered by many to have been the world's foremost medical statistician of the twentieth century.

He pioneered the randomized clinical trial, working with streptomycin in treating tuberculosis in the 1940s.

Hill's most famous contributions, however, were with Richard Doll.

Hill and Doll (pun intended) established the link between lung cancer and smoking in the 1950s.

The nine Bradford–Hill criteria for establishing causality were published by Hill in 1965 (*Proceedings of the Royal Society of Medicine*, 58, 295–300).

They are not meant as a checklist, and Hill himself noted: “None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required *sine qua non*”.

In any case, here they are:

Strength: A weak association does not imply that there is no causal effect; however, the stronger the association, the more likely it is causal.

Consistency: Consistent findings observed by different people in different places with different samples strengthens the likelihood of a causal effect.

Specificity: Causation is likely if a very specific population at a specific site has a disease with no other likely explanation. The more specific an association between a factor and an effect, the larger the probability of a causal relationship.

Temporality: The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, the effect must occur after that delay).

Biological gradient: Greater exposure should generally lead to greater incidence of the effect (that is, a dose-response relationship). In some cases, the mere presence of the factor (above some threshold, say) can trigger the effect; in others, an inverse proportion might even be observed, where greater exposure leads to lower incidence.

Plausibility: A plausible mechanism between cause and effect is helpful (but knowledge of the mechanism may be limited by current knowledge).

Coherence: Coherence between epidemiological and laboratory findings increases the likelihood of an effect. But as Hill noted, “lack of such [laboratory] evidence cannot nullify the epidemiological effect on associations”.

Experiment: Occasionally, it is possible to appeal to experimental evidence, such as when an agent is removed, a reduction in a particular disease is observed.

Analogy: The effect of similar factors may be considered.

We suggest three additional criteria:

(a) Have alternative explanations been considered thoroughly? possibilities include confounding with the presence of other variables; bias operating with respect to the subjects being studied; or by how exposure or disease status is assessed.

(b) Are there effects of ceasing or altering exposure?

(c) Are the findings consistent with other relevant knowledge?
For example, this may occur when a lessened (national) use of some agent (for example, smoking) appears associated with a reduced occurrence of some disease (for example, the prevalence of lung cancer).

Some Historical Health and Medical Conceptions of Disease Causality

Inferring
Causality

Psychology
(Statistics)
484

A later week will be concerned with the collection of data and experimental design and introduces the term *differential etiology*.

Simply put, this involves the investigation and reasoning that lies behind the determination of external causation, usually for a disease.

It typically proceeds by a process of elimination, and is directed toward identifying disease causes, but not to determining treatment.

A treatment emphasis falls under the notion of *differential diagnosis*.

In the early to late 1800s, the dominant idea used to explain infectious diseases was miasma theory.

Although diseases such as smallpox, syphilis, and measles were believed to be contagious, it was unclear what to think about the epidemic diseases such as cholera, typhoid, and typhus.

Miasma theory held that certain diseases, such as cholera, were caused by a noxious form of “bad air” (a miasma), filled with particles from decomposed matter (called miasmata). It could be identified by its foul smell.

The miasma theory of disease became a very appealing causal explanation in the middle 1800s for English public health reformers who could now focus on environmental problems as opposed to those tied only to infection and personal health.

We quote part of the Wikipedia article on “Miasma theory” in the required reading that mentions the work of two influential 19th century English statisticians, William Farr and Florence Nightingale.

There have been several historical instances where causes have been asserted for disease occurrence that later were proved wrong (usually, this happened after the germ theory of disease was developed in the late 1800s).

One particularly interesting historical tale involves Ignaz Semmelweis, who proposed a causal theory of puerperal (childbed) fever, which was rampant in the maternity wards of Europe in the middle 1800s.

Semmelweis believed that childbed fever was the result of doctors dissecting cadavers and then delivering babies without first washing their hands.

Arguing that cadaverous particles, similar to those in miasma theory, were being absorbed by the patients, Semmelweis institutionalized hand washing with a lime and chlorine solution (a disinfectant that removed the cadaverous odor), before reentering the maternity wards.

The results were dramatic in terms of the tenfold reduction in mortality from childbed fever.

The careful collection of time series mortality statistics before and after the introduction of the use of hand disinfectant lead Semmelweis to conclude that indeed it was the cadaverous material that was the cause of the childbed fever.

Although the effect of mortality reduction was present with the use of disinfectants, Semmelweis got the specific causal mechanism wrong;

it wasn't necessarily cadavers, but most often *Streptococcus pyogenes* present in unclean areas generally.

We should remember that this was before the germ theory of disease was established in the later 1800s, and therefore, we shouldn't blame Semmelweis for taking advantage of the "law of unintended consequences," where eliminating the cadaver odor also killed the *Streptococcus pyogenes*.

One could argue that Semmelweis was seduced by his strong belief in the biological causal mechanism of cadaver particles being resorbed by women giving birth.

And to some extent, this insistence on a single cause created difficulty in the adoption of the Semmelweis disinfectant theory.

Seduction by biological mechanism occurs to this day.

One example will be discussed in detail later involving hormone replacement therapy for women.

Just because a plausible biochemical mechanism could be advanced for why hormone therapy might be heart-protective doesn't necessarily mean that it has to be true.

In fact, hormone therapy is generally the opposite of heart-protective.

The reason the supposed benefits were observed in the women studied was likely due to the "healthy-user bias".

Women who took the hormone replacements regularly were generally different (for example, more health-conscious and diligent) than those who didn't, and this masked the severe downside of hormone therapy.

Medical Error as a (the) Causative Factor

Inferring
Causality

Psychology
(Statistics)

484

The term *iatrogenesis* refers to the inadvertent adverse effects or complications that result from medical treatment.

These errors can be widely defined: surgical complications, drug interactions, negligence, and so on.

As just discussed, the major iatrogenic catastrophe of the 19th century was puerperal fever, resulting from pathogen transfer from autopsy room to maternity patients.

More recently, major iatrogenic errors have been in the use of radiation, the antibiotic resistance in bacteria, and the various hospital anomalies against which the Institute for Healthcare Improvement have been campaigning for some time (discussed further in the required readings).