Defining the inquiry

Defining the inquiry

An inquiry is a summary of a theoretical model. Suppose your theory is that X affects Y. Using the framework provided in Pearl and Mackenzie (2018), one inquiry might be descriptive, or associational: what is the average level of Y when $X = 1$? A second might be about the effects of interventions: what is the average treatment effect of X on Y? A third is about counterfactuals: for what share of units would Y have been different if X were different? If a theory involves more variables, many more questions open up, for instance regarding how the effect of one variable passes through, or is modified by, another.

Your inquiry is your research question. Simple or complex, causal or descriptive, your inquiry can be thought of a summary of a data generating processes. Like models, inquiries—the questions you ask—are themselves theoretical objects. It is easy to confuse inquiries and the output of answer strategies. If our theory posits the existence of an Average Treatment Effect, we might use an answer strategy like difference-in-means to estimate it, but the estimate is fundamentally distinct from the inquiry. Estimates are empirical, inquiries are theoretical.

In general, an inquiry is a summary function I that operates on an instance of a model $m \in M$. When we summarize the model with the inquiry, we obtain an "answer under the model." We formalized this as $I(m) = a^m$. You can think of the difference between I and a^m as the difference between a question and its answer. I is the question we ask about the model and a^m is the answer. Alternatively, you can think of I as the "estimand" (that which is to be estimated) and a^m as the value of the estimand.

In this book when we talk about inquiries, we will usually be referring to single-number summaries of models. Some common estimands are descriptive, such as the means, conditional means, correlations, partial correlations, quantiles, and truth statements about variable in the model. Others are causal, such as the average difference in one variable when a second varible is set to two different values. You can think of a single-number inquiry as the atom of research question.

While most inquiries are "atomic" in this way, some inquiries are more complex than a single-number summary. For example, the best linear predictor of Y given X is a two-number summary: it is the pair of numbers (the slope and intercept) that minimizes the total squared distance between the line and each value of Y. Note that this is not a *causal* estimand but it is still well defined. No need to stop at two-number summaries though. We could imagine the best quadratic predictor of Y given X (three-number summary), and so on. See Figure @ref(fig:ch3polynomials). We could have an inquiry that is the full conditional expectation function of Y given X, no matter how wiggly, nonlinear, and nuanced the shape of that function – it could in principle be a 1,000 number summary of the model, or much more.

And of course it need not be a number at all: the answer to your question might be "blue" or "the normal distribution." Your inquiry could be "should I reject M as a useful model of the world?" A researcher might articulate a handful of important questions about the model that all have to come out a certain way or the model itself should be rejected. These complex inquires are made up of a series of atomic inquiries – we're interested in the sub-inquiries only insofar as they help us understand the real inquiry – is this model of the world a good one or not.

Kinds of Inquiries

Atomic inquiries I about a model M fall into two basic categories, descriptive and causal. Descriptive inquiries are about how the world was, is, and will be whereas causal inquiries are about how the world would have been, would be, or would be in the future if some variables were set to different levels. You might think of
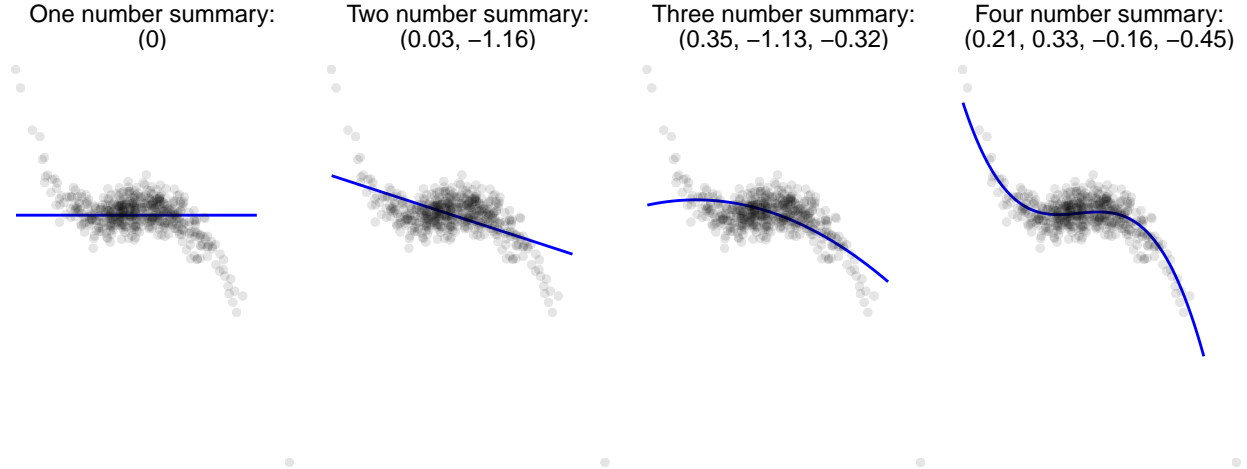
Figure 1: Inquiries based on different numbers of parameters

the distinction in terms of a DAG. Descripitive inquiries are about the nodes of a dag whereas causal inquires are about the edges.

Descriptive and causal inference have in common the difficulty of inferring unseen things from observed data. The fundamental problem of causal inference is well known. For the same reason that a unit cannot simultaneously be both treated and untreated, we can only every observe at most one potential outcome for any particular unit. The fundamental problem of descriptive inference is similar. The concepts we want to measure are latent constructs. Perfect measurement is impossible, so our measurements of the latent constructs always include some measurement error. Just like we can't know for sure how, counterfactually, a unit would have responded if the treatment had been set to a different level, we likewise can't know for sure whether our measurements accurately reflect the latent construct.

More specifically, a descriptive *inference* is a conclusion about features of a latent variable Y^* on the basis of observations of a measured variable Y. The feature we seek to describe—our inquiry—is some summary of Y^* like its mean or perhaps its covariance with a second latent variable X^*. When we do descripitive research, we draw inferences about features of the nodes of the latent causal model M. Our measurement can be imperfect for two reasons: that we do not get to observe the quantity of interest direcly for the units we study, and because the units we study are only a subset of the units of interest. It is these challenges that gives rise to the focus on descriptive *inference* rather than measurement alone.

A causal *inference* is a conclusion about the edge between two latent variables X^* and Y^*. Even if we do an exceptional job measuring X^* and Y^* with X and Y, we will still have trouble learning about edges because causal effects are quite literally unobservable. We have to *infer* causality on the basis of a strong research design because we can't just see it.

Descriptive inquiries

Descriptive inquiries are usually about latent variables since we mostly care about the true values of the variables in the models. Measured variables are often distinct from latent variables giving rise to a problem of descriptive inference. For the most part, we define our inquiries in terms of the true latent variables rather than in terms of their measured counterparts.

Table ZZZ enumerates some common descriptive estimands.

Inquiry	Description
$E_{i \in N}(Y)$	The average value of the variable Y

Inquiry	Description
$E[Y\|X=1]$	A conditional expectation of Y given X = 1.
$V[Y]$	The variance of Y
$\mathrm{Cov}(X,Y)$	The covariance of X and Y
$\mathrm{BLP}(Y\|X)$	The best linear predictor of Y given X
$\mathrm{CEF}(Y\|X)$	Conditional expectation function of Y given X
Truth status of X	Is X True or False

Causal inquiries

Causal inquiries involve a comparison of at least two possible worlds. For example, an inquiry might be the causal effect of X on Y for a single unit. In order to infer that causal effect, we would need to know the value of Y in two worlds: one world in which X is set to 1 and one in which X is set to 0.

Table ZZZ enumerates some common causal estimands.

- Estimand scope:What is the set of units which you want to learn the answer about? Know what ATE averages over

Inquiry	Description
$E[Y_i(1) - Y_i(0)]$	Average treatment effect (ATE)
$E[Y_i(1) - Y_i(0)\|X=1]$	Conditional average treatment effect (CATE)
$E[Y_i(1) - Y_i(0)\|d_i(1) \geq d_i(0)]$	Complier average causal effect (CACE) or local average treatment effect (LATE)
$E[Y_i(1) - Y_i(0)]$	PATE
$E[Y_i(1) - Y_i(0)\|S=1]$	SATE

Generations of students have been told to excise words that connote causality from their empirical writing. "Affects" becomes "is associated with" and "impacts" becomes "moves with." Being careful about causal language is of course very important (it's really true that correlation does not imply causation!). But this change in language is not usually accompanied by a change in inquiry. Many times we are faced with drawing causal inferences from than ideal data – but the deficiencies of the data strategy should not lead us too far away from our inferential targets. If the inquiry is a causal inquiry, then the move from "causes" to "is correlated with" might be a good description of the actual data analysis, but it doesn't move us closer to providing an answer to the inquiry.

Relationship between causal and descriptive inquiries

You can think of causal inference as repeated descriptive inference: we have to describe Y in multiple possible worlds.

Causal inquiries like the average effect of A on B in a causal model like $A->B<-U$ are helped enormously by good descriptive inference about the nodes A and B, but the focus is on the edge between them. You typically can't learn about the edge by doing desscripitve inference on the nodes only. If we measure A and B and find that they covary, we can't be sure that $A->B$ because it could be that $A<-U->B$. This problem goes by the familiar phrase that "correlation doesn't imply causation," which is true and is a problem that can't be wished away. But "correlation doesn't imply causation" also kind of misses the point. The point is that you can't see causality because it involves counterfactuals, which are imaginary and unseen. You have to infer causality on the basis of design.

Data-dependent inquiries

The inquiries we have introduced thus far depend on variables in the model, but not on features of the data and answer strategies. Many common inquiries do depend on realizations of the research design.

The first type depend on realizations of the data d: inquiries about units within a sample depend on which units enter the sample; inquiries about treated units depend on which are treated. The sample average treatment effect is a common inquiry used by experimental researchers who wish to not worry about how their effects generalize to a population but only about identifying the causal effect within the units in front of them. There is a true sample average treatment effect a^w for every possible sample that we could draw. But which of those fixed values is selected as our inquiry if we define I as the sample average effect depends on which sample is actually selected by our sampling procedure. The same is true when we condition inquiries on the set of treated units.

The second design-dependent inquiry class depends not on d but $A(d) = a^d$, the answer given the data. In an a^d-dependent inquiry, we only know the inquiry after seeing the results of our study. When we collect data on ten variables and calculate the correlation between all ten on a separate outcome variable, and then report on the magnitude of the correlation of the seventh variable - the only one that was statistically distinguishable from zero. If our inquiry is "which of these seven variables is significant" then there will be no problem, as long as we include a multiple comparisons correction to adjust for the probability of finding one significant under null effects due to random chance. If we started with the question of what is the correlation of variable seven and the outcome, we would also have no problem. However, if with our procedure we think of our inquiry as what is the correlation between variable seven and the outcome, then we have a problem: we are not accounting for the multi-step procedure in our answer strategy and we will not be able to provide good answers to that inquiry. In short, we typically should not be guided by the realization of our answer in *selecting* an inquiry, but there is no problem with procedure-based answer strategies – we just must be honest about our original inquiry. This problem comes up in descriptive inference in looking at multiple correlations as in this example, but also in many other places such as searching for heterogeneity in treatment effects in experiments and nested research designs that iterative between levels of data (Lieberman 2005).

Undefined and unanswerable inquiries

Declaring your design in terms of $MIDA$ may lead you to two awkward conclusions: your inquiry I returns $I(m) = a^m = \text{NA}$, i.e., the answer to your inquiry is undefined; or your answer may be (currently) unanswerable, that is, there is no feasible D and A that yield an answer to I. In both cases, one option is to change your inquiry. But often we selected I because of its importance, so we may want to try to find an answer. In the case of undefined inquiries, we have no option but to select a new one. In the case of unanswerable inquiries, we can work to identify a novel D and A that change the set of feasible designs and provide an answer. In some cases, however, the fact that the inquiry is unanswerable may be due to unchangeable limitations of research such as the fundamental problems of causal inference or descriptive inference.

How should you select inquiries?

It's hard to know where to start when picking a research question. We want to pick one that is interesting in its own right or one that would a facilitate real-world decision. We want to pick research questions that we can learn the answer to someday, with a lot of effort. Infeasible research questions should be abandoned as soon as possible, but of course that's hard to do. The trouble is that it's hard to know what research questions are feasible before you start looking into it, and it's really hard to quit research projects once you learn they are infeasible because of the sunk cost fallacy. Among feasible research questions, we want to select ones that we are likely to obtain the most informative answers, in terms of moving our priors the most. This last criteria will often help us select among related, feasible inquiries that are about the same DAG but for which we know we can learn more and less through research.

Sometimes, people give advice to students to follow a "theory-first" route to picking a research question. Read the literature, find an unsolved puzzle, then start choosing among the methodological approaches that might answer the problem. Others eschew the theory-first approach: "How on earth are you going to happen to land upon an unsolved – and yet somehow solvable – puzzle just by reading!?" These advice-givers emphasize a method-first route. Master the technical data-gathering and analysis procedures first, then set off to find opportunities to apply them. The theory-first people then say: "how would you know an interesting

theoretical question if it smacked you in the face!?"

Iteration between the two is typically necessary. In order to select research questions, empiricists have to be concerned about the entire research design. We have to develop empirical strategies to provide answers to our inquiries. We have to learn a lot about how to select data and answer strategies in ways that map on to inquiries about models. So empiricists have to learn both about about models and inquiries (theory) as well as about data strategies and answer strategies (empirics).

The first criteria is subjective importance of a question. The object of the importance may be a scientist, considering the value to building a theoretical understanding of the world; to a policymaker, deciding how to collect and allocate resources in a government; a private firm, who is making decisions about how to invest their own resources to maximize profit; or another individual or organization. The scientific enterprise is designed around the idea that importance is in the eye of the scientist and is not some objective quantity. This is for two reasons. First, the scientific or practical importance of a discovery may not be understood until decades later, when other pieces of the causal model are put together or the world faces new problems. Moreover, "importance" differs for different segments of society, and scientists must be able to study questions not judged important by groups in power in order to discover new ways to solve problems faced by left-out groups.

Among important questions, researchers can select the I they think they are likely to be able to learn the most about. We start with a prior distribution over a^w, the true answer about the inquiry, from past research, and a good research design will substantially update that prior distribution, either by moving the mean of the distribution or reducing our uncertainty about it.

The final criteria is that among important, probative research designs there must be a feasible M, D, and A that could answer I. That means picking a good question I does not just involve theory. You should study M to understand which Is are worth knowing. But you should also study D and A in order to learn how to demonstrate I. A central goal of research methodologists is to expand the feasible set of D and A that can provide informative answers to important questions.

Clearly research designs will vary how important, feasible, and probative they are, and so researchers so these criteria do not provide an immediate answer to how to select a design. Instead, the researcher must choose a weighting of the three, and norms in a research community may guide those weightings. In some disciplines, providing a minimally probative answer to a highly important question may be preferable to a highly probative answer to a less important question.
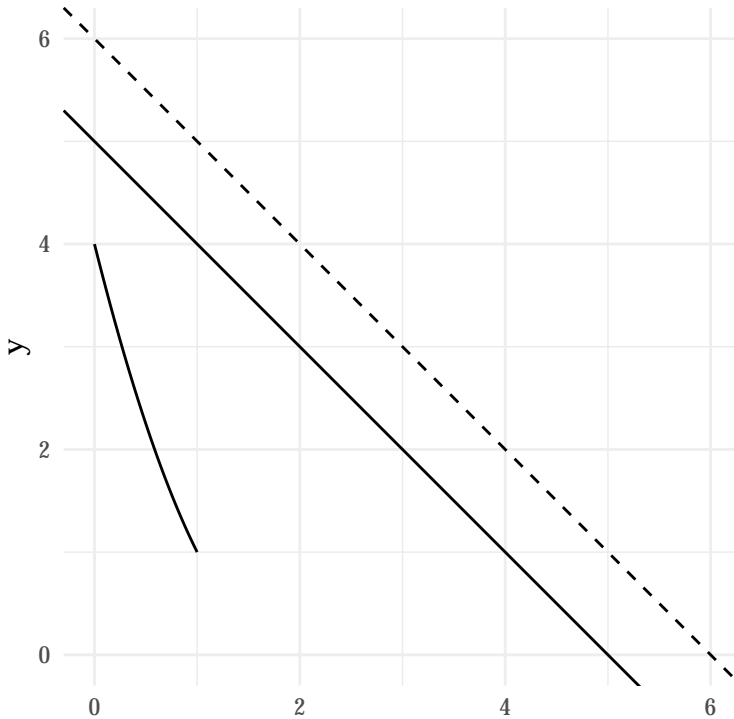
Here's our best advice for how to get started picking a research question: Write down the M and I of any causal model that you think is important to get started thinking about selecting strong Ds and As. The goal is to learn how to map $I(M)$ to $A(D)$. Then return to theory to find new important inquiries, then write down a new model inquiry data strategy and answer strategy. This process is how we make progress on M, that is, we bring M closer to W, thereby making M truer. In each iteration, consider how informative the answers you can provide: how much will they update what we know about the world?

People who are looking in the forest for mushrooms often don't see a mushroom for a long period of time. After a while, they acclimate. They get their "eyes on," and successful finds seem to be around every bend. In this analogy, the theory building process is going for a walk in the forest and methods training is learning to spot mushrooms – you need get your eyes on answerable research questions worth answering.

Example: Björkman and Svensson (2009)

Björkman and Svensson (2009) reports the results of a cluster-randomized trial of the effects of community-based monitoring of health clinics in Uganda to improve childrens health. The main inquiry was the average treatment effect on the child mortaility rate. The study showed that the program was a success: in the control group, the child mortality rate was 144 per 1000 live births, compared with 97 in the treatment group, for a 33% reduction in child mortality.

The study also considered a second inquiry: the average treatment effect on the weight-for-age of children under 18 months. This inquiry is harder to think about, precisely because we know that the community-based monitoring saved the lives of many children. To see problem, consider Table XXX, which shows four types of infants distinguished on the basis of their potential outcomes. Type A (for "Adverse") is alive if in control, but dies if in treatment. Type B ("Beneficial") is just the reverse: the child dies if untreated, but survives if treated. Type C ("Chronic") would die under either condition and Type D ("Destined") would live under either condition. For the first three types, the child dies under one condition, the other, or both. Since weight-for-age only exists if the child survives, the treatment effect on weight for A, B, and C types is undefined.

The main trouble here is that the average treatment effect averages over all four types so the ATE itself is

also undefined. Since this inquiry is undefined, we'll need to select a different one.

We might want switch our inquiry to be about the average effect among D types only. However this inquiry has a different problem – it's unanswerable. Even though this inquiry is not undefined, it's unanswerable because we won't be able to learn who the D types are. In the treatment group, we can't tell the Bs from the Ds and in the control group, we can't tell the As from the Ds. The reason we can't tell these types apart using realized data is the fundamental problem of causal inference.

Type	Alive($Z = 0$)	Alive($Z = 1$)	Weight($Z = 0$)	Weight($Z = 1$)	Estimand
A	1	0	exists	NA	undefined
B	0	1	NA	exists	undefined
C	0	0	NA	NA	undefined
D	1	1	exists	exists	$E[\text{Weight}(Z = 1) - \text{Weight}(Z = 0)]$

What can be done in this situation? We might try to find a covariate X that is correlated with being a D type and condition our analysis of the effect on weight on that covariate. Note that among this subgroup, the effect on mortality should be equal to zero. Finding such a covariate would be hard, though perhaps not impossible.

Alternateive, we might define our inquiry to be the average treatment on the average weight of living children within a cluster, but this inquiry requires careful interpretation. If the treatment saves the lives of children whose health is *marginal* in particular, then the effect on average weight could easily be negative, even though the treatment improves health. We specifically cannot interpret the answer to this inquiry to refer to the effect of treatment on health. That said, it might nevertheless be a useful number to know if we are in the position of admistering healthcare resourses – we'd want to send additional help to those treatment areas where the lives of unhealthy children (who otherwise would have perished) are being saved.

Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment of a Community-Based Monitoring Project in Uganda." *Quarterly Journal of Economics* 124 (2): 735–69.

Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (3): 435–52.

Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect.* Basic Books.