Causal Inference in Statistics: A Primer Solution Manual

Text Authors:
Judea Pearl, Madelyn Glymour, and Nicholas Jewell
Solution Authors:
Judea Pearl, Ang Li, Andrew Forney, and Johannes Textor

About This Manual

This document provides solutions, explanations, and intuition for the study questions posed in *Causality in Statistics: A Primer.* Students are encouraged to attempt each study question by hand before consulting the answers herein.

Online Access

As the authors make updates to the text and solution manual, changes and errata will be posted at the following links:

Textbook Information & Update site: http://bayes.cs.ucla.edu/PRIMER/
Solution Manual Information & Update site: http://bayes.cs.ucla.edu/PRIMER/Manual

Interactive Tutorial using DAGitty

The authors have collaborated with Johannes Textor, the maker of DAGitty (a browser-based environment for creating, editing, and analyzing causal models), to provide interactive tutorials for classroom use and self-study. We provide solutions to some exercises in the *R* environment for statistical computing, based on the DAGitty *R* package. Each question with an accompanying DAGitty explanation has been tagged accordingly, and students may find a complete list of these examples at the following link:

[DAGitty] Textbook Companion site: http://dagitty.net/primer/

Study Questions and Solutions for Chapter 1

Study question 1.2.1.
What is wrong with the following claims?

(a) *"Data show that income and marriage have a high positive correlation. Therefore, your earnings will increase if you get married."*

(b) *"Data show that as the number of fires increase, so does the number of fire fighters. Therefore, to cut down on fires, you should reduce the number of fire fighters."*

(c) *"Data show that people who hurry tend to be late to their meetings. Don't hurry, or you'll be late."*

Solution to study question 1.2.1

The three claims are obviously wrong, and in subsequent sections of this book we will acquire the tools to formally prove them wrong. At this stage, however, we will merely explain the observed correlations using alternative models which do not support the claims cited.

For each problem, we will explain the observed correlation using new variables in the answers below.

Part (a)

Consider an alternative explanation with a third variable, charm, which has a causal influence on both income and marriage (charming individuals have a higher propensity to marry and be promoted in their jobs), but where marriage has no causal influence on income. This explanation supports the observed data (that marriage and income are highly correlated) but does not allow us to conclude that marrying will increase one's income.

Part (b)

Consider that the number of fire fighters employed in a district is a direct response to the frequency of fires in that area. In a natural scenario, a higher frequency of fires causes additional fire fighters to be hired, and hiring fire fighters decreases the number of fires that would break out had they not been hired. Hence, hiring fewer fire fighters will actually increase the frequency of fires.

Part (c)

Let us consider the reason that an individual might hurry to an appointment: they believe that a slow pace will not allow them to arrive to the appointment on time because they woke up late. So, waking late is a common cause of hurrying and arriving late to the meeting. This will cause a high correlation between hurrying and arriving late, even though for a fixed waking time, hurrying would actually decrease one's likelihood of arriving late.

Study question 1.2.2.
A baseball batter Tim has a better batting average than his teammate Frank. However, someone notices that Frank has a better batting average than Tim against both right-handed and left-handed pitchers. How can this happen? (Present your answer in a table.)

Solution to study question 1.2.2

Observe that this problem requests that we create a Simpson's reversal in our data. [Hint: We can use Table 1.2 from the text to scaffold our answer.] Consider the following, somewhat unrealistic (but without loss of generality) batting averages for Frank and Tim against Right- and Left-handed pitchers:

	Frank	Tim
Right-handed	81 hits out of 87 at-bats (.931)	234 hits out of 270 at-bats (.867)
Left-handed	192 hits out of 263 at-bats (.730)	55 hits out of 80 at-bats (.688)
Combined Data	273 hits out of 350 at-bats (.780)	289 hits out of 350 at-bats (.826)

Study question 1.2.3.
Determine, for each of the following causal stories, whether you should use the aggregate or the segregated data to determine the true effect of treatment on recovery.

(a) *There are two treatments used on kidney stones: Treatment A and Treatment B. Doctors are more likely to use Treatment A on large (and therefore, more severe) stones and more likely to use Treatment B on small stones. Should a patient who doesn't know the size of his or her stone examine the general population data, or the stone size-specific data when deciding which treatment they would like to request?*

(b) *There are two doctors in a small town. Each has performed 100 surgeries in his career, which are of two types: one very difficult surgery, and one very easy surgery. The first doctor performs the easy surgery much more often than the difficult surgery, and the* *second doctor performs the difficult surgery more often than the easy surgery. You need surgery, but you do not know whether your case is easy or difficult. Should you consult the success rate of each doctor over all cases, or should you consult their success rates for the easy and difficult cases separately in choosing which surgeon to perform your operation.*

Solution to study question 1.2.3

To answer each of the questions in this section, and based on the structure of these relationships, we consider the causal relationships behind the described scenario to determine which interpretation of the data is valid.

Part (a)

Here, the size of the stone is a common cause of the treatment choice and its recovery outcome. In other words, the size of the stone both affects the likelihood of receiving one treatment over the other, and also the chance of recovery since larger stones are more severe. Moreover, treatment does not change the stone size. As such, the structure of this scenario is identical to that of Example 1.2.1, in which treatment does not affect sex. Similarly, whether or not the patient knows the Size of their stone, we should consult the segregated data conditioned on stone size to make a correct decision.

Part (b)

The same logic as above applies. Paralleling the structure of Example 1.2.1, Difficulty of surgery is a common cause of both doctor choice and recovery rates. In other words, the difficulty of a surgery affects both propensities for choosing one doctor over another as well as the chance of success, since more difficult cases could inherently have less chance of success. As such, whether or not the patient knows the difficulty of their surgery, we should consult the segregated data conditioned on difficulty to make a causally-correct decision.

Study question 1.2.4.
In an attempt to estimate the effectiveness of a new drug, a randomized experiment is conducted. In all, 50% of the patients are assigned to receive the new drug and 50% to receive a placebo. A day before the actual experiment, a nurse hands out lollipops to some patients who show signs of depression, mostly among those who have been assigned to treatment the next day (i.e., the nurse's round happened to take her through the treatment-bound ward). Strangely, the experimental data revealed a Simpson's reversal: Although the drug proved beneficial to the population as a whole, drug takers were less likely to recover than nontakers, among both lollipop receivers and lollipop nonreceivers. Assuming that lollipop sucking in itself has no effect whatsoever on recovery, answer the following questions:

(a) Is the drug beneficial to the population as a whole or harmful?

(b) Does your answer contradict our gender example, where sex-specific data was deemed more appropriate?

(c) Draw a graph (informally) that more or less captures the story. (Look ahead to Section 1.3 if you wish.)

(d) How would you explain the emergence of Simpson's reversal in this story?

(e) Would your answer change if the lollipops were handed out (by the same criterion) a day after the study?

[Hint: Use the fact that receiving a lollipop indicates a greater likelihood of being assigned to drug treatment, as well as depression, which is a symptom of risk factors that lower the likelihood of recovery.]

Solution to study question 1.2.4

The arguments behind this problem are somewhat intricate, but they can be made intuitive if we take an extreme case and assume that, among those in the treatment ward, patients received a lollipop regardless of their health, while among those in the placebo ward only extremely sick patients were given a lollipop. Under these circumstances, the group of lollipop-receiving patients would show a strong correlation between treatment and recovery even if the treatment is totally ineffective; treated individuals consist of typical patients while untreated individuals consist of only extremely sick people. Thus, the treatment will appear to improve chances of recovery even if it has no physical effect on recovery. The same applies to the group of lollipop-denied patients; a spurious correlation will appear between treatment and recovery, merely because the untreated patients where chosen among the very sick.

Such spurious correlations will not occur in the aggregated population because if we disregard the lollipop we find a perfectly randomized experiment; those chosen for treatment are chosen at random from the population, in total disregard of their health status. Another way to understand the difference in populations is to note that, when we compare treated and untreated patients in the lollipop-receiving group we are comparing apples and oranges; these two groups are not *exchangeable* in terms of their health status.

We conclude that, in this example, the aggregated data reveal the correct answer to our question, while the disaggregated data is biased. In Chapter 3 of this book we will see that, in stories of this nature, disaggregation is to be avoided regardless of the specific lollipop-handling strategy used by the nurse. We will further learn to identify such situations directly from the graph, without invoking arguments concerning exchangeability or apples and oranges.

Part (a)

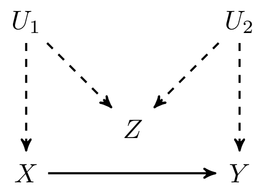
Per the above, we know that disaggregated data is biased, so we instead consult the aggregated data and conclude that the drug is beneficial to the population.

Part (b)

Our decision here does not contradict the gender example from Table 1.1 where we deemed it appropriate to consult the segregated data. In the gender example, gender was not merely correlated with treatment and recovery, it was actually a cause of both. Not so in the present story; lollipop receipt correlates with, but is not a cause of, either treatment or recovery. The two different stories warrant different treatments.

Part (c)

Let X indicate treatment receipt, Z indicate lollipop receipt, Y indicate recovery, and U_1, U_2 indicate two unobserved factors that correlate Z with X and Y, respectively. The causal graph illustrating our story can be modeled as:



Part (d)

Suppose we (incorrectly) decided to use segregated data where we condition on lollipop receipt. Simpson's Reversal could display benefit of the drug to the population (aggregate data) and harm to both lollipop-specific groups (segregated data) by a "trick" of the segregated group sizes, as we've seen many times in this chapter (Table 1.2, for example). Consider that the "got lollipop" group consists of a subset of the "treated" group and that if we got unlucky and gave lollipops to all of the treated individuals who were going to recover, it would give the impression of (negative) association between treatment and recovery even when there is no causal effect of treatment on recovery. The same argument applies for the "didn't get lollipop" group.

Part (e)

Our answer will not change since lollipop receipt is still only spuriously connected to treatment, even if the lollipops were distributed after the study. With these analyses, we always consult the "causal story" behind them.

Study question 1.3.1.
Identify the variables and events invoked in the lollipop story of Study question 1.2.4

Solution to study question 1.3.1

Variables: Let X indicate Treatment / Drug receipt, Z indicate Lollipop receipt, and Y indicate Recovery Status.
Events: "$X = 1$ *and* $Z = 1$ *and* $Y = 1$" indicates the event where an individual takes the drug, receives a lollipop, and recovers (the same applies for other values of each variable).

Table 1.5 The proportion of males and females achieving a given education level

Gender	Highest education achieved	Occurrence (in hundreds of thousands)
Male	Never finished high school	112
Male	High school	231
Male	College	595
Male	Graduate school	242
Female	Never finished high school	136
Female	High school	189
Female	College	763
Female	Graduate school	172

Study question 1.3.2.
Consider Table 1.5 showing the relationship between gender and education level in the U.S. adult population.

(a) Estimate $P(\text{High School})$

(b) Estimate $P(\text{High School OR Female})$

(c) Estimate $P(\text{High School} \mid \text{Female})$

(d) Estimate $P(\text{Female} \mid \text{High School})$.

Solution to study question 1.3.2

Using Table 1.5, for each of the specified quantities of interest, we simply sum over the cases in the matching attributes and divide by the appropriate population.

Part (a)

By marginalization, we can write:

$$\begin{aligned} P(High\ School) &= P(High\ School, Male) + P(High\ School, Female) \\ &= \frac{231 + 189}{112 + 231 + 595 + 242 + 136 + 189 + 763 + 172} \\ &= 0.1721 \end{aligned}$$

Part (b)

Summing over all cases falling in *either* the High School or Female categories, we have:

$$\begin{aligned} P(Female\ or\ High\ School) &= P(Female) + P(Male, High\ School) \\ &= \frac{189 + 136 + 763 + 172 + 231}{112 + 231 + 595 + 242 + 136 + 189 + 763 + 172} \\ &= 0.6111 \end{aligned}$$

Part (c)

By Bayes' conditioning, we can write:

$$\begin{aligned} P(High\ School|Female) &= P(High\ School, Female)/P(Female) \\ &= \frac{189}{136+189+763+172} \\ &= 0.15 \end{aligned}$$

Part (d)

Again by Bayes' conditioning, we can write:

$$\begin{aligned} P(Female|High\ School) &= P(Female, High\ School)/P(High\ School) \\ &= \frac{189}{231+189} \\ &= 0.45 \end{aligned}$$

Study question 1.3.3.

Consider the casino problem described in Section 1.3.7

(a) Compute $P(\text{“craps”}|\text{“11”})$ assuming that there are twice as many roulette tables as craps games at the casino.

(b) Compute $P(\text{“roulette”}|\text{“10”})$ assuming that there are twice as many craps games as roulette tables at the casino.

Solution to study question 1.3.3

Part (a)

Assuming that there are twice as many roulette tables as craps games at the casino, we have:

$$P(\text{“}roulette\text{”}) = 2/3$$

$$P(\text{“}craps\text{”}) = 1/3$$

So, by the law of total probability, we can write our target quantity $P(\text{“11”})$ in terms of what we know:

$$\begin{aligned} P(\text{“11”}) &= P(\text{“11”}|\text{“}craps\text{”})P(\text{“}craps\text{”}) + P(\text{“11”}|\text{“}roulette\text{”})P(\text{“}roulette\text{”}) \\ &= 1/18 * 1/3 + 1/38 * 2/3 \\ &= 37/1026 \\ &= 0.036 \\ P(\text{“}craps\text{”}|\text{“11”}) &= P(\text{“}craps\text{”}, \text{“11”})/P(\text{“11”}) \\ &= \frac{1/18 * 1/3}{37/1026} \\ &= 0.514 \end{aligned}$$

Part (b)

Assuming that there are twice as many craps games as roulette tables at the casino, we have:

$$P(\text{“}roulette\text{”}) = 1/3$$
$$P(\text{“}craps\text{”}) = 2/3$$

We can use the same tactic as in (a) (the law of total probability) to write our target quantity in terms of what we know:

$$\begin{aligned}
P(\text{“10”}) &= P(\text{“10”}|\text{“}craps\text{”})P(\text{“}craps\text{”}) + P(\text{“10”}|\text{“}roulette\text{”})P(\text{“}roulette\text{”}) \\
&= 1/12 * 2/3 + 1/38 * 1/3 \\
&= 11/171 \\
&= 0.064 \\
P(\text{“}roulette\text{”}|\text{“10”}) &= P(\text{“}roulette\text{”}, \text{“10”})/P(\text{“10”}) \\
&= \frac{1/38 * 1/3}{11/171} \\
&= 0.136
\end{aligned}$$

Study question 1.3.4.

Suppose we have three cards. Card 1 has two black faces, one on each side; Card 2 has two white faces; and Card 3 has one white face and one back face. You select a card at random and place it on the table. You find that it is black on the face-up side. What is the probability that the face-down side of the card is also black?

(a) *Use your intuition to argue that the probability that the face-down side of the card is also black is* $1/2$. *Why might it be greater than* $1/2$*?*

(b) *Express the probabilities and conditional probabilities that you find easy to estimate (for example,* $P(C_D = \text{Black})$*), in terms of the following variables:*

$$\begin{aligned}
I &= \text{Identity of the card selected (Card 1, Card 2, or Card 3)} \\
C_D &= \text{Color of the face-down side (Black, White)} \\
C_U &= \text{Color of the face-up side (Black, White)}
\end{aligned}$$

Find the probability that the face-down side of the selected card is black, using your estimates above.

(c) *Use Bayes' theorem to find the correct probability of a randomly selected card's back being black if you observe that its front is black.*

Solution to study question 1.3.4

Part (a)

The face-up side is black, so it is either card 1 or card 3. Given that cards have equal probabilities of being selected, the probability that the face-down side of the card is also black is 1/2. However, cards do not have equal probabilities conditioned on the evidence; if the face-up side is black, the card is more likely to be card 1, so the probability that the face-down side of the card is also black is greater than 1/2.

Part (b)

Since we don't know which card is face-up, we'll use the law of total probability indexing on the card number to compute our quantity of interest.

$$\begin{aligned}
P(C_D = Black) &= P(C_D = Black|I = 1)P(I = 1) + P(C_D = Black|I = 2)P(I = 2) \\
&\quad + P(C_D = Black|I = 3)P(I = 3) \\
&= 1 * 1/3 + 0 * 1/3 + 1/2 * 1/3 \\
&= 1/2
\end{aligned}$$

Part (c)

This is a straightforward application of Bayes' theorem:

$$\begin{aligned}
P(I = 1|C_U = Black) &= \frac{P(C_U = Black|I = 1)P(I = 1)}{P(C_U = Black)} \\
&= \frac{1 * 1/3}{1/2} \\
&= 2/3
\end{aligned}$$

Study question 1.3.5 (Monty Hall)**.**
Prove, using Bayes' theorem, that switching doors improves your chances of winning the car in the Monty Hall problem.

Solution to study question 1.3.5

We'll adopt the variable labeling used in the text, where each may have values indicating doors A, B, C: Let X indicate the door chosen by the player, Y indicate the door hiding the car, and Z indicate the door opened by the host. We want to prove that:

$$P(Y = A|X = A, Z = C) < P(Y = B|X = A, Z = C)$$

So, we'll compute the components of this expression necessary to illustrate this inequality, and then combine them.

$$\begin{aligned}
P(Z = C|X = A) &= P(Z = C|X = A, Y = A)P(Y = A) \\
&\quad + P(Z = C|X = A, Y = B)P(Y = B) \\
&\quad + P(Z = C|X = A, Y = C)P(Y = C) \\
&= 1/2 * 1/3 + 1 * 1/3 + 0 * 1/3 \\
&= 1/2 \\
P(Y = A|X = A, Z = C) &= \frac{P(Z = C|X = A, Y = A)P(Y = A|X = A)}{P(Z = C|X = A)} \\
&= \frac{1/2 * 1/3}{1/2} \\
&= 1/3 \\
P(Y = B|X = A, Z = C) &= 1 - P(Y = A|X = A, Z = C) - P(Y = C|X = A, Z = C) \\
&= 1 - 1/3 - 0 \\
&= 2/3
\end{aligned}$$

Thus, switching doors doubles the chances of winning the car.

Study question 1.3.6.

(a) Prove that, in general, both σ_{XY} and ρ_{XY} vanish when X and Y are independent. [Hint: Use Eqs. (1.16) and (1.17).]

(b) Give an example of two variables that are highly dependent and, yet, their correlation coefficient vanishes.

Solution to study question 1.3.6

Part (a)

By assumption, X and Y are independent, allowing us to write:

$$\begin{aligned}
E(XY) &= \sum_{xy} xy * P(xy) \\
&= \sum xyxy * P(x)P(y) \\
&= \sum_{x} xP(x) * \sum_{y} yP(y) \\
&= E(X)E(Y)
\end{aligned}$$

Using this decomposition, we can now show:

$$\begin{aligned}
\therefore \sigma_{XY} &= E[(X - E(X))(Y - E(Y))] \\
&= E(X)E(Y) - 2E(X)E(Y) + E(XY) \\
&= E(XY) - E(X)E(Y) \\
&= 0 \\
&= \rho_{XY}
\end{aligned}$$

Part (b)

Consider an abstract gambling game with a player and "the house" (e.g., a casino dealer). Let X represent the possible winnings/losses of the player and Y represent the winnings/losses of the house such that $X \in \{-1, 1\}, Y \in \{-1, 0, 1\}$. In this game, the winnings of the house depend on the winnings of the player, as illustrated in the table that follows.
Furthermore, let $P(X = -1) = P(X = 1) = 0.5 \Rightarrow E(X) = 0$

$P(Y\|X)$	$X = -1$	$X = 1$
$Y = -1$	0.5	0.5
$Y = 0$	0.5	0.5
$Y = 1$	0	1

Above, we see that X and Y are dependent ($P(Y = 1|X = -1) \neq P(Y = 1|X = 1)$), yet:

$$\begin{aligned}
E(XY) &= \sum_x \sum_y XY * P(XY) \\
&= \sum_x \sum_y XY * P(Y|X)P(X) \\
&= (-1)(-1) * 0.5 * 0.5 + (-1)(0) * 1 * 0.5 + (-1)(1) * 0 * 0.5 \\
&\quad + (1)(-1) * 0.5 * 0.5 + (1)(0) * 0.5 * 0.5 + (1)(1) * 0 * 0.5 \\
&= 0
\end{aligned}$$

$E(XY) = 0$, $E(X)E(Y) = 0$, so $Cov(X, Y) = 0$.
Thus, we have found a scenario where two variables are dependent, but their correlation coefficient vanishes.

Study question 1.3.7.
Two fair coins are flipped simultaneously to determine the payoffs of two players in the town's casino. Player 1 wins a dollar if and only if at least one coin lands on head. Player 2 receives a dollar if and only if the two coins land on the same face. Let X stand for the payoff of Player 1 and Y for the payoff of Player 2.

(a) Find and describe the probability distributions

$$P(x), P(y), P(x, y), P(y|x) \text{ and } P(x|y)$$

(b) Using the descriptions in (a), compute the following measures:

$$E[X], E[Y], E[Y|X = x], E[X|Y = y]$$

$$Var(X), Var(Y), Cov(X, Y), \rho_{XY}$$

(c) Given that Player 2 won a dollar, what is your best guess of Player 1's payoff?

(d) Given that Player 1 won a dollar, what is your best guess of Player 2's payoff?

(e) Are there two events, $X = x$ and $Y = y$, that are mutually independent?

Solution to study question 1.3.7

Let X and Y stand for the winnings of Player 1 and Player 2, respectively. We have:

Part (a)

The descriptions of these distributions are as follows:
$P(x)$: The probability that player 1 gets x dollars.
$P(y)$: The probability that player 2 gets y dollars.
$P(x, y)$: The probability that player 1 gets x dollars and player 2 gets y dollars.
$P(y|x)$: The probability that player 2 gets y dollars given that player 1 gets x dollars.
$P(x|y)$: The probability that player 1 gets x dollars given that player 2 gets y dollars.

Part (b)

We'll compute each measure by its definition, using the fact that each coin flip is fair and independent:

First, observe that Player 1 wins a dollar if at least 1 of the coins lands on heads. Another way to think about this scenario is that Player 1 loses if both coins land on tails, which we can subtract from 1 to find the probability of them winning. Specifically:

$$P(X = 1) = 1 - P(X = 0) = 1 - P(tails_1)P(tails_2) = 1 - 1/2 * 1/2 = 3/4$$

Computing the expected value follows from Eq. (1.10), summing over all outcomes and their associated probabilities:

$$E[X] = \sum_x x * P(x) = 1 * P(X = 1) + 0 * P(X = 0) = 3/4$$

We'll use a similar approach to computing the winning probability for Player 2 as well as the expected value of their winnings. Observe that the winning conditions for Player 2 are when both coins land on the same face, specifically:

$$P(Y = 1) = P(heads_1)P(heads_2) + P(tails_1)P(tails_2) = 1/2 * 1/2 + 1/2 * 1/2 = 1/2$$

$$E[Y] = \sum_y y * P(y) = 1 * P(Y = 1) + 0 * P(Y = 0) = 1/2$$

To compute the conditional expected values, we will use Eq. (1.13), which intuitively sums over all possible values of the query and weights by the conditional probability of each:

$$\begin{aligned}
E[Y|X=x] &= \sum_y P(y|X=x) \\
&= 1 * P(Y=1|X=x) + 0 * P(Y=0|X=x) \\
&= P(Y=1|X=x) \\
E[X|Y=y] &= \sum_x P(x|Y=y) \\
&= 1 * P(X=1|Y=y) + 0 * P(X=0|Y=y) \\
&= P(X=1|Y=y)
\end{aligned}$$

Next, we can compute the variances of each variable using Eq. (1.15), their covariance using Eq. (1.16), and their correlation coefficient using Eq. (1.17).

$$\begin{aligned}
Var(X) &= E((X-3/4)^2) \\
&= (1-3/4)^2 * P(X=1) + (0-3/4)^2 * P(X=0) \\
&= 1/16 * 3/4 + 9/16 * 1/4 \\
&= 3/16 \\
Var(Y) &= E((Y-1/2)^2) \\
&= (1-1/2)^2 * P(Y=1) + (0-1/2)^2 * P(X=0) \\
&= 1/4 * 1/2 + 1/4 * 1/2 \\
&= 1/4 \\
Cov(X,Y) &= E[(X-3/4)(Y-1/2)] \\
&= 1/4 * 1/2 * P(X=1,Y=1) - 3/4 * 1/2 * P(X=0,Y=1) \\
&\quad + 1/4 * -1/2 * P(X=1,Y=0) - 3/4 * -1/2 * P(X=0,Y=0) \\
&= 1/4 * 1/2 * 1/4 - 3/4 * 1/2 * 1/4 + 1/4 * -1/2 * 1/2 - 3/4 * -1/2 * 0 \\
&= -1/8 \\
\rho_{XY} &= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \\
&= \frac{-1/8}{\sqrt{3/16}\sqrt{1/4}} \\
&= -1/\sqrt{3}
\end{aligned}$$

Part (c)

To answer this query, we know that if both $X = 1$ and $Y = 1$, then the outcome of the two coins must have been both heads, meaning that $P(X = 1, Y = 1) = 1/4$. Furthermore, we can phrase our query as $E[X|Y = 1]$, since we are interested in the expectation of Player 1's winnings having observed that Player 2 won a dollar. Combining this knowledge with our solution to each conditional expected value from part (b) above, we have:

$$\begin{aligned} E[X|Y = 1] &= P(X = 1|Y = 1) \\ &= \frac{P(X = 1, Y = 1)}{Y = 1} \\ &= \frac{1/4}{1/2} \\ &= 1/2 \end{aligned}$$

Part (d)

We use the same strategy as in part (c) above, and have:

$$\begin{aligned} E[Y|X = 1] &= P(Y = 1|X = 1) \\ &= \frac{P(X = 1, Y = 1)}{X = 1} \\ &= \frac{1/4}{3/4} \\ &= 1/3 \end{aligned}$$

Part (e)

Consider what we know about the joint events:

$$\begin{aligned} P(X = 1, Y = 1) = 1/4 \\ P(X = 0, Y = 1) = 1/4 \\ P(X = 1, Y = 0) = 1/2 \\ P(X = 0, Y = 0) = 0 \end{aligned}$$

Now, examining their priors, we have:

$$\begin{aligned} P(X = 1) = 3/4 \\ P(X = 0) = 1/4 \\ P(Y = 1) = P(Y = 0) = 1/2 \end{aligned}$$

Plainly, there are no two values for X and Y such that the product of their priors will equal their joint, i.e., for no two values $X = x, Y = y$ do we have: $P(Y = y, X = x) = P(Y = y) * P(X = x)$. Therefore, we conclude that there are no two mutually independent events.

Study question 1.3.8.

Compute the following theoretical measures of the outcome of a single game of craps (one roll of two independent dice), where X stands for the outcome of Die 1, Z for the outcome of Die 2, and Y for their sum.

(a)

$$E[X], E[Y], E[Y|X = x], E[X|Y = y], \text{ for each value of } x \text{ and } y, \text{ and}$$
$$Var(X), Var(Y), Cov(X, Y), \rho_{XY}, Cov(X, Z)$$

Table 1.6 describes the outcomes of 12 craps games.

Table 1.6 Results of 12 rolls of two fair dice

	X Die 1	Z Die 2	Y Sum
Roll 1	6	3	9
Roll 2	3	4	7
Roll 3	4	6	10
Roll 4	6	2	8
Roll 5	6	4	10
Roll 6	5	3	8
Roll 7	1	5	6
Roll 8	3	5	8
Roll 9	6	5	11
Roll 10	3	5	8
Roll 11	5	3	8
Roll 12	4	5	9

(b) Find the sample estimates of the measures computed in (a), based on the data from Table 1.6. [Hint: Many software packages are available for doing this computation for you.]

(c) Use the results in (a) to determine the best estimate of the sum, Y, given that we measured $X = 3$.

(d) What is the best estimate of X, given that we measured $Y = 4$?

(e) What is the best estimate of X, given that we measured $Y = 4$ and $Z = 1$? Explain why it is not the same as in (d).

Solution to study question 1.3.8

Part (a)

Because we are playing craps, the outcomes of each dice X and Z are, by assumption, independent. However, we know that Y is not independent of either X nor Z, since it is representative of their sum. So, let us first deduce the expected values of each variable individually, which exploits the fact that each dice has a possible outcome of equally likely integers between 1 and 6.

$$\begin{aligned}
E[X] &= 1*1/6 + 2*1/6 + 3*1/6 + 4*1/6 + 5*1/6 + 6*1/6 \\
&= 3.5 \\
&= E[Z] \\
E[Y] &= E[X+Z] \\
&= E[X] + E[Z] \\
&= 7
\end{aligned}$$

Next we consider, without loss of generality, $E[Y|X=x]\ \forall x$. To determine these quantities, we again exploit the facts that $Y = X + Z$ and that X and Z are independent to write:

$$\begin{aligned}
E[Y|X=1] &= E[X+Z|X=1] = E[X|X=1] + E[Z|X=1] = 1 + 3.5 = 4.5 \\
E[Y|X=2] &= E[X+Z|X=2] = E[X|X=2] + E[Z|X=1] = 2 + 3.5 = 5.5 \\
E[Y|X=3] &= E[X+Z|X=3] = E[X|X=3] + E[Z|X=1] = 3 + 3.5 = 6.5 \\
E[Y|X=4] &= E[X+Z|X=4] = E[X|X=4] + E[Z|X=1] = 4 + 3.5 = 7.5 \\
E[Y|X=5] &= E[X+Z|X=5] = E[X|X=5] + E[Z|X=1] = 5 + 3.5 = 8.5 \\
E[Y|X=6] &= E[X+Z|X=6] = E[X|X=6] + E[Z|X=1] = 6 + 3.5 = 9.5
\end{aligned}$$

By a similar token, we consider quantities $E[X|Y=y]\ \forall y$. These can be computed by the same method above, exploiting the knowledge that:

$$\begin{aligned}
E[X|Y=y] &= E[Y-Z|Y=y] \\
&= E[Y|Y=y] - E[Z|Y=y] \\
&= y - E[X|Y=y] \\
2*E[X|Y=y] &= y \\
E[X|Y=y] &= y/2
\end{aligned}$$

So, using this derivation, we see that:

$$\begin{aligned}
E[X|Y=2] &= 1 \\
E[X|Y=3] &= 1.5 \\
E[X|Y=4] &= 2 \\
E[X|Y=5] &= 2.5 \\
E[X|Y=6] &= 3 \\
&etc.
\end{aligned}$$

Next we compute the variances of our variables, which is a simple application of Eq. (1.15):

$$
\begin{aligned}
Var(X) &= E[(X - E[X])^2] \\
&= E[(X - 3.5)^2] \\
&= (1 - 3.5)^2 * 1/6 + (2 - 3.5)^2 * 1/6 + (3 - 3.5)^2 * 1/6 \\
&\quad + (4 - 3.5)^2 * 1/6 + (5 - 3.5)^2 * 1/6 + (6 - 3.5)^2 * 1/6 \\
&= 2.917 \\
&= Var(Z) \\
Var(Y) &= E[(Y - E[Y])^2] \\
&= E[(Y - 7)^2] \\
&= E[Y^2 - 14Y + 49] \\
&= E[Y^2] - 98 + 49 \\
&= E[X^2 + 2XZ + Z^2] - 49 \\
&= 2(E[X^2] + E[XZ]) - 49 \\
&= 2(\frac{91}{6} + \frac{21 * 21}{36}) - 49 \\
&= 5.833
\end{aligned}
$$

Now knowing our variances, we can compute $Cov(X, Y)$ and $Cov(X, Z)$ through application of Eq. (1.16):

$$
\begin{aligned}
Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\
&= E[(X - 3.5)(Y - 7)] \\
&= E[XY - 3.5Y - 7X + 24.5] \\
&= E[XY] - 24.5 - 24.5 + 24.5 \\
&= E[X(X + Z)] - 24.5 \\
&= E[X^2] + E[XZ] - 24.5 \\
&= 91/6 + 21 * 21/36 - 24.5 \\
&= 2.917 \\
Cov(X, Z) &= E[(X - 3.5)(Z - 3.5)] \\
&= E[XZ] - 3.5^2 \\
&= 21 * 21/36 - 3.5^2 \\
&= 0
\end{aligned}
$$

Intuitively, we can check our answer that $Cov(X, Z) = 0$ because X and Z are independent dice rolls. Finally, we compute the correlation between X and Y by Eq. (1.17):

$$\begin{aligned}\rho_{XY} &= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \\ &= \frac{2.917}{\sqrt{2.917}\sqrt{5.833}} \\ &= 0.707\end{aligned}$$

Part (b)

You can use programming packages in R (see DAGitty package), Matlab, Python(Numpy), etc. to calculate the quantities of interest from the sample in Table 1.6. The same computational strategies we used in part (a) apply, except that now our frequencies come from the data rather than our analysis of craps. Specifically, we get:

$$\begin{aligned}E[X] &= 4.33 \\ E[Y] &= 8.5 \\ Var(X) &= 2.389 \\ Var(Y) &= 1.75 \\ Cov(X,Y) &= 1.545 \\ Cov(X,Z) &= -1.06 \\ \rho_{XY} &= 0.756\end{aligned}$$

Part (c)

From our computations in part (a), we have $E[Y|X = 3] = 6.5$

Part (d)

From our computations in part (a), we have $E[X|Y = 4] = 2$

Part (e)

We can compute $E[X|Y = 4, Z = 1]$ by application of Eq. (1.13), summing over the six possible values of X and the probabilities of $Y = 4, Z = 1$ associated with each:

$$\begin{aligned}E[X|Y=4,Z=1] &= \sum_x x\, P(X=x|Y=4,Z=1) \\ &= 1 * P(X=1|Y=4,Z=1) + 2 * P(X=2|Y=4,Z=1) \\ &\quad + 3 * P(X=3|Y=4,Z=1) + 4 * P(X=4|Y=4,Z=1) \\ &\quad + 5 * P(X=5|Y=4,Z=1) + 6 * P(X=6|Y=4,Z=1) \\ &= 1*0 + 2*0 + 3*1 + 4*0 + 5*0 + 6*0 \\ &= 3\end{aligned}$$

Intuitively, this quantity will not be the same as in part (d) because knowing that $Z = 1$ precludes the possibilities of some values of X given that $Y = X + Z = 4$. For example, in (d), we allowed for the possibility that $X = 2$ (and that therefore $Z = 2$ in order to sum to $Y = 4$), which is impossible in this problem given that $Z = 1$.

Study question 1.3.9.

(a) Prove Eq. (1.22) using the orthogonality principle. [Hint: Follow the treatment of Eq. (1.26).]

(b) Find all partial regression coefficients

$$R_{YX \cdot Z}, R_{XY \cdot Z}, R_{YZ \cdot X}, R_{ZY \cdot X}, R_{XZ \cdot Y}, \text{ and } R_{ZX \cdot Y}$$

for the craps game described in Study question 1.3.8. [Hint: Apply Eq. (1.27) and use the variances and covariances computed for part (a) of this question.]

Solution to study question 1.3.9

Part (a)

By assumption, from Eq. (1.21), we have:

$$Y = a + bX$$

So, using the linear property of expected value, we know that:

$$E[Y] = E[a + bX] = a + bE[X]$$

Thus, using the hint to follow the treatment of Eq. (1.26), we have:

$$E[XY] = aE[X] + bE[X^2]$$

Finally, the above allows us to prove Eq. (1.22):

$$b = \frac{E[XY] - E[X]E[Y]}{E[X^2] - E^2[X]} = \frac{\sigma_{XY}}{\sigma_X^2}$$

Part (b)

From our answer to study question 1.3.8, we have:
$\sigma_{XY} = \sigma_{ZY} = 2.917$, $\sigma_{XZ} = 0$, $\sigma_Y^2 = 5.833$, $\sigma_X^2 = \sigma_Z^2 = 2.917$.

Using the above, we can use Eq. (1.27) to compute each partial regression coefficient:

$$R_{YX\cdot Z} = \frac{\sigma_Z^2\sigma_{YX} - \sigma_{YZ}\sigma_{ZX}}{\sigma_X^2\sigma_Z^2 - \sigma_{XZ}^2} = \frac{2.917^2 - 0}{2.917^2 - 0} = 1$$
$$R_{XY\cdot Z} = \frac{\sigma_Z^2\sigma_{XY} - \sigma_{XZ}\sigma_{ZY}}{\sigma_Y^2\sigma_Z^2 - \sigma_{YZ}^2} = \frac{2.917^2 - 0}{5.833 * 2.917 - 2.917^2} = 1$$
$$R_{YZ\cdot X} = \frac{\sigma_X^2\sigma_{YZ} - \sigma_{YX}\sigma_{XZ}}{\sigma_Z^2\sigma_X^2 - \sigma_{ZX}^2} = \frac{2.917^2 - 0}{2.917^2 - 0} = 1$$
$$R_{ZY\cdot X} = \frac{\sigma_X^2\sigma_{ZY} - \sigma_{ZX}\sigma_{XY}}{\sigma_Y^2\sigma_X^2 - \sigma_{YX}^2} = \frac{2.917^2 - 0}{5.833 * 2.917 - 2.917^2} = 1$$
$$R_{XZ\cdot Y} = \frac{\sigma_Y^2\sigma_{XZ} - \sigma_{XY}\sigma_{YZ}}{\sigma_Z^2\sigma_Y^2 - \sigma_{ZY}^2} = \frac{0 - 2.917^2}{5.833 * 2.917 - 2.917^2} = -1$$
$$R_{ZX\cdot Y} = \frac{\sigma_Y^2\sigma_{ZX} - \sigma_{ZY}\sigma_{YX}}{\sigma_Y^2\sigma_X^2 - \sigma_{YX}^2} = \frac{0 - 2.917^2}{5.833 * 2.917 - 2.917^2} = -1$$

Study question 1.4.1.
Consider the graph shown in Figure 1.8:

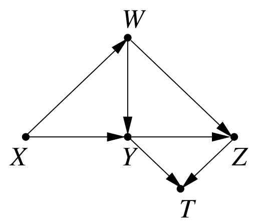


Figure 1.8: A directed graph used in Study question 1.4.1

(a) Name all of the parents of Z.

(b) Name all the ancestors of Z.

(c) Name all the children of W.

(d) Name all the descendants of W.

(e) Draw all (simple) paths between X and T (i.e., no node should appear more than once).

(f) Draw all the directed paths between X and T.

Solution to study question 1.4.1

A basic interactive tutorial where students can practice their knowledge of graph terminology is provided at dagitty.net/learn/graphs/

A more advanced tutorial where students can apply these terms to recognize causal concepts like "mediator" and "confounder" is provided at dagitty.net/learn/graphs/roles.html

An *R* solution of this exercise is provided at dagitty.net/primer/1.4.1

Part (a)

Parents of Z: W, Y

Part (b)

Ancestors of Z: X, W, Y

Part (c)

Children of W: Y, Z

Part (d)

Descendants of W: Y, Z, T

Part (e)

Assuming cycles are not allowed, the simple paths between X and T are:
$\{X,Y,T\}, \{X,Y,Z,T\}, \{X,Y,W,Z,T\}, \{X,W,Y,T\}, \{X,W,Y,Z,T\}, \{X,W,Z,T\}, \{X,W,Z,Y,T\}$

Part (f)

Assuming cycles are not allowed, the directed paths between X and T are: $\{X,Y,T\}, \{X,Y,Z,T\}, \{X,W,Y,T\}, \{X,W,Y,Z,T\}, \{X,W,Z,T\}$

Study question 1.5.1.
Suppose we have the following SCM. Assume all exogenous variables are independent and that the expected value of each is 0.
SCM 1.5.1.

$$V = \{X,Y,Z\}, \quad U = \{U_X, U_Y, U_Z\}, \quad F = \{f_X, f_Y, f_Z\}$$
$$f_X : X = u_X$$
$$f_Y : Y = \frac{X}{3} + U_Y$$
$$f_Z : Z = \frac{Y}{16} + U_Z$$

(a) Draw the graph that complies with the model.

(b) Determine the best guess of the value (expected value) of Z, given that we observe $Y = 3$.

(c) *Determine the best guess of the value of Z, given that we observe $X = 3$.*

(d) *Determine the best guess of the value of Z, given that we observe $X = 1$ and $Y = 3$.*

(e) *Assume that all exogenous variables are normally distributed with zero means and unit variance, that is, $\sigma = 1$.*

(i) *Determine the best guess of X, given that we observed $Y = 2$.*

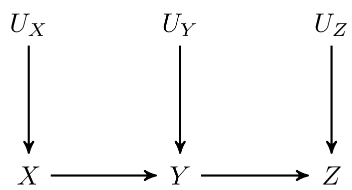
(ii) *(Advanced) Determine the best guess of Y, given that we observed $X = 1$ and $Z = 3$. [Hint: You may wish to use the technique of multiple regression, together with the fact that, for every three normally distributed variables, say X, Y, and Z, we have $E[Y|X = x, Z = z] = R_{YX \cdot Z}x + R_{YZ \cdot X}z$.]*

Solution to study question 1.5.1

An *R* solution of this exercise is provided at dagitty.net/primer/1.5.1

Part (a)

The following graph complies with the given model:



Part (b)

Assuming that U_X, U_Y, U_Z are independent and have 0 means, we have:

$$
\begin{aligned}
E[Z|Y = 3] &= E[Y/16 + U_Z|Y = 3] \\
&= E[Y|Y = 3]/16 + E[U_Z|Y = 3] \\
&= 3/16 + 0 \\
&= 3/16
\end{aligned}
$$

Part (c)

$$
\begin{aligned}
E[Z|X = 3] &= E[Y/16 + U_Z|X = 3] \\
&= E[Y|X = 3]/16 + E[U_Z|X = 3] \\
&= E[X/3 + U_Y|X = 3]/16 + 0 \\
&= E[X|X = 3]/3/16 + E[U_Y|X = 3]/16 \\
&= 3/3/16 + 0/16 \\
&= 1/16
\end{aligned}
$$

Part (d)

$$\begin{aligned} E[Z|X=1,Y=3] &= E[Y/16+U_Z|X=1,Y=3] \\ &= E[Y|X=1,Y=3]/16+E[U_Z|X=1,Y=3] \\ &= 3/16+0 \\ &= 3/16 \end{aligned}$$

(e.i)

$$\begin{aligned} E[X|Y=2] &= E[3Y-3U_Y|Y=2] \\ &= 3E[Y|Y=2]-3E[U_Y|Y=2] \\ &= 3*2-3*0 \\ &= 6 \end{aligned}$$

(e.ii)

$$\begin{aligned} E[Y|X=1,Z=3] &= \alpha x+\beta z \\ &= R_{YX\cdot Z}x+R_{YZ\cdot X}z \\ &= \frac{\sigma_Z^2\sigma_{YX}-\sigma_{YZ}\sigma_{XZ}}{\sigma_X^2\sigma_Z^2-\sigma_{XZ}^2}x+\frac{\sigma_X^2\sigma_{YZ}-\sigma_{YX}\sigma_{XZ}}{\sigma_X^2\sigma_Z^2-\sigma_{XZ}^2}z \\ &= \frac{\sigma_Z^2\sigma_{YX}-\sigma_{YZ}\sigma_{XZ}}{\sigma_X^2\sigma_Z^2-\sigma_{XZ}^2}+\frac{\sigma_X^2\sigma_{YZ}-\sigma_{YX}\sigma_{XZ}}{\sigma_X^2\sigma_Z^2-\sigma_{XZ}^2}3 \end{aligned}$$

$\because \sigma_X^2=1, \sigma_Y^2=10/9, \sigma_Z^2=10/(256*9)+1, \sigma_{XY}=1/3, \sigma_{YZ}=5/72, \sigma_{XZ}=1/48$
$\therefore E[Y|X=1,Z=3]=400/771$

Study question 1.5.2.
Assume that a population of patients contains a fraction r of individuals who suffer from a certain fatal syndrome Z, which simultaneously makes it uncomfortable for them to take a life-prolonging drug X (Figure 1.8).

Let $Z=z_1$ and $Z=z_0$ represent, respectively, the presence and absence of the syndrome, $Y=y_1$ and $Y=y_0$ represent death and survival, respectively, and $X=x_1$ and $X=x_0$ represent taking and not taking the drug. Assume that patients not carrying the syndrome, $Z=z_0$, die with probability p_2 if they take the drug and with probability p_1 if they don't. Patients carrying the syndrome, $Z=z_1$, on the other hand, die with probability p_3 if they do not take the drug and with probability p_4 if they do take the drug. Further, patients having the syndrome are more likely to avoid the drug, with probabilities $q_1=P(x_1|z_0)$ and $q_2=p(x_1|z_1)$.

(a) Based on this model, compute the joint distributions $P(x,y,z), P(x,y), P(x,z)$ and $P(y,z)$ for all values of $x, y,$ and z, in terms of the parameters $(r,p_1,p_2,p_3,p_4,q_1,q_2)$. [Hint: Use the product decomposition of Section 1.5.2.]

(b) Calculate the difference $P(y_1|x_1) - P(y_1|x_o)$ for three populations: (1) those carrying the syndrome, (2) those not carrying the syndrome and (3) the population as a whole.

(c) Using your results for (b), find a combination of parameters that exhibits Simpson's reversal.

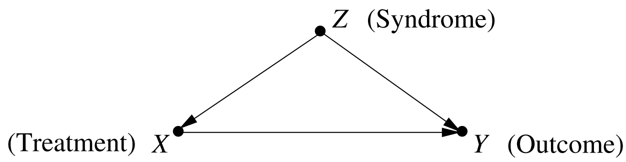


Figure 1.8: Model showing an unobserved syndrome, Z, affecting both treatment (X) and outcome (Y)

Solution to study question 1.5.2

Part (a)

The following two tables describe our distributions of interest:

$P(Y|X,Z)$	y	x	z
p_1	1	0	0
p_2	1	1	0
p_3	1	0	1
p_4	1	1	1

$P(X|Z)$	x	z
q_1	1	0
q_2	1	1

We also have that: $P(Z_1) = r$

By the chain rule, we know that: $P(x, y, z) = P(y|x, z)P(x|z)P(z)$

So, substituting the table rows into the above factorization, we have:

$$P(x_0, y_0, z_0) = (1 - p_1)(1 - q_1)(1 - r)$$
$$P(x_0, y_0, z_1) = (1 - p_3)(1 - q_2)(r)$$
$$P(x_0, y_1, z_0) = (p_1)(1 - q_1)(1 - r)$$
$$P(x_0, y_1, z_1) = (p_3)(1 - q_2)(r)$$
$$P(x_1, y_0, z_0) = (1 - p_2)(q_1)(1 - r)$$
$$P(x_1, y_0, z_1) = (1 - p_4)(q_2)(r)$$
$$P(x_1, y_1, z_0) = (p_2)(q_1)(1 - r)$$
$$P(x_1, y_1, z_1) = (p_4)(q_2)(r)$$

By marginalization, we know that: $P(x, y) = \Sigma_z P(x, y, z)$
So, again substituting our table rows into the above, the joint distribution of X, Y reads:

$$P(x_0, y_0) = P(x_0, y_0, z_0) + P(x_0, y_0, z_1) = (1 - p_1)(1 - q_1)(1 - r) + (1 - p_3)(1 - q_2)(r)$$
$$P(x_0, y_1) = P(x_0, y_1, z_0) + P(x_0, y_1, z_1) = (p_1)(1 - q_1)(1 - r) + (p_3)(1 - q_2)(r)$$
$$P(x_1, y_0) = P(x_1, y_0, z_0) + P(x_1, y_0, z_1) = (1 - p_2)(q_1)(1 - r) + (1 - p_4)(q_2)(r)$$
$$P(x_1, y_1) = P(x_1, y_1, z_0) + P(x_1, y_1, z_1) = (p_2)(q_1)(1 - r) + (p_4)(q_2)(r)$$

Similarly, for X, Z we know that: $P(x, z) = \Sigma_y P(x, y, z)$

$$P(x_0, z_0) = (1 - p_1)(1 - q_1)(1 - r) + (p_1)(1 - q_1)(1 - r)$$
$$P(x_0, z_1) = (1 - p_3)(1 - q_2)(r) + (p_3)(1 - q_2)(r)$$
$$P(x_1, z_0) = (1 - p_2)(q_1)(1 - r) + (p_2)(q_1)(1 - r)$$
$$P(x_1, z_1) = (1 - p_4)(q_2)(r) + (p_4)(q_2)(r)$$

And for Y, Z, we know that: $P(y, z) = \Sigma_x P(x, y, z)$

$$P(y_0, z_0) = (1 - p_1)(1 - q_1)(1 - r) + (1 - p_2)(q_1)(1 - r)$$
$$P(y_0, z_1) = (1 - p_3)(1 - q_2)(r) + (1 - p_4)(q_2)(r)$$
$$P(y_1, z_0) = (p_1)(1 - q_1)(1 - r) + (p_2)(q_1)(1 - r)$$
$$P(y_1, z_1) = (p_3)(1 - q_2)(r) + (p_4)(q_2)(r)$$

Part (b)

(b.1)

$$P(y_1|x_1, z_1) - P(y_1|x_0, z_1) = p_4 - p_3$$

(b.2)

$$P(y_1|x_1, z_0) - P(y_1|x_0, z_0) = p_2 - p_1$$

(b.3)

$$\begin{aligned}
P(y_1|x_1) - P(y_1|x_0) &= \frac{P(y_1, x_1)}{P(x_1)} - \frac{P(y_1, x_0)}{P(x_0)} \\
&= \frac{P(y_1, x_1)}{P(x_1, y_1) + P(x_1, y_0)} - \frac{P(y_1, x_0)}{P(x_0, y_1) + P(x_0, y_0)} \\
&= \frac{p_2 q_1 (1-r) + p_4 q_2 r}{r q_2 + (1-r) q_1} - \frac{p_1 (1-q_1)(1-r) + p_3 (1-q_2) r}{r(1-q_2) + (1-r)(1-q_1)}
\end{aligned}$$

Part (c)

To elicit Simpson's reversal, we want to find a combination of parameters such that parts (b.1) and (b.2) above have a different sign than (b.3). As such, consider the following parameterization:

$$p_1 = 0.1, p_2 = 0, p_3 = 0.3, p_4 = 0.2, q_1 = 0, q_2 = 1, r = 0.1$$

Now, substituting the above into (b.1), (b.2), and (b.3), we have:

$$\begin{aligned}
P(y_1|x_1, z_1) - P(y_1|x_0, z_1) &= p_4 - p_3 = -0.1 \\
P(y_1|x_1, z_0) - P(y_1|x_0, z_0) &= p_2 - p_1 = -0.1 \\
P(y_1|x_1) - P(y_1|x_0) &= \frac{P(y_1, x_1)}{P(x_1)} - \frac{P(y_1, x_0)}{P(x_0)} \\
&= \frac{P(y_1, x_1)}{P(x_1, y_1) + P(x_1, y_0)} - \frac{P(y_1, x_0)}{P(x_0, y_1) + P(x_0, y_0)} \\
&= \frac{p_2 q_1 (1-r) + p_4 q_2 r}{r q_2 + (1-r) q_1} - \frac{p_1 (1-q_1)(1-r) + p_3 (1-q_2) r}{r(1-q_2) + (1-r)(1-q_1)} \\
&= 0.2 - 0.1 \\
&= 0.1
\end{aligned}$$

Study question 1.5.3.
Consider a graph $X_1 \to X_2 \to X_3 \to X_4$ of binary random variables, and assume that the conditional probabilities between any two consecutive variables are given by

$$\begin{aligned}
P(X_i = 1|X_{i-1} = 1) &= p \\
P(X_i = 1|X_{i-1} = 0) &= q \\
P(X_1 = 1) &= p_0
\end{aligned}$$

Compute the following probabilities

$$P(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0)$$
$$P(X_4 = 1|X_1 = 1)$$
$$P(X_1 = 1|X_4 = 1)$$
$$P(X_3 = 1|X_1 = 0, X_4 = 1)$$

Solution to study question 1.5.3

Part (1)

We'll begin the first problem using the chain rule and factor it as:

$$\begin{aligned} P(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0) &= P(X_1 = 1)P(X_2 = 0|X_1 = 1) \\ &\quad * P(X_3 = 1|X_2 = 0)P(X_4 = 0|X_3 = 1) \\ &= p_0(1-p)q(1-p) \\ &= p_0(1-p)^2 q \end{aligned}$$

We can compute the next three quantities using marginalization and Bayes' conditioning.

Part (2)

$$\begin{aligned} P(X_4 = 1|X_1 = 1) &= \frac{P(X_4 = 1, X_1 = 1)}{P(X_1 = 1)} \\ &= \frac{\Sigma_{X_2,X_3} P(X_4 = 1, X_1 = 1, X_2, X_3)}{P(X_1 = 1)} \\ &= \frac{p_0 p^3 + 2p_0 pq(1-p) + p_0 q(1-p)(1-q)}{p_0} \end{aligned}$$

Part (3)

$$\begin{aligned} P(X_1 = 1|X_4 = 1) &= \frac{P(X_1 = 1, X_4 = 1)}{P(X_4 = 1)} \\ &= \frac{\Sigma_{X_2,X_3} P(X_4 = 1, X_1 = 1, X_2, X_3)}{\Sigma_{X_1,X_2,X_3} P(X_1, X_2, X_3, X_4 = 1)} \\ &= \frac{p_0 p^3 + 2p_0 pq(1-p) + p_0 q(1-p)(1-q)}{\begin{array}{l} p_0 p^3 + 2p_0 pq(1-p) + p_0 q(1-p)(1-q) \\ \quad + (1-p_0)(qp^2 + q^2(1-p) + qp(1-q) + q(1-q)^2) \end{array}} \end{aligned}$$

Part (4)

$$\begin{aligned} P(X_3 = 1|X_1 = 0, X_4 = 1) &= \frac{P(X_1 = 0, X_3 = 1, X_4 = 1)}{P(X_1 = 0, X_4 = 1)} \\ &= \frac{\Sigma_{X_2} P(X_2, X_1 = 0, X_3 = 1, X_4 = 1)}{\Sigma_{X_2,X_3} P(X_2, X_3, X_1 = 0, X_4 = 1)} \\ &= \frac{(1 - p_0)(1 - q)pq + (1 - p_0)qp^2}{(1 - p_0)(qp^2 + q^2(1 - p) + qp(1 - q) + q(1 - q)^2)} \end{aligned}$$

Study question 1.5.4.
Define the structural model that corresponds to the Monty Hall problem, and use it to describe the joint distribution of all variables.

Solution to study question 1.5.4

Again, we'll adopt the variables from the text: let X indicate the door chosen by the player, Y indicate the door hiding the car, and Z indicate the door opened by the host.

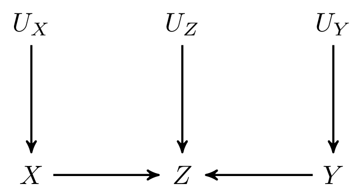
From the story, we know that each door has an equal chance of being the winner, and that each has an equal chance of being selected by the player. This suggests that both X and Y are selected independently from one another as a function of some unmodeled factors.

Furthermore, we know that the door revealed by the host, Z, will be neither the one opened by the player, X, nor the winning door, Y, and in the event that $X = Y$, then Z has an equal chance of being one of the two remaining doors. These observations suggest that Z must be a function of not only X, Y, but also of some unmodeled factors whenever $X = Y$.

Combining our observations from above gives us the following model specification:

$$V = \{X, Y, Z\}, U = \{U_X, U_Y, U_Z\}, F = \{f\},$$
$$X = U_X, Y = U_Y, Z = f(X, Y) + U_Z$$

We can also depict this model graphically:



And lastly, the joint distribution can be factorized by using the chain rule and independence relations (i.e., the Markovian factorization, Eq. (1.29), which decomposes a joint distribution into family factors) to write:

$$P(X, Y, Z) = P(Z|X, Y)P(Y)P(X)$$

So, to explore some example queries, we begin by acknowledging that each door has an equal chance of being the winner and the one chosen by the player, namely:

$$P(Y) = 1/3 \, \forall y \in Y \text{ and } P(X) = 1/3 \, \forall x \in X$$

Furthermore, we know that the host cannot reveal the winning door or the one chosen by the player, so:

$$P(Z|X,Y) = 0 \, \forall z = x \; or \; z = y$$

Lastly, we know that the host *must* open the last remaining door if X and Y are different, and has an equal chance of choosing one of the remaining two when $X = Y$, giving us:

$$P(Z|X,Y) = 1 \, \forall z \neq x \; and \; z \neq y \; and \; x \neq y$$
$$P(Z|X,Y) = 0.5 \, \forall z \neq x \; and \; z \neq y \; and \; x = y$$

These observations allow us to compute arbitrary probability queries because we know the decomposition of the joint distribution. For example, for doors A, B, and C:

$$\begin{aligned} P(X = A, Y = B, Z = C) &= P(Z = C|X = A, Y = B)P(Y = B)P(X = A) \\ &= 1 * 1/3 * 1/3 \\ &= 1/9 \end{aligned}$$

Study Questions and Solutions for Chapter 2

Study question 2.3.1.

$$X \rightarrow R \rightarrow S \rightarrow T \leftarrow U \leftarrow V \rightarrow Y$$

Figure 2.5: A directed graph for demonstrating conditional independence (error terms are not shown explicitly)

(a) List all pairs of variables in Figure 2.5 that are independent conditional on the set $Z = \{R, V\}$.

(b) For each pair of nonadjacent variables in Figure 2.5, give a set of variables that, when conditioned on, renders that pair independent.

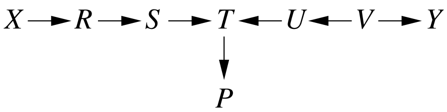


Figure 2.6: A directed graph in which P is a descendant of a collider

(c) List all pairs of variables in Figure 2.6 that are independent conditional on the set $Z = \{R, P\}$.

(d) For each pair of nonadjacent variables in Figure 2.6, give a set of variables that, when conditioned on,renders that pair independent.

(e) Suppose we generate data by the model described in Figure 2.5, and we fit them with the linear equation $Y = a + bX + cZ$. Which of the variables in the model may be chosen for Z so as to guarantee that the slope b would be equal to zero? [Hint: Recall, a non zero slope implies that Y and X are dependent given Z.]

(f) Continuing question (e), suppose we fit the data with the equation:

$$Y = a + bX + cR + dS + eT + fP$$

which of the coefficients would be zero?

Solution to study question 2.3.1

An interactive tutorial explaining *d*-separation is provided at dagitty.net/learn/dsep/

An *R* solution of this exercise is provided at dagitty.net/primer/2.3.1

To determine if two variables are independent in a network, we consider all simple paths between them and determine if all paths are "blocked" by Rules 1, 2, and 3 in this chapter that detail the graphical criteria for conditional independence. If a single path is not blocked, then the two variables are dependent. They are therefore independent (or conditionally independent, when conditioning on some other variables) whenever all simple paths between them are blocked.

Part (a)

In the table below, each row may be read as: "X is independent of Y given Z because..." where X is the variable listed in the first column of every row, Y is listed in the second column, Z in the third, and an explanation in the fourth. Note that for each row, X and Y may be swapped with the same valid claim of independence.

X	***Y***	***Z***	**Reason**
X	S	$\{R, V\}$	$X \rightarrow R \rightarrow S$ is blocked at chain $X \rightarrow R \rightarrow S$ (R is given)
X	T	$\{R, V\}$	$X \rightarrow R \rightarrow S \rightarrow T$ is blocked at chain $X \rightarrow R \rightarrow S$ (R is given)
X	U	$\{R, V\}$	$X \rightarrow R \rightarrow S \rightarrow T \leftarrow U$ is blocked at chain $X \rightarrow R \rightarrow S$ (R is given)
X	Y	$\{R, V\}$	$X \rightarrow R \rightarrow S \rightarrow T \rightarrow U \leftarrow V \leftarrow Y$ is blocked at chain $X \rightarrow R \rightarrow S$ (R is given)
S	U	$\{R, V\}$	$S \rightarrow T \leftarrow U$ is blocked at collider $S \rightarrow T \leftarrow U$ (neither T nor descendants of T given)
S	Y	$\{R, V\}$	$S \rightarrow T \leftarrow U \leftarrow V \rightarrow Y$ is blocked at collider $S \rightarrow T \leftarrow U$ (neither T nor descendants of T given)
T	Y	$\{R, V\}$	$T \leftarrow U \leftarrow V \rightarrow Y$ is blocked at fork $U \leftarrow V \rightarrow Y$ (V is given)
U	Y	$\{R, V\}$	$U \leftarrow V \rightarrow Y$ is blocked at fork $U \leftarrow V \rightarrow Y$ (V is given)

Part (b)

Answers to part (b) can be found in the following table, formatted in the same way as (a) above:

X	***Y***	***Z***	**Reason**
X	S	$\{R\}$	$X \to R \to S$ is blocked at chain $X \to R \to S$ (R is given)
X	T	$\{R\}$	$X \to R \to S \to T$ is blocked at chain $X \to R \to S$ (R is given)
X	U	$\{R\}$	$X \to R \to S \to T \leftarrow U$ is blocked at chain $X \to R \to S$ (R is given)
X	V	$\{R\}$	$X \to R \to S \to T \leftarrow U \leftarrow V$ is blocked at chain $X \to R \to S$ (R is given)
X	Y	$\{R\}$	$X \to R \to S \to T \leftarrow U \leftarrow V \to Y$ is blocked at chain $X \to R \to S$ (R is given)
R	T	$\{S\}$	$R \to S \to T$ is blocked at chain $R \to S \to T$ (S is given)
R	U	$\{S\}$	$R \to S \to T \leftarrow U$ is blocked at chain $R \to S \to T$ (S is given)
R	V	$\{S\}$	$R \to S \to T \leftarrow U \leftarrow V$ is blocked at chain $R \to S \to T$ (S is given)
R	Y	$\{S\}$	$R \to S \to T \leftarrow U \leftarrow V \to Y$ is blocked at chain $R \to S \to T$ (S is given)
S	U	$\{\}$	$S \to T \leftarrow U$ is blocked at collider $S \to T \leftarrow U$ (neither T nor descendants of T are given)
S	V	$\{\}$	$S \to T \leftarrow U \leftarrow V$ is blocked at collider $S \to T \leftarrow U$ (neither T nor descendants of T are given)
S	Y	$\{\}$	$S \to T \leftarrow U \leftarrow V \to Y$ is blocked at collider $S \to T \leftarrow U$ (neither T nor descendants of T are given)
T	V	$\{U\}$	$T \leftarrow U \leftarrow V$ is blocked at chain $T \leftarrow U \leftarrow V$ (U is given)
T	Y	$\{U\}$	$T \leftarrow U \leftarrow V \to Y$ is blocked at chain $T \leftarrow U \leftarrow V$ (U is given)
U	Y	$\{V\}$	$U \leftarrow V \to Y$ is blocked at fork $U \leftarrow V \to Y$ (V is given)

Part (c)

Observe that conditioning on $\{R, P\}$ blockes only the chain $X \to R \to S$ (R is given) and opens the collider $S \to T \leftarrow U$ (P, a descendant of T, is given). Thus, we render only X independent of all other variables in the model; specifically, the pairs of independent variables conditioned on $\{R, P\}$ are: $(X, R), (X, S), (X, T), (X, P), (X, U), (X, V), (X, Y)$

Part (d)

Now that we're familiar with Figure 2.6, we can summarize independence relationships in the following table, which is similar to the previous two in parts (a) and (b) except that every row may be read, "variable X is independent of all variables in set Y given set Z."

X	**Y**	**Z**	**Reason**
X	$\{S, T, U, V, Y, P\}$	$\{R\}$	Chain blocked at $X \rightarrow R \rightarrow S$ (R is given)
R	$\{T, U, V, Y, P\}$	$\{S\}$	Chain blocked at $R \rightarrow S \rightarrow T$ (S is given)
S	$\{U, V, Y\}$	$\{\}$	Collider blocked at $S \rightarrow T \leftarrow U$ (Neither T nor descendants of T are given)
P	$\{S, U, V, Y\}$	$\{T\}$	Chains blocked at $S \rightarrow T \rightarrow P$ (T is given) and $U \rightarrow T \rightarrow P$ (T is given)
T	$\{V, Y\}$	$\{U\}$	Chain blocked at $V \rightarrow U \rightarrow T$ (U is given)
U	$\{Y\}$	$\{V\}$	Fork blocked at $U \leftarrow V \rightarrow Y$ (V is given)

Part (e)

Since Y and X are independent conditional on any member of the set $\{R, S, U, V\}$, we may choose Z to be any of these variables.

Part (f)

To determine which slopes will be equal to zero, we can again consider if Y is independent of each variable given the other variables on the RHS of our equation. Specifically:

1. b (the slope associated with X) will be zero, since Y and X are independent given R, S, T, P.
2. c (the slope associated with R) will be zero, since Y and R are independent given X, S, T, P.
3. f (the slope associated with P) will be zero, since Y and P are independent given X, R, S, T.

Study question 2.4.1.

Figure 2.9 below represents a causal graph from which the error terms have been deleted. Assume that all those errors are mutually independent.

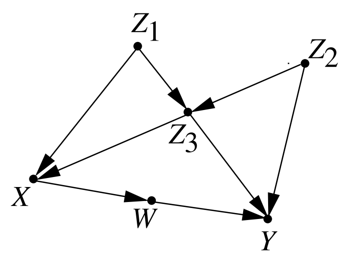


Figure 2.9: A causal graph used in study question 2.4.1, all U terms (not shown) are assumed independent

(a) *For each pair of nonadjacent nodes in this graph, find a set of variables that d-separates that pair. What does this list tell us about independencies in the data?*

(b) *Repeat question (a) assuming that only variables in the set $\{Z_3, W, X, Z_1\}$ can be measured.*

(c) *For each pair of nonadjacent nodes in the graph, determined whether they are independent conditional on all other variables.*

(d) *For every variable V in the graph find a minimal set of nodes that renders V independent of all other variables in the graph.*

(e) *Suppose we wish to estimate the value of Y from measurements taken on all other variables in the model. Find the smallest set of variables that would yield as good an estimate of Y as when we measured all variables.*

(f) *Repeat question (e) assuming that we wish to estimate the value of Z_2.*

(g) *Suppose we wish to predict the value of Z_2 from measurements of Z_3. Would the quality of our prediction improve if we add measurement of W? Explain.*

Solution to study question 2.4.1

An *R* solution of this exercise is provided at dagitty.net/primer/2.4.1

Part (a)

Recall that two variables are *d*-separated if *all* simple paths between them are blocked by the given set of variables (see Definition 2.4.1, and Rules 1, 2, and 3 for the graphical criteria for conditional independence). Below, we list all pairs of variables that are *d*-separated, along with the paths that must be blocked between them.
(X, Y) are independent conditioned on set $\{Z_1, Z_3, W\}$:

1. $X \to W \to Y$ is blocked at chain $X \to W \to Y$ (W is given).
2. $X \leftarrow Z_3 \to Y$ is blocked at fork $X \leftarrow Z_3 \to Y$ (W is given).
3. $X \leftarrow Z_3 \leftarrow Z_2 \to Y$ is blocked at fork $Z_3 \leftarrow Z_2 \to Y$ (Z_2 is given).
4. $X \leftarrow Z_1 \leftarrow Z_3 \to Y$ is blocked at chain $Z_1 \to Z_3 \to Y$ (Z_3 is given).
5. $X \leftarrow Z_1 \to Z_3 \leftarrow Z_2 \to Y$ is blocked at fork $X \leftarrow Z_1 \to Z_3$ (Z_1 is given).

(X, Z_2) are independent conditioned on set $\{Z_1, Z_3\}$:

1. $X \leftarrow Z_1 \to Z_3 \leftarrow Z_2$ is blocked at fork $X \leftarrow Z_1 \to Z_3$ (Z_1 is given).
2. $X \leftarrow Z_1 \to Z_3 \to Y \leftarrow Z_2$ is blocked at fork $X \leftarrow Z_1 \to Z_3$ (Z_1 is given).
3. $X \leftarrow Z_3 \leftarrow Z_2$ is blocked at chain $X \leftarrow Z_3 \leftarrow Z_2$ (Z_3 is given).
4. $X \leftarrow Z_3 \to Y \leftarrow Z_2$ is blocked at fork $X \leftarrow Z_3 \to Y$ (Z_3 is given).
5. $X \to W \to Y \leftarrow Z_2$ is blocked at collider $W \to Y \leftarrow Z_2$ (Y is not given).

6. $X \to W \to Y \leftarrow Z_3 \leftarrow Z_2$ is blocked at collider $W \to Y \leftarrow Z_3$ (Y is not given).

Viewing the above *d*-separation path analysis, we can similarly deduce the remaining independent pairs.

- (Y, Z_1) are independent conditioned on set $\{X, Z_2, Z_3\}$.
- (W, Z_1) are independent conditioned on set $\{X\}$.
- (W, Z_2) are independent conditioned on set $\{X\}$.
- (W, Z_3) are independent conditioned on set $\{X\}$.
- (Z_1, Z_2) are independent conditioned on set $\{\}$.

What does this list of independencies tell us about those in the data? Assuming that the data was generated from this model, then the independencies will also be manifest in the data.

Part (b)

Using the path analyses we performed in (a) above, we can determine if each pair of variables can be *d*-separated conditioning only on variables in the set $\{Z_3, W, X, Z_1\}$.

- (X, Y) are independent conditioned on set $\{Z_1, Z_3, W\}$.
- (X, Z_2) are independent conditioned on set $\{Z_1, Z_3\}$.
- (W, Z_1) are independent conditioned on set $\{X\}$.
- (W, Z_2) are independent conditioned on set $\{X\}$.
- (W, Z_3) are independent conditioned on set $\{X\}$.
- (Z_1, Z_2) are independent conditioned on set $\{\}$.

(Y, Z_1) were independent when we could condition on $\{X, Z_2, Z_3\}$, but now there is no such set that will render (Y, Z_1) independent. Since we can no longer condition on Z_2, we must address two paths:

1. $Z_1 \to Z_3 \leftarrow Z_2 \to Y$ is blocked if we do not condition on Z_3, but open if we do.
2. $Z_1 \to Z_3 \to Y$ is blocked if we condition on Z_3, but open if we do not.

Observe that these two requirements are mutually exclusive, so one of these two paths remains open, and so (Y, Z_1) cannot be *d*-separated using the covariates available.

Part (c)

Again using our path analysis from (a), we have:

- (X, Y) are independent given $\{Z_1, Z_2, Z_3, W\}$.
- (X, Z_2) are independent given $\{Z_1, Z_3, W, Y\}$.
- (Y, Z_1) are independent given $\{Z_2, Z_3, X, W\}$.
- (W, Z_1) are independent given $\{Z_2, Z_3, X, Y\}$.

- (W, Z_2) are dependent given $\{Z_1, Z_3, X, Y\}$; there is an open path $W \to Y \leftarrow Z_2$ ($W \to Y \leftarrow Z_2$ is a collider and Y is given).
- (W, Z_3) are dependent given $\{Z_1, Z_2, X, Y\}$: there is an open path $W \to Y \leftarrow Z_3$ ($W \to Y \leftarrow Z_3$ is a collider and Y is given).
- (Z_1, Z_2) are dependent given $\{Z_3, X, W, Y\}$: there is an open path $Z_1 \to Z_3 \leftarrow Z_2$ ($Z_1 \to Z_3 \leftarrow Z_2$ is a collider and Z_3 is given).

Part (d)

Consider the set of variables **Z** comprised of the parents, children, and spouses of V: **Z** $= \{parents(V),\ children(V),\ spouses(V)\}$.
Let us first convince ourselves that conditioning on **Z** is guanteed to render V independent of all other variables in the graph.

1. $parents(V)$: conditioning on the parents of V will block any forks and chains incumbent to V.
2. $children(V)$: conditioning on the children of V will block any forks and chains emanating from V.
3. $spouses(V)$: conditioning on the spouses of V will block paths that were opened at a collider formed on a child of V (which was opened when we conditioned on $children(V)$).

Conditioning on **Z** is guaranteed to *d*-separate V from all other nodes in the model (which is referred to as the Markov Blanket for a node V, sometimes denoted $MB(V)$).

$$MB(X) = \{W, Z_1, Z_3\}$$
$$MB(Y) = \{Z_2, Z_3, W\}$$
$$MB(W) = \{X, Y, Z_2, Z_3\}$$
$$MB(Z_1) = \{X, Z_2, Z_3\}$$
$$MB(Z_2) = \{Z_1, Z_3, Y, W\}$$
$$MB(Z_3) = \{Z_1, Z_3, X, W, Y\}$$

Part (e)

The minimal set would be the Markov Blanket of Y, $MB(Y) = \{Z_2, Z_3, W\}$, since $\{X, Z_1\}$ are independent from Y given $MB(Y)$, and so their addition would not improve our estimate.

Part (f)

The minimal set would be the Markov Blanket of Z_2, $MB(Z_2) = \{Z_1, Z_3, Y, W\}$, since $\{X\}$ is independent from Z_2 given $MB(Z_2)$. Observe that here we include Y since it improves our estimate of Z_2 (they are dependent), but this inclusion opens a path that was not open when we conditioned on all variables: $Y \leftarrow W \leftarrow X \leftarrow Z_1 \to Z_3 \leftarrow Z_2$. We then block this path by conditioning on W.

Part (g)

Yes, since Z_2 and W are dependent given Z_3 (there is an open path $W \leftarrow X \leftarrow Z_1 \rightarrow Z_3 \leftarrow Z_2$ such that information about W provides information about Z_2).

Study question 2.5.1.

(a) Which of the arrows in Figure 2.9 can be reversed without being detected by any statistical test? [Hint: Use the criterion for equivalence class.]

(b) List all graphs that are observationally equivalent to the one in Figure 2.9.

(c) List the arrows in Figure 2.9 whose directionality can be determined from non-experimental data.

(d) Write down a regression equation for Y such that, if a certain coefficient in that equation is non-zero the model of Figure 2.9 is wrong.

(e) Repeat question (d) for variable Z_3.

(f) Repeat question (e) assuming the X is not measured.

(g) How many regression equations of the type described in (d) and (e) are needed to ensure that the model is fully tested, namely, that if it passes all these tests it cannot be refuted additional tests of these kind. [Hint: Ensure that you test every vanishing partial regression coefficient that is implied by the product decomposition (1.29).]

Solution to study question 2.5.1

An *R* solution of this exercise is provided at dagitty.net/primer/2.5.1

Part (a)

To determine which arrows can be reversed without being detectable by a statistical test, we consider models that are in the *equivalence class* of Figure 2.9 (see paragraphs previous to these study questions in the text). Accordingly, we first find the *v*-structures in the graph (i.e., colliders whose parents are not adjacent), which are:

- $Z_1 \rightarrow Z_3 \leftarrow Z_2$
- $Z_3 \rightarrow Y \leftarrow W$
- $Z_2 \rightarrow Y \leftarrow W$

So, to find other models in this equivalence class, we may flip the direction of edges such that the resulting model abides by two criteria:

- We neither create nor destroy any *v*-structures.
- We must not introduce a cycle into the resulting graph. Note that while *d*-separation is valid in linear cyclical models, it is not valid in general, namely, for any non-linear functions.

With these constraints, we conclude that there are *no such edges* that can be reversed within the model to find another within its equivalence class. We would verify this claim by testing our constraints against each edge reversal. Here is a complete list of our tests and a reason (of which there might be several) why each reversal fails to produce a model in the equivalence class of Figure 2.9 (note, we define the "old model" as the one pre-edge-reversal, and the "new model" as the one resulting from the reversal):

Edge	**Reason**
$Z_1 \to X$	Creates a cycle in new model, $Z_1 \to Z_3 \to X \to Z_1$
$Z_1 \to Z_3$	Destroys a *v*-structure in old model, $Z_1 \to Z_3 \leftarrow Z_2$
$Z_3 \to X$	Creates a *v*-structure in new model, $X \to Z_3 \leftarrow Z_2$
$Z_3 \to Y$	Creates a *v*-structure in new model, $Y \to Z_3 \leftarrow Z_1$
$X \to W$	Creates a *v*-structure in new model, $W \to X \leftarrow Z_1$
$W \to Y$	Destroys a *v*-structure in old model, $W \to Y \leftarrow Z_2$
$Z_2 \to Z_3$	Destroys a *v*-structure in old model, $Z_1 \to Z_3 \leftarrow Z_2$
$Z_2 \to Y$	Destroys a *v*-structure in old model, $Z_2 \to Y \leftarrow W$

Part (b)

There are no additional models in the equivalence class of Figure 2.9 for the reasons stated in part (a) above.

Part (c)

No edge may be reversed to produce a model in the equivalence class of Figure 2.9 (see explanation in part (a)). Therefore, all edge directionalities in the graph may be determined from non-experimental data.

Part (d)

The model in 2.4.1(d) implies that Y is independent of Z_1 given $\{Z_2, Z_3, W\}$.
So, suppose we fit the data with the equation:

$$y = r_2 z_2 + r_3 z_3 + r_w w + r_1 z_1$$

If r_1 is non-zero in the fitted equation, then the model of Figure 2.9 is wrong since the data violates the conditional independence between Y and Z_1 as claimed by the model.

Part (e)

The model in 2.4.1(d) implies that Z_3 is independent of W given $\{X\}$.
So, suppose we fit the data with the equation:

$$z_3 = r_x x + r_w w$$

If r_w is non-zero in the fitted equation, then the model of Figure 2.9 is wrong since the data violates the conditional independence between Z_3 and W as claimed by the model.

Part (f)

No such regression exists because no variable can be separated from Z_3 by any set of observed variables.

Part (g)

According to Equation 1.29, the joint probability distribution can be factorized as:
$P(Z_1, Z_2, Z_3, X, W, Y) = P(Z_1)P(Z_2)P(Z_3|Z_1, Z_2)P(X|Z_1, Z_3)P(W|X)P(Y|W, Z_2, Z_3)$
So, to fully test the model, we need to examine every factor in this factorization and establish that, conditional on its parents, every variable V is independent of its non-descendants. The corresponding regression equations necessary to perform these tests are:

1. $z_2 = r_1 z_1$ with vanishing r_1.
2. $x = r_1 z_1 + r_2 z_2 + r_3 z_3$ with vanishing r_2.
3. $w = r_x x + r_1 z_1 + r_2 z_2 + r_3 z_3$ with vanishing r_1, r_2, r_3.
4. $y = r_w w + r_x x + r_1 z_1 + r_2 z_2 + r_3 z_3$ with vanishing r_x, r_1.

So, in total, to fully test the model we would need 4 regression equations, through which we would perform $1 + 1 + 3 + 2 = 7$ tests for vanishing regression coefficients.

Study Questions and Solutions for Chapter 3

Study question 3.2.1.
Referring to Study question 1.5.2 (Figure 1.8) and the parameters listed therein,

(a) *Compute $P(y|\text{do}(x))$ for all values of x and y, by simulating the intervention $\text{do}(x)$ on the model.*

(b) *Compute $P(y|\text{do}(x))$ for all values of x and y, using the adjustment formula (3.5).*

(c) *Compute the ACE*

$$ACE = P(y_1|do(x_1)) - P(y_1|do(x_0))$$

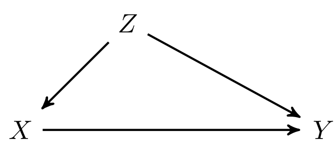
and compare it to the Risk Difference

$$RD = P(y_1|x_1) - P(y_1|x_0).$$

(d) *Find a combination of parameters that exhibit Simpson's reversal (as in Study question 1.5.2(c)) and show explicitly that the overall causal effect of the drug is obtained from the aggregate data.*

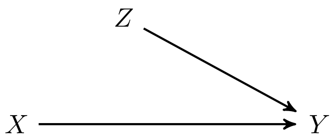
Solution to study question 3.2.1

First, let's save the graph for ease of reference:



Part (a)

Now, to compute $P(y|do(x))$ for all values of x and y, we can consider the mutilated model m wherein all causal influences to X are severed, and X is forced to some value x:



With this mutilated model in hand, we can write the product decomposition, Eq. (1.29), for m to solve for our quantities of interest. Observe three facts that will help us with our task: $P(Y|Z,X) = P_m(Y|Z,X)$, $P(Z) = P_m(Z)$, and $P_m(Z|X) = P_m(Z) = P(Z)$.

$$\begin{aligned}
P(Y = y|do(X = x)) &= P_m(Y = y|X = x) \\
&= \sum_z P_m(Y = y, Z = z|X = x) \\
&= \sum_z P_m(Y = y|Z = z, X = x)P_m(Z = z|X = x) \\
&= \sum_z P_m(Y = y|Z = z, X = x)P_m(Z = z) \\
&= \sum_z P(Y = y|Z = z, X = x)P(Z = z)
\end{aligned}$$

$$\begin{aligned}
P(y_1|do(x_1)) &= P(y_1|x_1, z_1)P(z_1) + P(y_1|x_1, z_0)P(z_0) = rp_4 + (1 - r)p_2 \\
P(y_1|do(x_0)) &= P(y_1|x_0, z_1)P(z_1) + P(y_1|x_0, z_0)P(z_0) = rp_3 + (1 - r)p_1 \\
P(y_0|do(x_1)) &= 1 - P(y_1|do(x_1)) = 1 - (rp_4 + (1 - r)p_2) \\
P(y_0|do(x_0)) &= 1 - P(y_1|do(x_0)) = 1 - (rp_3 + (1 - r)p_1)
\end{aligned}$$

Part (b)

By the adjustment formula, Eq. (3.6), we have the same as in (a):

$$\begin{aligned}
P(y_1|do(x_1)) &= P(y_1|x_1, z_1)P(z_1) + P(y_1|x_1, z_0)P(z_0) = rp_4 + (1 - r)p_2 \\
P(y_1|do(x_0)) &= P(y_1|x_0, z_1)P(z_1) + P(y_1|x_0, z_0)P(z_0) = rp_3 + (1 - r)p_1 \\
P(y_0|do(x_1)) &= 1 - (rp_4 + (1 - r)p_2) \\
P(y_0|do(x_0)) &= 1 - (rp_3 + (1 - r)p_1)
\end{aligned}$$

Part (c)

To find the ACE, we simply substitute our computations from (a) into the ACE formula, giving us:

$$\begin{aligned} ACE &= P(y_1|do(x_1)) - P(y_1|do(x_0)) \\ &= rp_4 + (1-r)p_2 - rp_3 - (1-r)p_1 \end{aligned}$$

To compute RD, we use study question 1.5.2(d), and obtain:

$$\begin{aligned} RD &= P(y_1|x_1) - P(y_1|x_0) \\ &= \frac{[rp_4q_2 + (1-r)p_2q_1]}{[rq_2 + (1-r)q_1]} - \frac{[rp_3(1-q_2) + (1-r)p_1(1-q_1)]}{[r(1-q_2) + (1-r)(1-q_1)]} \end{aligned}$$

We see that, in general, the $ACE \neq RD$: the Average Causal Effect (ACE) measures the effect on Y from *intervening* and forcing X to change from x_0 to x_1. In contrast, the Risk Difference (RD) measures the effect on Y from *observing* change from x_0 to x_1.

Comparing the expressions for ACE and RD, we see that when $r = 0, q_1 \neq 0, q_1 \neq 1$, the ACE is equivalent to RD, namely:

$$ACE - RD = p_2 - p_1 - p_2 + p_1 = 0$$

Part (d)

Using our answer to study question 1.5.2(c), we note that the following parameterization will yield Simpson's reversal:

$$p_1 = 0.1, p_2 = 0, p_3 = 0.3, p_4 = 0.2, q_1 = 0, q_2 = 1, r = 0.1$$

Recall that this parameterization yields Simpson's reversal because, for each value of z, the difference $P(y|x_1, z) - P(y|x_0, z)$ has the same sign as the ACE. We also know that, because Z is the only confounder, to determine the overall causal effect of the drug on recovery, we're interested in its average influence across all Z-specific conditions. To aid us in this analysis, we can use the ACE from part (c) above, and consult the segregated data, namely:

$$\begin{aligned} ACE &= P(y_1|do(x_1)) - P(y_1|do(x_0)) \\ &= rp_4 + (1-r)p_2 - rp_3 - (1-r)p_1 \\ &= 0.02 + 0 - 0.03 - 0.09 \\ &= -0.1 \end{aligned}$$

Study question 3.3.1.
Consider the graph in Figure 3.8:

(a) List all of the sets of variables that satisfy the backdoor criterion to determine the causal effect of X on Y.

(b) List all of the minimal sets of variables that satisfy the backdoor criterion to determine the causal effect of X on Y (i.e., any set of variables such that, if you removed any one of the variables from the set, it would no longer meet the criterion).

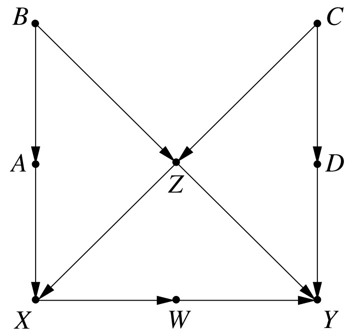


Figure 3.8: Causal graph used to illustrate the backdoor criterion in the following study questions

(c) List all minimal sets of variables that need be measured in order to identify the effect of D on Y. Repeat, for the effect of $\{W, D\}$ on Y.

Solution to study question 3.3.1

The graph of this exercise is available at dagitty.net/m331. Students can interactively modify this graph and see the implications for the backdoor adjustment sets.

An *R* solution of this exercise is provided at dagitty.net/primer/3.3.1

Part (a)

Let us use the abbreviation "backdoor admissible" to denote a set of variables that satisfy the backdoor criterion of Definition 3.3.1; for the present model, a backdoor admissible set ***Z*** blocks all spurious paths between X and Y while leaving all directed paths from X to Y open. We can easily verify that the following sets satisfy the backdoor criterion to determine the causal effect of X on Y.

1. *Sets of 2 nodes:* $\{Z, A\}, \{Z, B\}, \{Z, C\}, \{Z, D\}$
2. *Sets of 3 nodes:* $\{Z, A, B\}, \{Z, A, C\}, \{Z, A, D\}, \{Z, B, C\}, \{Z, B, D\}, \{Z, C, D\}$
3. *Sets of 4 nodes:* $\{Z, A, B, C\}, \{Z, A, B, D\}, \{Z, A, C, D\}, \{Z, B, C, D\}$
4. *Sets of 5 nodes:* $\{Z, A, B, C, D\}$

Part (b)

According to (a), the following 4 sets are minimal, since in every other set, a node could be removed and still ensure that the backdoor criterion is satisfied:
$\{Z, A\}, \{Z, B\}, \{Z, C\}, \{Z, D\}$.

Part (c)

For identifying the effect of D on Y:
We want to find a set **Z** that blocks all backdoor paths from D to Y. Notice that the set $\{C\}$ is one solution, so any other set that contains C is not minimal. Also, if the set does not contain C, then it must contain Z, otherwise, we have an open path $Y \leftarrow Z \leftarrow C \rightarrow D$. However, by including Z, the backdoor path $Y \leftarrow W \leftarrow X \leftarrow A \leftarrow B \rightarrow Z \leftarrow C \rightarrow D$ is open. To block this path, we add any of A, B, X, or W. So, there are a total of 5 minimal sets:
$\{C\}, \{Z, A\}, \{Z, B\}, \{Z, X\}, \{Z, W\}$.

For identifying the effect of $\{W, D\}$ on Y:
Again, we want to verify that the only open paths from $\{W, D\}$ to Y are the direct edges. So similar to above, if the set **Z** contains Z, then all backdoor paths from D to Y or from W to Y will be blocked. Also if the set **Z** does not contain Z, then it must contain C and X, otherwise, we have an open path through $D \leftarrow C \rightarrow Z \rightarrow Y$ or $W \leftarrow X \leftarrow Z \rightarrow Y$. So there are a total of 2 minimal sets as follows:
$\{Z\}, \{C, X\}$.

Study question 3.3.2. *(Lord's Paradox).*
At the beginning of the year, a boarding school offers its students a choice between two meal plans for the year: Plan A and Plan B. The students' weights are recorded at the beginning and the end of the year. To determine how each plan affects students' weight gain, the school hired two statisticians who, oddly, reached different conclusions. The first statistician calculated the difference between each student's weight in June (W_F) and in September (W_I) and found that the average weight gain in each plan was zero.

The second statistician divided the students into several subgroups, one for each initial weight, W_I. He finds that for each initial weight, the final weight for Plan B is higher than the final weight for Plan A.

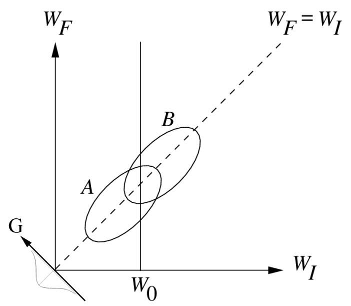


Figure 3.9: Scatter plot with students' initial weights on the x-axis and final weights on the y-axis. The vertical line indicates students whose initial weights are the same, and whose final weights are higher (on average) for plan B compared with plan A

So, the first statistician concluded that there was no effect of diet on weight gain and the second concluded there was.

Figure 3.9 illustrates data sets that can cause the two statisticians to reach conflicting conclusions. Statistician-1 examined the weight gain $W_F - W_I$ which, for each student, is represented by the shortest distance to the 45° line. Indeed, the average gain for each diet plan is zero; the two groups are each situated symmetrically relative to the zero-gain line, $W_F = W_I$. Statistician-2, on the other hand, compared the final weights of plan A students to those of plan-B students who entered school with the same initial weight W_0 and, as the vertical line in the figure indicates, plan B students are situated above plan A students along this vertical line. The same will be the case for any other vertical line, regardless of W_0.

(a) Draw a causal graph representing the situation.

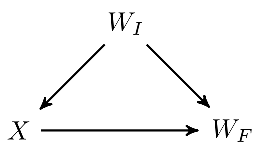
(b) Determine which statistician is correct.

(c) How is this example related to Simpson's paradox?

Solution to study question 3.3.2

Part (a)

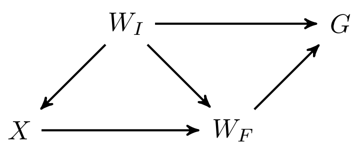
First, let us configure the variables we'll use in our model. Let X be the students' meal plan choice, W_F be the students' final weights, and W_I be the students' initial weights. We hypothesize that a student's initial weight influences both their choice of meal plan and their final weight. Additionally, meal plan influences final weight. So, the causal graph is as follows:



Part (b)

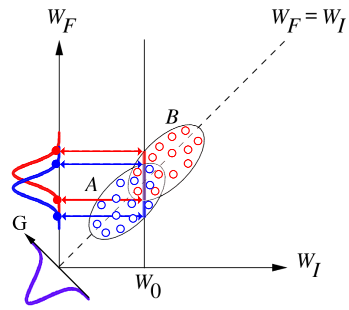
The 2nd statistician is correct, since the initial weight W_I is the common cause of plan choice X and final weight W_F. As such, when we estimate the effect of X on W_F, we should segregate data on the initial weight. Also, W_I satisfies the backdoor criterion to determine the causal effect of X on W_F. The first statistician mistakenly used the aggregated data, failing to account for the confounder W_I.

Statistician 1's argument sounds compelling only because it is expressed in terms of the gain $G = W_F - W_I$, which people perceive to be the quantity of interest, and which Figure 3.9 clearly shows to have the same mean in Diet A as in Diet B. However, once we add G to the graph (see below) the error in this argument becomes clear: to compute the effect of X on G we still need to adjust for W_I.



Part (c)

Comparing the model in Part (a) to the standard model in Simpson's paradox (e.g., Fig. 1.8) we see that the structures of the two models are the same, with a slightly different causal story. The difference is that, in Simpson's paradox, we have complete reversal while in Lord's paradox, we are going from inequality to equality. To visualize this transition, we can examine the W_I-specific distributions of W_F for each of the diets, and ask whether the two distributions differ. This we do by projecting samples corresponding to an initial weight W_0 on to the W_F axis, as shown in the graph below.



We see that for individuals having the same initial weight, W_0, their final weight will be higher in Plan B than in Plan A (on the average). The distributions corresponding to the two scatter plots are shifted. On the other hand, if we project the two scatter plots onto the G axis, the two distributions coincide. Thus, the segregated data (conditioned on W_I) yields preference of one diet over the other, while the unsegregated data (unconditioned on W_I) claims equality for the two diets.

In Simpson's paradox, on the other hand, we encounter sign reversal as we go from the segregated to the unsegregated data. This is shown, for example, in the age-specific slopes of Figure 1.1 in the text, which have opposite sign to the slope of the aggregated ellipse.

Study question 3.3.3.

Revisit the lollipop story of Study question 1.2.4 and answer the following questions:

(a) Draw a graph that captures the story.

(b) Determine which variables must be adjusted for by applying the backdoor criterion. variables need to be adjusted for.

(c) Write the adjustment formula for the effect of the drug on recovery.

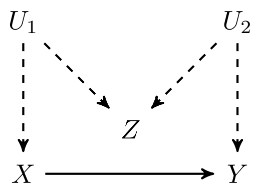
(d) Repeat questions (a)–(c) assuming that the nurse gave lollipops a day after the study, still preferring patients who received treatment over those who received placebo.

Solution to study question 3.3.3

Students who wish to study Simpson's paradox in more detail can use the interactive simulator of the "Simpson Machine" at dagitty.net/learn/simpson/

Part (a)

As with any modeling problem, we begin by formalizing our variable choices. Let X indicate Treatment receipt, Z indicate Lollipop receipt, Y indicate Recovery, and U_1, U_2 indicate 2 unobserved factors that correlate Z with X and Y, respectively. The causal graph will be:



Part (b)

By Definition 3.3.1, the backdoor criterion, to estimate the effect of X on Y, we need not adjust for any variable, since $U_1 \to Z \leftarrow U_2$ is a collider that is closed when Z is *not* given. As we discussed in the solution to study question 1.2.4, the structure of the graph permits us to skip considerations of exchangeability (i.e., comparing apples and oranges), and get the answer mechanically and reliably.

Part (c)

According to (b), since we need not adjust for any covariates to block any spurious paths, we may simply say that: $P(y|do(x)) = P(y|x)$

Part (d)

Our answers do not change; timing of the Lollipop receipt does not change the causal structure of the model, as long as receiving a Lollipop is assumed to have no effect on either treatment or outcome. In other words, Z is not an ancestor of either X or Y.

Study question 3.4.1.
Assume that in Figure 3.8, only X, Y, and one additional variable can be measured. Which variable would allow the identification of the effect of X on Y? What would that effect be?

Solution to study question 3.4.1

Variable W satisfies the Front-Door criterion in Definition 3.4.1, so W would allow the identification of the effect of X on Y, namely:

$$P(y|do(x)) = \Sigma_w P(w|x) \Sigma_{x'} P(y|x', w) P(x')$$

Study question 3.4.2.

I went to a pharmacy to buy a certain drug, and I found that it was available in two different bottles: one priced at $1, the other at $10. I asked the druggist, "What's the difference?" and he told me, "The $10 bottle is fresh, whereas the $1 bottle one has been on the shelf for 3 years. But, you know, data shows that the percentage of recovery is much higher among those who bought the cheap stuff. Amazing isn't it?" I asked if the aged drug was ever tested. He said, "Yes, and this is even more amazing; 95% of the aged drug and only 5% of the fresh drug has lost the active ingredient, yet the percentage of recovery among those who got bad bottles, with none of the active ingredient, is still much higher than among those who got good bottles, with the active ingredient."

Before ordering a cheap bottle, it occurred to me to have a good look at the data. The data were, for each previous customer, the type of bottle purchased (aged or fresh), the concentration of the active ingredient in the bottle (high or low), and whether the customer recovered from the illness. The data perfectly confirmed the druggist's story. However, after making some additional calculations, I decided to buy the expensive bottle after all; even without testing its content, I could determine that a fresh bottle would offer the average patient a greater chance of recovery.

Based on two very reasonable assumptions, the data show clearly that the fresh drug is more effective. The assumptions are as follows:

- *(i) Customers had no information about the chemical content (high or low) of the specific bottle of the drug that they were buying; their choices were influenced by price and shelf-age alone.*
- *(ii) The effect of the drug on any given individual depends only on its chemical content, not on its shelf age (fresh or aged).*

(a) Describe this scenario in a causal graph.

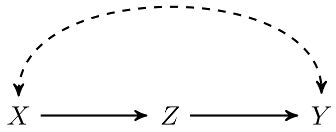
(b) Construct a data set compatible with the story and the decision to buy the expensive bottle.

(c) Determine the effect of choosing the fresh versus the aged drug by using assumptions (i) and (ii) and the data given in (b).

Solution to study question 3.4.2

Part (i)

Let X denote Drug Price (the cheap / old vs. expensive / fresh), Z indicate chemical Potency of the drug, and Y indicate Recovery. We hypothesize that the age of the Drug affects its chemical Potency, which in turn affects Recovery. We also hypothesize the existence of an unobserved confounder that makes people influenced by Price to have more health problems. The graph is as follows:



Part (ii)

(Hint: the data is the same as in table 3.2 in the book, and they share the same causal relations and corresponding graphical model)

	Cheap	Expensive	All Subjects
	400	400	800
	Low | High	Low | High	Low | High
	380 | 20	20 | 380	400 | 400
Recovery	323 | 18	1 | 38	324 | 56
Not Recovery	57 | 2	19 | 324	76 | 344

Part (iii)

Using (ii) above, our knowledge that Z satisfies the Front-Door criterion, and Theorem 3.4.1, we can compare the causal effects of the two drug types to see which is superior. Let y_1 denote recovery, x_1 denote choosing the expensive drug, and z_1 denote high chemical content.

$$\begin{aligned}
P(y_1|do(x_1)) &= \Sigma_z P(z|x_1)\Sigma_{x'} P(y_1|x',z)P(x') \\
&= P(z_1|x_1)[P(y_1|x_1,z_1)P(x_1) + P(y_1|x_0,z_1)P(x_0)] \\
&\quad + P(z_0|x_1)[P(y_1|x_1,z_0)P(x_1) + P(y_1|x_0,z_0)P(x_0)] \\
&= 0.95 * (0.1 * 0.5 + 0.9 * 0.5) + 0.05 * (0.05 * 0.5 + 0.85 * 0.5) \\
&= 0.4975 \\
P(y_1|do(x_0)) &= \Sigma_z P(z|x_0)\Sigma_{x'} P(y_1|x',z)P(x') \\
&= P(z_1|x_0)[P(y_1|x_1,z_1)P(x_1) + P(y_1|x_0,z_1)P(x_0)] \\
&\quad + P(z_0|x_0)[P(y_1|x_1,z_0)P(x_1) + P(y_1|x_0,z_0)P(x_0)] \\
&= 0.05 * (0.1 * 0.5 + 0.9 * 0.5) + 0.95 * (0.05 * 0.5 + 0.85 * 0.5) \\
&= 0.4525
\end{aligned}$$

$$\begin{aligned}
ACE &= P(y_1|do(x_1)) - P(y_1|do(x_0)) \\
&= 0.045 > 0
\end{aligned}$$

So, we see that recovery is indeed more likely when using the expensive drug than the cheap one.

We can rationalize these findings by first remembering that our data is observational, and that an unobserved confounder between drug choice and recovery might create the illusion that the cheap drug is more effective. This would happen, for instance, if the more frugal customers are also on a more healthy diet. To eliminate such illusions we must evaluate the ACE which, given our story, is obtained from the front door formula.

We can also foster an intuition for how the front-door formula is useful. First, we can compute the causal effect of the drug's active ingredients on recovery. Let p (respectively, p') represent the recovery probability of a randomly chosen person forced to take a drug high (respectively, low) in the active ingredients, i.e.,

$$p = P(Y = recover|do(Z = high))$$
$$p' = P(Y = recover|do(Z = low))$$

To counteract possible confounding between recovery and active ingredient potency, we adjust for price. We find the recovery probability for those who bought cheap bottles, then for those who bought expensive bottles, and we take a weighted average of the two based on the relative size of the two groups, doing the same with those given a bottle with low potency. Taking the difference gives us the causal effect difference, $p - p'$, averaged over the entire population.

Now that we know the causal effect of the ingredient, we reason as follows: If I choose a cheap bottle, I stand a 5% chance of getting a good bottle, with recovery probability p, and 95% chance of getting a bad bottle with a recovery probability of p'. Thus, the average probability of recovery on choosing a cheap bottle is $.05p + .95p'$. Things turn around if I buy an expensive bottle, giving me an average probability of recovery of $.05p' + .95p$. Thus the difference between the two choices amounts to:

$$.05(p - p') + .95(p' - p) = .9(p' - p)$$

So, if $p > p'$, it is clearly advantageous to buy the expensive bottle.

Study question 3.5.1.
Consider the causal model of Figure 3.8.

(a) *Find an expression for the c-specific effect of X on Y.*

(b) *Identify a set of four variables that need to be measured in order to estimate the z-specific effect of X on Y, and find an expression for the size of that effect.*

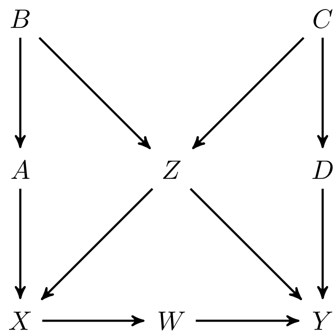
(c) *Using your answer to part (b), determine the expected value of Y under a Z-dependent strategy where X is set to 0 when Z is smaller or equal to 2 and X is set to 1 when Z is larger than 2. (Assume Z takes on integer values from 1 to 5.)*

Solution to study question 3.5.1

The graph of this exercise is available at dagitty.net/m331. Students can solve parts (a) and (b) interactively in class by forcing adjustment for single covariates (move mouse pointer over the variable and press "a" key).

An *R* solution of this exercise is provided at dagitty.net/primer/3.5.1

Repeating Figure 3.8 for ease of reference:



Part (a)

By Rule 2, we must adjust for a set of variables that satisfies the backdoor criterion, conditional on C. We observe that when we condition on C, there is still an open backdoor from $X \leftarrow Z \rightarrow Y$, which we can block by conditioning on Z. So, we may claim that:

$$P(Y = y|do(X = x), C = c) = \sum_z P(Y = y|X = x, Z = z, C = c)P(Z = z)$$

Above, Rule 2 is applicable because the set $\{Z, C\}$ satisfies the backdoor criterion to assess the c-specific effect of X on Y.

Part (b)

Again using Rule 2, we see that $\{X, Y, Z, C\}$ is such a set since $\{Z, C\}$ satisfies the backdoor criterion. We can then write:

$$P(Y = y|do(X = x), Z = z) = \sum_c P(y|x, z, c)P(c)$$

Advanced students may be challenged to show that $\{X, Y, Z, W\}$ is also such a set, since W satisfies the front-door criterion when Z is specified.

Part (c)

Since our choice of X relies upon the value of Z, we need to adopt the conditional policy $do(X = g(Z))$, where:

$$g(Z) = \begin{cases} 0 & Z \leq 2 \\ 1 & Z > 2 \end{cases}$$

We then assume that $Z \in \{1, 2, 3, 4, 5\}$, so by Eq. (3.17) we have:

$$\begin{aligned}
P(Y = y|do(X = g(Z))) &= \sum_z P(Y = y|do(X = g(z)), Z = z)P(Z = z) \\
&= P(Y = y|do(X = 0), Z = 1)P(Z = 1) \\
&\quad + P(Y = y|do(X = 0), Z = 2)P(Z = 2) \\
&\quad + P(Y = y|do(X = 1), Z = 3)P(Z = 3) \\
&\quad + P(Y = y|do(X = 1), Z = 4)P(Z = 4) \\
&\quad + P(Y = y|do(X = 1), Z = 5)P(Z = 5)
\end{aligned}$$

So, for each term above in the format $P(Y = y|do(X = x), Z = z)$, we can substitute our findings from (b) to find an expression free of the *do*-operator.

Study question 3.8.1.

Consider the causal model of Figure 3.10.

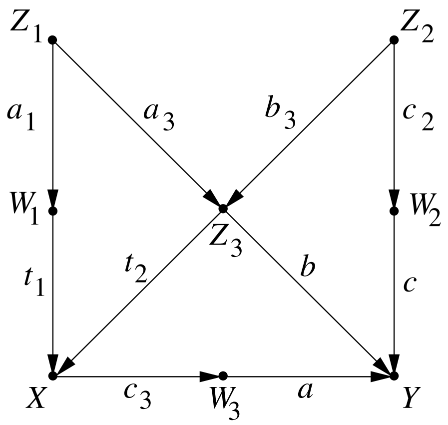


Figure 3.10

(a) *Identify three testable implications of this model.*

(b) *Identify a testable implication assuming that only X, Y, W_3, and Z_3 are observed.*

(c) *For each of the parameters in the model, write a regression equation in which one of the coefficients is equal to that parameter. Identify the parameters for which more than one such equation exists.*

(d) *Suppose X, Y and W_3 are the only variables observed. Which parameters can be identified from the data? Can the total effect of X on Y be estimated?*

(e) *If we regress Z_1 on all other variables in the model, which regression coefficient will be zero?*

(f) *The model in Figure 3.10 implies that certain regression coefficients will remain invariant when an additional variable is added as a regressor. Identify five such coefficients with their added regressors.*

(g) *Assume that variables Z_2 and W_2 cannot be measured. Find a way to estimate b using regression coefficients. [Hint: Find a way to turn Z_1 into an instrumental variable for b.]*

Solution to study question 3.8.1

Part (a)

Testable implications are conditional independence relationships implied by the structure of the graph. These conditional independences translate into vanishing regression coefficients in the data. Examining Figure 3.10, three regression equations that could be used to test the model could be:

1. W_3 is independent of W_1 given X, giving us the regression equation: $W_3 = r_X X + r_{W_1} W_1$ with $r_{W_1} = 0$. This means that if we fit the data to the line $W_3 = r_X X + r_{W_1} W_1$, we expect to find $r_{W_1} = 0$, or else the model is wrong.
2. W_1 is independent of Z_3 given Z_1, giving us the regression equation: $W_1 = r_{Z_1} Z_1 + r_{Z_3} Z_3$ with $r_{Z_3} = 0$
3. Y is independent of Z_1 given W_1, Z_2, and Z_3, giving us the regression equation: $Y = r_{Z_1} Z_1 + r_{W_1} W_1 + r_{Z_2} Z_2 + r_{Z_3} Z_3$ with $r_{Z_1} = 0$

Part (b)

The only conditional independence that involves the measured variables is the one between Z_3 and W_3 given X, which leads to $r_{Z_3} = 0$ in the corresponding regression equation:

$$W_3 = r_{Z_3} Z_3 + r_X X \text{ with } r_{Z_3} = 0$$

Part (c)

(i) If we regress a variable on its parents, we get a regression equation whose coefficients equal the model parameters. Therefore:

1. $a = r_{W_3}, b = r_{Z_3}, c = r_{W_2}$ in the equation:
$$Y = r_{W_3} W_3 + r_{Z_3} Z_3 + r_{W_2} W_2$$
2. $a_1 = r_{Z_1}$ in:
$$W_1 = r_{Z_1} Z_1$$
3. $a_3 = r_{Z_1}, b_3 = r_{Z_2}$ in:
$$Z_3 = r_{Z_1} Z_1 + r_{Z_2} Z_2$$
4. $c_2 = r_{Z_2}$ in:
$$W_2 = r_{Z_2} Z_2$$
5. $c_3 = r_X$ in:
$$W_3 = r_X X$$
6. $t_1 = r_{W_1}, t_2 = r_{Z_3}$:
$$X = r_{W_1} W_1 + r_{Z_3} Z_3$$

(ii) The "Regression Rule for Identification" tells us that, if G_α has several backdoor sets, each would lead to a regression equation in which α is a coefficient. Therefore, a, b, c can be identified by:

1. $a = r_{W_2}, b = r_{Z_3}, c = r_{W_2}$ in:

$$Y = r_{W_3} W_3 + r_{Z_3} Z_3 + r_{W_2} W_2$$

   Or, by $a = r_{W_2}, b = r_{Z_3}, c = r_{Z_2}$ in:

$$Y = r_{W_3} W_3 + r_{Z_3} Z_3 + r_{Z_2} Z_2$$

2. Likewise, t_1 can be identified either by $t_1 = r_{W_1}$ in:

$$X = r_{W_1} W_1 + r_{Z_2} Z_2$$

   Or, by $t_2 = r_{W_1}$ in:

$$X = r_{W_1} W_1 + r_{Z_2} Z_3$$

Part (d)

To determine which parameters are estimable from data, we consult "The Regression Rule for Identification." For example, the parameter c_3 can be estimated from data because $W_3 = r_X X + U_3' = c_3 X + U_3'$, since W_3 is d-separated from Y given X in G_{W_3}. Likewise, $a = r_{YW_3 \cdot X}$.

Lastly, we note that W_3 is a front-door admissible variable for attaining the total effect of X on Y, and so the effect is estimable. Indeed the total effect of X on Y is simply the product of $a * c_3$, which we identified above.

Part (e)

Regressing Z_1 on all other variables in the model gives:

$$Z_1 = r_{Z_2} Z_2 + r_{Z_3} Z_3 + r_X X + r_{w_1} W_1 + r_{W_2} W_2 + r_{W_3} W_3 + r_Y Y$$

By d-separation, we see that Z_1 is independent of $\{X, W_3, Y, W_2\}$ given W_1, Z_3, Z_2. Therefore, $r_X = 0, r_{W_2} = 0, r_{W_3} = 0, r_Y = 0$

Part (f)

In order for a coefficient to remain invariant under the addition of a new regressor, the dependent variable must be independent of the added regressor given all of the old regressors.

Thus, for example, if we regress W_1 on Z_3 and X, adding W_3 would keep all regression coefficients in tact, but adding Y or Z_2 would change them, because of the path: $Y \leftarrow W_2 \leftarrow Z_2 \rightarrow Z_3 \leftarrow Z_1 \rightarrow W_1$ is opened by conditioning on Z_3. If we regress W_1 on Z_1, then we can add Z_3, Z_2, or W_2 without changing the regression coefficient.

Part (g)

Note that if we condition on W_1, we turn Z_1 into an instrument relative to the effect τ of Z_3 on Y. Using this idea, we can write the regression of Y on Z_1 given W_1, as the product τa_3 where $\tau = t_2 c_3 a + b$. Since each of t_2, c_3, a can be separately identified (see Part a above), we can then solve for b. Formally, we have:

$$t_2 c_3 a + b = r_{Z_1} / r_{Z_1'}$$

Where r_{Z_1} and $r_{Z_1'}$ are the regression coefficients in the following equations:

$$Y = r_{Z_1} Z_1 + r_{W_1} W_1 + \epsilon$$
$$Z_3 = r_{Z_1'} Z_1 + \epsilon$$

Study Questions and Solutions for Chapter 4

Study question 4.3.1.

Consider the model in Figure 4.3 and assume that U_1 and U_2 are two independent Gaussian variables, each with zero mean and unit variance.

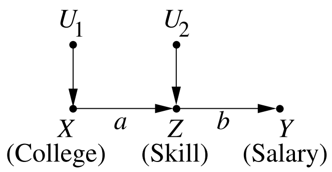


Figure 4.3: A model representing Eq. (4.7), illustrating the causal relations between college education (X), skills (Z), and salary (Y)

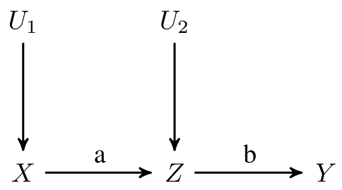
(a) Find the expected salary of workers at skill level $Z = z$ had they received x years of college education. [Hint: Use Theorem 4.3.2, with $e : Z = z$, and the fact that for any two Gaussian variables, say X and Z, we have $E[X|Z = z] = E[x] + R_{XZ}(z - E[Z])$. Use the material in Sections 3.8.2 and 3.8.3 to express all regression coefficients in terms of structural parameters, and show that $E[Y_x|Z = z] = abx + bz/(1 + a^2)$.]

(b) Based on the solution for (a), show that the skill-specific effect of education on salary is independent on the skill level.

Solution to study question 4.3.1

Part (a)

The quantity of interest will be $E[Y_x|Z = z]$ in the linear model:



Using the counterfactual formula of Theorem 4.3.2

$$E[Y_x|e] = E[Y|e] + \tau[x - E(X|e)]$$

we insert $e = \{Z = z\}$, and obtain

$$E[Y_x|Z = z] = E[Y|z] + \tau(x - E[X|z])$$

Assuming U_1 and U_2 are standardized, we have

$$\begin{aligned} E[X|z] &= \beta_{xz} z = \beta_{zx} \frac{\sigma_X^2}{\sigma_Z^2} z = a \frac{1}{cov(aX + U_2)} z \\ &= z \frac{a}{(1 + a^2)} \end{aligned}$$

which gives

$$\begin{aligned} E[Y_x|Z = z] &= bz + ab(x - \frac{za}{(1 + a^2)}) \\ &= abx + \frac{bz}{1 + \alpha^2} \end{aligned}$$

Part (b)

We want to show that $E[Y_1 - Y_0|Z = z] = E[Y_1 - Y_0]$. According to (a), we know that:

$$\begin{aligned} E[Y_1 - Y_0|Z = z] &= E[Y_1|Z = z] - E[Y_0|Z = z] \\ &= ab + \frac{bz}{1 + a^2} - 0 - \frac{bz}{1 + a^2} \\ &= ab \end{aligned}$$

Also, we know that $E[Y_x] = E[Y] + R_{YX}(x - E[X])$, so we may conclude:

$$\begin{aligned} E[Y_1 - Y_0] &= E[Y_1] - E[Y_0] \\ &= E[Y] + ab(1 - E[X]) - E[Y] - ab(0 - E[X]) \\ &= ab \\ &= E[Y_1 - Y_0 | Z = z] \end{aligned}$$

Study question 4.3.2.

(a) Describe how the parameters a, b, c of Model 2 (Figure 4.1) can be estimated from nonexperimental data.

(b) In the model of Figure 4.3, find the effect of education on those students whose salary is $Y = 1$. [Hint: use Theorem 4.3.2 to compute $E[Y_1 - Y_0 | Y = 1]$.]

(c) Estimate τ and the $ETT = E[Y_1 - Y_0 | X = 1]$ for the model described in Eq. (4.19) [Hint: Use the basic definition of counterfactuals, Eq. (4.5) and the equality $E[Z|X = x'] = R_{ZX}x'$.]

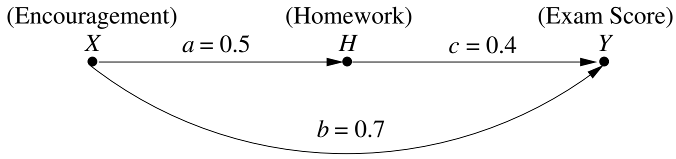


Figure 4.1: A model depicting the effect of Encouragement (X) on student's score

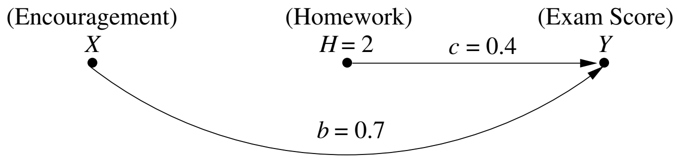


Figure 4.2: Answering a counterfactual question about a specific student's score, predicated on the assumption that homework would have increased to $H = 2$

Solution to study question 4.3.2

Part (a)

If the model is correct, then we can estimate the parameters from non-experimental data by simple regression. To estimate a, we regress H on X and compute the slope via Eq. (1.22):

$$H = \alpha + ax + \epsilon_h$$

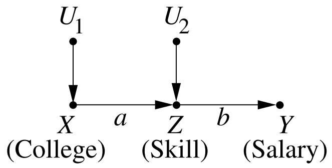


Figure 4.3: A model representing Eq. (4.7), illustrating the causal relations between college education (X), skills (Z), and salary (Y)

In other words, our estimate of a is the coefficient of x in the equation $H = \alpha + ax$ which fits the data best. Similarly (using Eqs. (1.27) and (1.28)), our estimates of b and c are the coefficients of x and h, respectively, in the regression equation $Y = \gamma + bx + ch$.

These "best fit" coefficients can be computed by efficient "least-square" algorithms.

Part (b)

By Theorem 4.3.2 we know that:

$$
\begin{aligned}
E[Y_1 - Y_0|Y=1] &= E[Y|Y=1] + \tau(1 - E[X|Y=1]) - E[Y|Y=1] - \tau(0 - E[X|Y=1]) \\
&= \tau \\
&= ab
\end{aligned}
$$

Part (c)

First, we define the model M as follows:

$$
\begin{aligned}
M &: \\
H &= U_H \\
X &= aH + U_X \\
Y &= bX + cH + \delta XH + U_Y
\end{aligned}
$$

From M, we may also describe the mutilated model M_x, representing the intervention $X = x$, as given by:

$$
\begin{aligned}
M_x &: \\
H &= U_H \\
X &= x \\
Y &= bx + cH + \delta xH + U_Y
\end{aligned}
$$

So by definition, to compute the counterfactual quantities needed for the ETT, we use Eq. (4.5) and write:

$$
\begin{aligned}
E[Y_x|X = x'] &= E[bx + cH + \delta xH + U_Y|X = x'] \\
&= bx + (c + \delta x)E[H|X = x'] \\
&= bx + (c + \delta x)\beta_{HX}x' \\
&= bx + (c + \delta x)\beta_{XH}\frac{\sigma_H^2}{\sigma_X^2}x' \\
&= bx + (c + \delta x)a\frac{1}{Var(aH + U_X)}x' \\
&= bx + (c + \delta x)\frac{ax'}{1 + a^2}
\end{aligned}
$$

Consequently, we can compute the ETT using Eq. (4.18) as:

$$
\begin{aligned}
ETT &= E[Y_1 - Y_0|X = 1] \\
&= E[Y_1|X = 1] - E[Y_0|X = 1] \\
&= b + (c + \delta)E[H|X = 1] - \frac{ca}{1 + a^2} \\
&= b + (c + \delta)\frac{a}{1 + a^2} - \frac{ca}{1 + a^2} \\
&= b + \frac{\delta a}{1 + a^2}
\end{aligned}
$$

Knowing also that:

$$
\begin{aligned}
\tau &= E[b + cH + \delta H + U_Y] - E[cH + U_Y] \\
&= b + \delta E[H] \\
&= b
\end{aligned}
$$

We may conclude:

$$
ETT - \tau = \frac{\delta a}{1 + a^2}
$$

Study question 4.3.3.

(a) Prove that, if X is binary, the effect of treatment on the treated can be estimated from both observational and experimental data. Hint: Decompose $E[Y_x]$ into

$$
E[Y_x] = E[Y_x|x']P(x') + E[Y_x|x]P(x)
$$

(b) Apply the result of Question (a) to Simpson's story with the nonexperimental data of Table 1.1, and estimate the effect of treatment on those who used the drug by choice. [Hint: Estimate $E[Y_x]$ assuming that gender is the only confounder.]

(c) Repeat Question (b) using the fact that Z in Figure 3.3, satisfies the backdoor criterion. Show that the answers to (b) and (c) coincide.

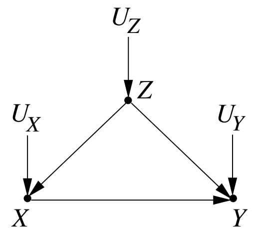


Figure 3.3: A graphical model representing the effects of a new drug, with Z representing gender, X standing for drug usage, and Y standing for recovery

Solution to study question 4.3.3

Part (a)

We begin by noting that, since X is binary, we can simply use the law of total probability and write:

$$E[Y_x] = E[Y_x|X = x]P(X = x) + E[Y_x|X = x']P(X = x')$$

By the *consistency* axiom, we also know that:

$$E[Y_x|X = x] = E[Y|X = x]$$
$$\therefore E[Y_x] = E[Y|X = x]P(X = x) + E[Y_x|X = x']P(X = x')$$

The consistency axiom intuitively follows from the notion that a counterfactual predicated on an actual observation is not counterfactual (here we observed $X = x$ and were hypothesizing about Y_x). The term $E[Y|X = x]P(X = x)$ is already estimable from observational data, so it remains to address the other. Solving for $E[Y_x|X = x']$, gives:

$$E[Y_x|X = x'] = \frac{E[Y_x] - E[Y|X = x]P(X = x)}{P(X = x')}$$

Now, substituting back into our equation for the ETT:

$$\begin{aligned} ETT &= E[Y_x - Y_{x'}|X = x] \\ &= E[Y|X = x] - E[Y_{x'}|X = x] \\ &= E[Y|X = x] - \frac{E[Y|do(X = x')] - E[Y|X = x']P(X = x')}{P(X = x)} \end{aligned}$$

We see that the effect of treatment on the treated can be estimated from non-experimental (*do*-free expressions) and experimental data (expressions with the *do*-operator).

Part (b)

Because Gender is the only confounder, we can adjust for it and substitute the values of Table 1.1, specifically:

$$\begin{aligned}
E[Y_{X=0}] &= P(Y=1|do(X=0)) \\
&= \sum_z P(Y=1|X=0, Z=z)P(Z=z) \\
&= P(Y=1|X=0, Z=1)P(Z=1) + P(Y=1|X=0, Z=0)P(Z=0) \\
&= 0.87 * 357/700 + 0.69 * 343/700 \\
&= 0.78
\end{aligned}$$

Now, we can use this estimation in concert with our findings from (a):

$$\begin{aligned}
ETT &= E[Y=1|X=1] - \frac{E[Y=1|do(X=0)] - E[Y=1|X=0]P(X=0)}{P(X=1)} \\
&= 0.78 - \frac{0.78 - 0.83 * 0.5}{0.5} \\
&= 0.05
\end{aligned}$$

Part (c)

Because Z satisfies the backdoor criterion, we can directly write:

$$\begin{aligned}
ETT &= E[Y_1 - Y_0|X=1] \\
&= E[Y_{X=1}|X=1] - E[Y_{X=0}|X=1] \\
&= E[Y|X=1] - \Sigma_z P(Y=1|X=0, Z=z)P(Z=z|X=1) \\
&= E[Y|X=1] - P(Y=1|X=0, Z=1)P(Z=1|X=1) \\
&\quad - P(Y=1|X=0, Z=0)P(Z=0|X=1) \\
&= 0.78 - 0.87 * 87/350 - 0.69 * 263/350 \\
&= 0.05
\end{aligned}$$

We see that this method agrees with our findings from (b).

Study question 4.3.4. *Joe has never smoked before but, as a result of peer pressure and other personal factors he decided to start smoking. He buys a pack of cigarettes, comes home and asks himself: "I am about to start smoking, should I?"*

(a) Formulate Joe's question mathematically, in terms of ETT, assuming that the outcome of interest is lung cancer.

(b) What type of data would enable Joe to estimate his chances of getting cancer given that he goes ahead with the decision to smoke, versus refraining from smoking.

(c) Use the data in Table 3.1 to estimate the chances associated with the decision in (b).

Solution to study question 4.3.4

Part (a)

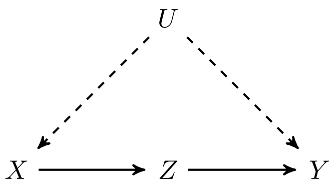
Let Y stand for lung cancer, with $Y = 1$ denoting that it is present. Let X be Joe's choice, with $X = 1$ indicating that he has decided to start smoking. So, we want to compute the effect of treatment on the treated to determine if Joe should start smoking *given* that he was about to, or specifically: $ETT = E[Y_1 - Y_0|X = 1]$, if $ETT > 0$, it means that smoking yields a higher chance of lung cancer for Joe (given that he was about to start smoking) than refraining from smoking.

Part (b)

Referencing our findings from study question 4.4.1, if we can find a set of variables that satisfy the backdoor or front-door criterion of the effect of X on Y, then we only need non-experimental data (see study question 4.4.1c); otherwise, we need both non-experimental and experimental data (see study question 4.4.1b).

Part (c)

First, let us recall the graphical model associated with this problem from Figure 3.10 in the book, wherein Z indicates the presence of Tar deposits in the lung, and U, an unmeasured Genotype that influences individuals to both smoke and get cancer:



Since Z satisfies the front-door criterion, by Theorem 3.4.1, and again referencing Table 3.1, we have:

$$
\begin{aligned}
E[Y|do(X=0)] &= \sum_z P(Z=z|X=0) \sum_{x'} P(Y=1|X=x', Z=z)P(X=x') \\
&= P(Z=1|X=0)[P(Y=1|X=1, Z=1)P(X=1) \\
&\quad + P(Y=1|X=0, Z=1)P(X=0)] \\
&\quad + P(Z=0|X=0)[P(Y=1|X=1, Z=0)P(X=1) \\
&\quad + P(Y=1|X=0, Z=0)P(X=0)] \\
&= 20/400 * [0.15 * 0.5 + 0.95 * 0.5] + 380/400 * [0.1 * 0.5 + 0.9 * 0.5] \\
&= 0.5025
\end{aligned}
$$

So, from our findings in 4.4.1 (b), we know:

$$\begin{aligned}
ETT &= E[Y_1 - Y_0|X = 1] \\
&= E[Y|X = 1] - E[Y_0|X = 1] \\
&= E[Y|X = 1] - \frac{E[Y|do(X = 0)] - E[Y|X = 0]P(X = 0)}{P(X = 1)} \\
&= 0.15 - \frac{0.5025 - 0.9025 * 0.5}{0.5} \\
&= 0.0475 > 0
\end{aligned}$$

Since our ETT is greater than 0, we know that the chance of cancer from smoking (given that Joe was about to start smoking) is greater than the alternative (no cancer). Therefore, Joe should refrain from smoking.

In this solution, we relied on two assumptions: (1) X is binary and (2) the model satisfies the front-door criterion. A more advanced analysis shows that assumtion (2) is sufficient for estimating ETT; X need not be binary once we have a front-door structure (Shpitser and Pearl, 2009).

Study question 4.5.1.
Consider the dilemma faced by Ms. Jones, as described in Example 4.4.3. Assume that, in addition to the experimental results of Fisher et al. (2002), she also gains access to an observational study, according to which the probability of recurrent tumor in all patients (regardless of therapy) is 30%, whereas among the recurrent cases, 70% did not choose therapy. Use the bounds provided in Eq. (4.30) to update her estimate that her decision was necessary for remission.

Solution to study question 4.5.1

First, according to the problem description, let X represent treatment (with x' representing lumpectomy alone and x representing Ms. Jones' decision: lumpectomy plus raditation) and Y represent recovery (with y' representing recurrence of cancer, and y representing the outcome for Ms. Jones: no recurrence). We also are given that:

$$\begin{aligned}
P(y') &= 0.3 \\
P(x'|y') &= 0.7 \\
P(y|do(x)) &= 0.39 \\
P(y|do(x')) &= 0.14
\end{aligned}$$

Our goal is to determine if Ms. Jones' decision was *necessary* for remission. So, we'll see if *PN* is more probable than not using the lower bound afforded by Eq. (4.30).

$$\begin{aligned} PN &\geq \frac{P(y) - P(y|do(x'))}{P(x,y)} \\ &= \frac{P(y) - P(y|do(x'))}{P(y|x)P(x)} \\ &= \frac{P(y) - P(y|do(x'))}{(1 - \frac{P(x|y')P(y')}{P(x)})P(x)} \\ &= \frac{P(y) - P(y|do(x'))}{P(x) - P(x|y')P(y')} \end{aligned}$$

At this point we consider whether we have all of the components for our computation, and see that we can find all parameters except for $P(x)$. However, because we are computing a lower bound for the PN, we can consider the parameterization that would yield its smallest value, namely, when the denominator is as large as possible with $P(x) = 1$. So, with this assumption, we write:

$$\begin{aligned} PN &\geq \frac{P(y) - P(y|do(x'))}{P(x) - P(x|y')P(y')} \\ &\geq \frac{0.7 - 0.14}{1 - (1 - 0.7) * 0.3} \\ &= 0.62 > 0.5 \end{aligned}$$

Since *PN* is greater than 0.5, we conclude that Ms. Jones' decision was likely necessary for remission.

Study question 4.5.2.
Consider the structural model:

$$y = \beta_1 m + \beta_2 t + u_y \tag{4.53}$$

$$m = \gamma_1 t + u_m \tag{4.54}$$

(a) Use the basic definition of the natural effects (Eqs. (4.46)–(4.47)) to determine TE, NDE and NIE.

(b) Repeat (a) assuming that u_y is correlated with u_m.

Solution to study question 4.5.2

Part (a)

To compute the NDE, we measure the expected increase in Y as the treatment T changes from $T = 0$ to $T = 1$ while the mediator M is set to the value it would have attained prior to the change (i.e., under $T = 0$), so by Eq. (4.46):

$$\begin{aligned} NDE &= E[Y_{1,M_0} - Y_{0,M_0}] \\ &= E[Y_{1,M_0}] - E[Y_{0,M_0}] \\ &= (\beta_1[\gamma_1 * 0 + u_m] + \beta_2 * 1 + u_y) - (\beta_1[\gamma_1 * 0 + u_m] + \beta_2 * 0 + u_y) \\ &= \beta_2 - 0 \\ &= \beta_2 \end{aligned}$$

Similarly, the NIE is defined as the expected increase in Y when the treatment is held constant at $T = 0$ and M changes to whatever value it would have attained under $T = 1$. By Eq. (4.47), we have:

$$\begin{aligned} NIE &= E[Y_{0,M_1} - Y_{0,M_0}] \\ &= E[Y_{0,M_1}] - E[Y_{0,M_0}] \\ &= (\beta_1[\gamma_1 * 1 + u_m] + \beta_2 * 0 + u_y) - (\beta_1[\gamma_1 * 0 + u_m] + \beta_2 * 0 + u_y) \\ &= \gamma_1\beta_1 - 0 \\ &= \gamma_1\beta_1 \end{aligned}$$

Finally, the total effect can be computed using Eq. (4.48):

$$\begin{aligned} TE &= E[Y_1 - Y_0] \\ &= NDE + NIE \\ &= \beta_2 + \gamma_1\beta_1 \end{aligned}$$

Note that, in this question, we did not have to assume that the treatment is randomized or, equivalently, that u_t is not correlated with u_y or y_m. This is because we have the functional form of the equations (linear), and we take the structural parameters as given.

Part (b)

The computations will remain the same since none of the above require that u_y is uncorrelated with u_m. This is because we are dealing with a linear system with given parameters.

Study question 4.5.3.
Consider the structural model:

$$y = \beta_1 m + \beta_2 t + \beta_3 tm + \beta_4 w + u_y \quad (4.55)$$

$$m = \gamma_1 t + \gamma_2 w + u_m \quad (4.56)$$

$$w = \alpha t + u_w \quad (4.57)$$

with $\beta_3 tm$ representing an interaction term.

(a) Use the basic definition of the natural effects (Eqs. (4.46) and (4.47)) (treating M as the mediator), to determine the portion of the effect for which mediation is necessary *($TE - NDE$) and the portion for which mediation is* sufficient *(NIE). Hint: Show that:*

$$NDE = \beta_2 + \alpha\beta_4 \tag{4.58}$$

$$NIE = \beta_1(\gamma_1 + \alpha\gamma_2) \tag{4.59}$$

$$TE = \beta_2 + (\gamma_1 + \alpha\gamma_2)(\beta_3 + \beta_1) + \alpha\beta_4 \tag{4.60}$$

$$TE - NDE = (\beta_1 + \beta_3)(\gamma_1 + \alpha\gamma_2) \tag{4.61}$$

(b) Repeat, using W as the mediator.

Solution to study question 4.5.3

Part (a)

We can use the same definitions and strategy as in Study Question 4.5.2. Let us first compute some values of our covariates for ease of reference within our solution:

$$E[W_0] = \alpha * 0 = 0$$
$$E[W_1] = \alpha * 1 = \alpha$$
$$E[M_0] = \gamma_1 * 0 + \gamma_2 * 0 = 0$$
$$E[M_1] = \gamma_1 * 1 + \gamma_2\alpha = \gamma_1 + \gamma_2\alpha$$

Now we can compute our target quantities:

$$\begin{aligned}
NDE &= E[Y_{1,M_0} - Y_{0,M_0}] \\
&= E[Y_{1,M_0}] - E[Y_{0,M_0}] \\
&= (\beta_1 * 0 + \beta_2 * 1 + \beta_3 * 1 * 0 + \beta_4\alpha) - (\beta_1 * 0 + \beta_2 * 0 + \beta_3 * 0 * 0 + \beta_4 * 0) \\
&= \beta_2 + \alpha\beta_4 \\
NIE &= E[Y_{0,M_1} - Y_{0,M_0}] \\
&= E[Y_{0,M_1}] - E[Y_{0,M_0}] \\
&= (\beta_1[\gamma_1 + \gamma_2\alpha] + \beta_2 * 0 + \beta_3 * 0 + \beta_4 * 0) - (\beta_1 * 0 + \beta_2 * 0 + \beta_3 * 0 * 0 + \beta_4 * 0) \\
&= \beta_1(\gamma_1 + \alpha\gamma_2) \\
TE &= E[Y_1 - Y_0] \\
&= E[Y_1] - E[Y_0] \\
&= (\beta_1[\gamma_1 + \gamma_2\alpha] + \beta_2 * 1 + \beta_3[\gamma_1 + \gamma_2\alpha] + \beta_4\alpha) - (\beta_1 * 0 + \beta_2 * 0 + \beta_3 * 0 * 0 + \beta_4 * 0) \\
&= \beta_2 + (\gamma_1 + \alpha\gamma_2)(\beta_1 + \beta_3) + \alpha\beta_4
\end{aligned}$$

So, combining our computations from above, we find that the portion of the effect for which mediation is necessary is:

$$\begin{aligned} TE - NDE &= \beta_2 + (\gamma_1 + \alpha\gamma_2)(\beta_1 + \beta_3) + \alpha\beta_4 - (\beta_2 + \alpha\beta_4) \\ &= (\gamma_1 + \alpha\gamma_2)(\beta_1 + \beta_3) \end{aligned}$$

Part (b)

We'll use the same strategy that we did for part (a), first computing the values of W, M under the changing treatment T.

$$\begin{aligned} E[W_0] &= \alpha * 0 = 0 \\ E[W_1] &= \alpha * 1 = \alpha \\ E[M_{0,W_0}] &= \gamma_1 * 0 + \gamma_2 * 0 = 0 \\ E[M_{0,W_1}] &= \gamma_1 * 0 + \gamma_2 * \alpha = \gamma_2 * \alpha \\ E[M_{1,W_0}] &= \gamma_1 * 1 + \gamma_2 * 0 = \gamma_1 \\ E[M_{1,W_1}] &= \gamma_1 * 1 + \gamma_2 * \alpha \end{aligned}$$

Once more, using the above to compute our target quantities, we have:

$$\begin{aligned} NDE &= E[Y_{1,W_0} - Y_{0,W_0}] \\ &= E[Y_{1,W_0}] - E[Y_{0,W_0}] \\ &= (\beta_1[\gamma_1] + \beta_2 * 1 + \beta_3 * 1[\gamma_1] + \beta_4 * 0) - (\beta_1 * 0 + \beta_2 * 0 + \beta_3 * 0 + \beta_4 * 0) \\ &= \beta_1\gamma_1 + \beta_2 + \beta_3\gamma_1 \\ NIE &= E[Y_{0,W_1} - Y_{0,W_0}] \\ &= E[Y_{0,W_1}] - E[Y_{0,W_0}] \\ &= (\beta_1[\gamma_2\alpha] + \beta_2 * 0 + \beta_3 * 0 + \beta_4 * \alpha) - (\beta_1 * 0 + \beta_2 * 0 + \beta_3 * 0 + \beta_4 * 0) \\ &= \alpha\beta_1\gamma_2 + \alpha\beta_4 \\ TE &= E[Y_1 - Y_0] \\ &= E[Y_1] - E[Y_0] \\ &= (\beta_1[\gamma_1 + \gamma_2\alpha] + \beta_2 * 1 + \beta_3[\gamma_1 + \gamma_2\alpha] + \beta_4\alpha) - (\beta_1 * 0 + \beta_2 * 0 + \beta_3 * 0 * 0 + \beta_4 * 0) \\ &= \beta_2 + (\gamma_1 + \alpha\gamma_2)(\beta_1 + \beta_3) + \alpha\beta_4 \end{aligned}$$

So, combining our computations from above, we know that the portion of the effect for which mediation is necessary is:

$$\begin{aligned} TE - NDE &= \beta_2 + (\gamma_1 + \alpha\gamma_2)(\beta_1 + \beta_3) + \alpha\beta_4 - [\beta_1\gamma_1 + \beta_2 + \beta_3\gamma_1] \\ &= \beta_2 + \beta_1\gamma_1 + \beta_1\alpha\gamma_2 + \beta_3\gamma_1 + \beta_3\alpha\gamma_2 + \alpha\beta_4 - \beta_1\gamma_1 - \beta_2 - \beta_3\gamma_1 \\ &= \alpha\gamma_2(\beta_1 + \beta_3) + \alpha\beta_4 \end{aligned}$$

Study question 4.5.4.

Apply the mediation formulas provided in this section to the discrimination case discussed in Section 4.4.4, and determine the extent to which ABC International practiced discrimination in their hiring criteria. Use the data in Tables 4.6 and 4.7, with $T = 1$ standing for male applicants, $M = 1$ standing for highly qualified applicants, and $Y = 1$ standing for hiring. (Find the proportion of the hiring disparity that is due to gender, and the proportion that could be explained by disparity in qualification alone.)

Solution to study question 4.5.4

[Hint: this is precisely the numerical problem on homework mediation as presented in the text; see Tables 4.5, 4.7 and computations that follow]. Our goal is to find the proportion of hiring disparity that is due to gender, and the proportion that could be explained by disparity in qualification alone. Using the same strategies as the homework-training program example: the quantity NIE/TE tells us what proportion of the disparity is due to qualification alone and the quantity $1 - NDE/TE$ tells us what proportion of the disparity is due to gender. Assuming that there exists no unobserved confounding, we'll compute our quantities of interest using Eqs. (4.51), (4.52), (4.44):

$$
\begin{aligned}
NDE &= \sum_m [E[Y|T=1,M=m] - E[Y|T=0,M=m]]P(M=m|T=0) \\
&= [E[Y|T=1,M=0] - E[Y|T=0,M=0]]P(M=0|T=0) \\
&\quad + [E[Y|T=1,M=1] - E[Y|T=0,M=1]]P(M=1|T=0) \\
&= (0.4 - 0.2)(1 - 0.4) + (0.8 - 0.3)0.4 \\
&= 0.32 \\
NIE &= \sum_m E[Y|T=0,M=m][P(M=m|T=1) - P(M=m|T=0)] \\
&= E[Y|T=0,M=0][P(M=0|T=1) - P(M=0|T=0)] \\
&\quad + E[Y|T=0,M=1][P(M=1|T=1) - P(M=1|T=0)] \\
&= (0.75 - 0.4)(0.3 - 0.2) \\
&= 0.035
\end{aligned}
$$

$$
\begin{aligned}
TE &= E[Y_1 - Y_0] \\
&= E[Y_1] - E[Y_0] \\
&= E[Y|do(T=1)] - E[Y|do(T=1)] \\
&= \sum_m E[Y|do(T=1), M=m]P(M=m|do(T=1)) \\
&\quad - \sum_m E[Y|do(T=0), M=m]P(M=m|do(T=0)) \\
&= \sum_m E[Y|T=1, M=m]P(M=m|T=1) - \sum_m E[Y|T=0, M=m]P(M=m|T=0) \\
&= [E[Y|T=1, M=0]P(M=0|T=1) + E[Y|T=1, M=1]P(M=1|T=1)] \\
&\quad - [E[Y|T=0, M=0]P(M=0|T=0) + E[Y|T=0, M=1]P(M=1|T=0)] \\
&= [0.4 * (1-0.75) + 0.8 * 0.75] - [0.2 * (1-0.4) + 0.3 * 0.4] \\
&= 0.46
\end{aligned}
$$

Now, we have all the components needed to make claims about the hiring disparity; in particular:

$$NIE/TE = 0.035/0.46 = 0.07$$

$$1 - NDE/TE = 1 - 0.32/0.46 = 0.304$$

So, from the above, we conclude that 30.4% of the hiring disparity is due to gender and 7% of the proportion could be explained by disparity in qualification alone.