Philosophy 12: Introduction to Causal Reasoning
Answer key to the study questions

Section 1.

1. Categorize the following: The St. Louis Rams win the Superbowl on January 30th, 2000.
   - [✓ (a)] A particular event
   - (b) A type of event
2. Categorize the following: Catching the Flu in the Winter.
   - (a) A particular event
   - [✓ (b)] A type of event
3. A response structure is:
   - (a) A specification of the value of each variable besides the effect.
   - (b) A set of variables.
   - [✓ (c)] A causal system with a designated effect, a list of every possible casual situation for that effect, and whether the effect occurs in each.
   - (d) A set of potential causal factors.
   - (e) All of the above.
   - (f) None of the above
4. Response structure uniformity is:
   - [✓ (a)] The assumption that all individuals in the population have the same response structure.
   - (b) The assumption that all individuals in the population are in the same causal assignment.
   - (c) The assumption that all individuals in the population have the same potential causal factors.
   - (d) None of the above.
5. A deterministic system is one in which:
   - [✓ (a)] The individual's causal assignment determines whether the effect occurs
   - (b) The values of the observed causal factors determine whether the effect occurs.
   - (c) We have determined all of the actual causal factors.
   - (d) All of the above
6. Consider the claim "People with cats are more likely to say that they are happy than people without cats." This is:
   - [✓ (a)] An associational claim.
   - (b) A causal claim.
   - (c) Both a causal claim and an associational claim.
   - (d) Neither a causal claim nor an associational claim.

7. Assume all of the following:

    - a randomly chosen San Diegan has a 50% chance of living within 2 miles of the ocean;
    - a randomly chosen San Diegan has a 40% chance of living in an apartment.
    - Living within 2 miles of the ocean and living in an apartment are positively associated, since living near the ocean is a cause of living in an apartment.

    If we pick a person at random, and we find out that the person lives in an apartment, what is the chance that the person lives within two miles of the ocean?

    [✓ (a)] Above 50%
    (b) Exactly 50%
    (c) Below 50%
    (d) Not enough information to tell

8. If a property P_1 is *necessary* for an effect E, then (circle all that apply):

    (a) every time P_1 occurs, E also occurs.
    [✓ (b)] every time E occurs, P_1 also occurs.
    (c) whenever P_1 doesn't occur, E does occur.
    (d) whenever E doesn't occur, P_1 never occurs

9. Consider the following table:

CAUSAL FACTOR 1	CAUSAL FACTOR 2	CAUSAL FACTOR 3	EFFECT
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	Yes	Yes
Yes	No	No	No
No	Yes	Yes	No
No	Yes	No	No
No	No	Yes	No
No	No	No	Yes

    Which causal factors are jointly sufficient for the effect to occur?

    (a) Causal Factors 1 and 2
    (b) Causal Factors 2 and 3
    [✓ (c)] Causal Factors 1 and 3
    (d) None of the above

10. An ideal intervention: (circle all that apply)

    [✓ (a)] targets only one variable
    (b) must be practical
    (c) is not actually implemented
    [✓ (d)] completely determines the value of the target variable

11. Consider the variables STUDENT and CAR OWNER. STUDENT can take on the values { Student, Not a student } and CAR OWNER can take on the values { Owns car, Doesn't own car }. Assume we're interested in the effect CAR OWNER = Owns car, and we take data on 500 people:

	Owns Car	Doesn't own car	Total
Student	100	100	200
Not a student	300	0	300
Total	400	100	500

Does this system appear deterministic (relative to these variables)?

(a) Yes
[✓ (b)] No
(c) Not enough information to tell

12. Pseudo-indeterministic systems are: (circle all that apply)

(a) systems that include all of the actual causal factors.
(b) systems where we don't see all of the causal assignments.
[✓ (c)] systems that appear indeterministic, but are actually deterministic when all the relevant variables are included.
(d) systems that appear deterministic, but are actually indeterministic when all the relevant variables are included.

For questions 13 and 14, consider the following data table:

Individual	Sex	Hair Color	Smoker
1	Male	Blond	Yes
2	Female	Dark	Yes
3	Female	Blond	No
4	Male	Blond	No
5	Female	Dark	Yes
6	Male	Blond	No
7	Female	Blond	No
8	Male	Dark	Yes

13. What is the relative frequency of blond males?

(a) $\frac{1}{2}$
[✓ (b)] $\frac{3}{8}$
(c) $\frac{3}{4}$
(d) $\frac{3}{5}$

14. What is the relative frequency of female smokers?

(a) $\frac{1}{2}$
(b) $\frac{1}{3}$
[✓ (c)] $\frac{1}{4}$
(d) $\frac{1}{8}$

15. If property A is independent of property B given property C, and $\mathrm{Fr}_S(A) > 0$, $\mathrm{Fr}_S(B) > 0$, and $\mathrm{Fr}_S(C) > 0$, then which of the following will *always* be true? (circle all that apply)

(a) $\mathrm{Fr}_S(A \mid C) = \mathrm{Fr}_S(B \mid C)$
[✓ (b)] $\mathrm{Fr}_S(B \mid C) = \mathrm{Fr}_S(B \mid A, C)$
[✓ (c)] $\mathrm{Fr}_S(A \bullet B \mid C) = \mathrm{Fr}_S(A \mid C) \cdot \mathrm{Fr}_S(B \mid C)$
(d) $\mathrm{Fr}_S(A \mid B, {\sim}S) = \mathrm{Fr}_S(A \mid {\sim}C)$

For questions 16 and 17, consider the following contigency table:

	Recovered	Died
Experimental Drug	65	30
Control	55	25

16. What is the relative frequency of death overall?

    (a) $\frac{30}{65}$
    (b) $\frac{30}{95}$
    (c) $\frac{55}{120}$
    [✓ (d)] $\frac{55}{175}$

17. What is the relative frequency of recovery among those who took the experimental drug?

    (a) $\frac{30}{65}$
    [✓ (b)] $\frac{65}{95}$
    (c) $\frac{65}{120}$
    (d) $\frac{95}{175}$

18. If two properties A and B are independent, which of the following are true?
    (Boxed choices are true.)

    [✓ (a)] Not having A and not having B are independent.
    (b) The relative frequency of A is the same as the relative frequency of B among individuals who have A.
    [✓ (c)] $\text{Fr}(A \bullet B) = \text{Fr}(A) \cdot \text{Fr}(B)$.
    (d) $\text{Fr}(A) = \text{Fr}(B)$.

Consider the variables Education, Book Reader, and Income. Suppose that Education can take on the values { High School, College }, Book Reader can take on the values { Often, Rarely }, and Income can take on the values { High, Low }. Consider the following contigency table:

	Often, High	Often, Low	Rarely, High	Rarely, Low	Total
High School	30	20	20	30	100
College	40	60	60	40	200
Total	70	80	80	70	300

19. Is Education independent of Book Reader?

    [✓ (a)] Yes
    (b) No
    (c) Not enough information to tell

20. Is Education independent of Book Reader given Income?

    (a) Yes
    [✓ (b)] No
    (c) Not enough information to tell

21. A non-collider is: (circle all that apply)

    [✓ (a)] A mediator
    [✓ (b)] A common cause
    (c) A common effect
    (d) None of the above

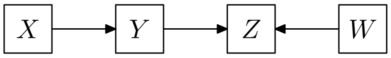
22. Consider the following response structure:

Situation	Causal Factor 1	Causal Factor 2	Causal Factor 3	Effect
1	Yes	Yes	Yes	Yes
2	Yes	Yes	No	Yes
3	Yes	No	Yes	Yes
4	Yes	No	No	Yes
5	No	Yes	Yes	No
6	No	Yes	No	Yes
7	No	No	Yes	No
8	No	No	No	Yes

Should there be a direct arrow from Causal Factor 2 to Effect?

(a) Yes
[✓ (b)] No
(c) Not enough information to tell

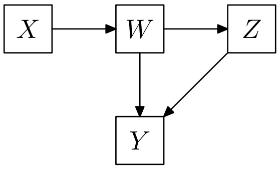
23. Consider the following causal graph:



Suppose that this causal graph accurately depicts the causal relations among the variables X, Y, Z, and W. Is X a direct cause of Z relative to the variables $\{X, Z, W\}$?

[✓ (a)] Yes
(b) No
(c) Not enough information to tell

24. Consider the following causal graph:



An intervention on W changes the causal relationship between: (circle all that apply)

[✓ (a)] X and Y
[✓ (b)] X and W
(c) Y and W
(d) Z and W

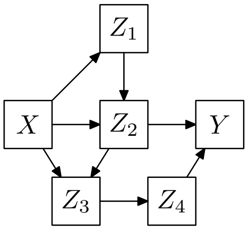
25. Given a group of associations and independencies, we can ______________ determine the causal graph that produced the group. (Which choice correctly fills in the blank.)

(a) Always
[✓ (b)] Sometimes
(c) Never

26. If A is associated with B, and B is associated with C, then A is ______________ associated with C. (Which choice correctly fills in the blank.)

(a) Always
[✓ (b)] Sometimes
(c) Never

For questions 27–30, consider the following causal graph.



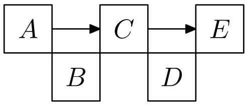
27. Which variables are common causes of Z_4 and Y? (Circle all that apply)
    - [✓ (a)] X
    - (b) Z_1
    - [✓ (c)] Z_2
    - (d) Z_3

28. How many causal connections are there between X and Y?
    - (a) 1
    - (b) 2
    - (c) 3
    - (d) 4
    - [✓ (e)] 5

29. Which of the following independencies hold? (Circle all that apply)
    - [✓ (a)] $Z_2 \perp\!\!\!\perp Z_4 \mid Z_3$
    - (b) $X \perp\!\!\!\perp Y \mid Z_2$
    - [✓ (c)] $X \perp\!\!\!\perp Y \mid Z_2, Z_3$
    - (d) $Z_1 \perp\!\!\!\perp Y \mid Z_2$

30. How many undirected paths are there between X and Z_2?
    - (a) 0
    - (b) 1
    - (c) 2
    - (d) 3
    - [✓ (e)] 4

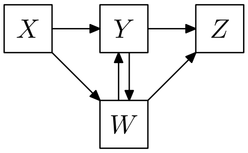
31. Consider the following causal graph:



Which of the following independencies hold? (Circle all that apply)

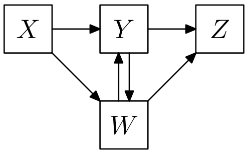
- (a) $A \perp\!\!\!\perp B$
- [✓ (b)] $C \perp\!\!\!\perp D$
- (c) $C \perp\!\!\!\perp E$
- [✓ (d)] $B \perp\!\!\!\perp D$

32. Assume that the causal graph below describes the relations among the variables X, Y, Z, and W in the pre-manipulated state:

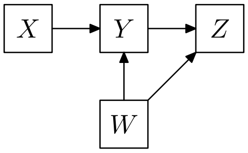


If we ideally intervene on Y, which of the following is the post-manipulation graph?

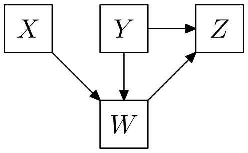
(a)



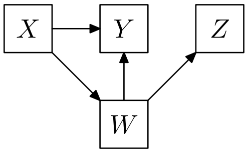
(b)



[✓ (c)]



(d)



(e) None of the above

33. Which of the following kinds of undirected paths, $A \cdots C \cdots B$ (where $\cdots$ stands for some kind of connection between nodes) produce an association between A and B? (Circle all that apply)

(a) C is a common cause of A and B
(b) C is a mediator between A and B
(c) C is a common effect of A and B
(d) None of the above

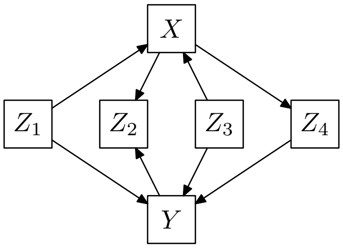
34. Suppose that the Food and Drug Administration commissioned an extensive randomized test of a new drug and found that it had *no* effect on weight loss. With respect to just the variables Take Drug and Lose Weight, which of the following frequency relations would you expect to hold? (Circle all that apply)

(a) Fr(Take Drug • Lose Weight) = Fr(Take Drug)
(b) Fr(Take Drug • Lose Weight) = Fr(Lose Weight)
(c) Fr(Take Drug • Lose Weight) = Fr(Take Drug) + Fr(Lose Weight)
[✓ (d)] Fr(Take Drug • Lose Weight) = Fr(Take Drug) · Fr(Lose Weight)
(e) None of the above

35. In an experiment to test whether an area has more plant growth when seeded with a variety of plants, it was found that areas seeded with a variety of plants yielded more plant growth than did areas seed with a single variety because the areas seeded with a variety of plants were more likely to include one dominant species which flourished in the environment. Which of the following frequency relations would you expect to hold? (using V for a variety of plants seeded, D for a dominant species, and G for a large plant growth)

    [✓ (a)] In areas with D, V is not associated with G
    (b) In areas with G, V is not associated with D
    (c) V is not associated with G
    (d) None of the above

36. Consider the following causal graph:



If we want to judge the influence of X on Y, which variables do we need to control for? (Circle all that apply)

    [✓ (a)] Z_1
    (b) Z_2
    [✓ (c)] Z_3
    (d) Z_4

37. The association between two variables X and Y is confounded if: (Circle all that apply)

    [✓ (a)] there is a common cause of X and Y
    (b) we cannot intervene on X
    (c) an effect of X is associated with Y
    [✓ (d)] a cause of X is associated with Y

38. Assume we have three variables X, Y, and Z. If we want to determine the causal influence X has on Y, and we control for Z, we will get the wrong determination when:

    (a) Z is a common cause of X and Y
    [✓ (b)] Z is a common effect of X and Y
    (c) Z is not causally connected to either X or Y
    (d) Z directly causes X and indirectly causes Y

39. If we measure all of the confounding variables we know about, we can ________________ determine the actual causal influence of one variable on another. (Which choice correctly fills in the blank?)

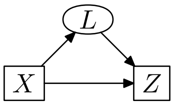
    (a) Always
    [✓ (b)] Sometimes
    (c) Never

40. After an intervention on W, which of the following causal connections can hold between W and U? (Circle all that apply)

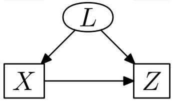
(a) common cause of W and U
(b) common effect of W and U
[✓ (c)] causal path from W to U
(d) confounding between W and U

41. If we ideally intervene on X in the following graphs, then which post-manipulation graphs make the same prediction about the association of X and Z?

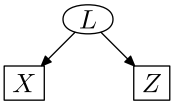
[✓ (a)]



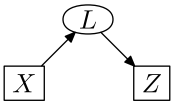
[✓ (b)]



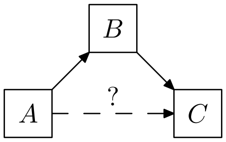
(c)



[✓ (d)]



42. Consider the following graph:



If we want to find out whether A is also a direct cause of C, we should experimentally control B

[✓ (a)] True
(b) False
(c) Not enough information to tell

43. Suppose the Center for Disease Control (CDC) announces that they have found a negative association between how much oat bran people consume and their cholesterol level. People who consume high amounts of oat bran have lower cholesterol than people who consume low amounts of oat bran. A student in epidemiology 101 announces that this association might be confounded and not due to a causal relation in which oat bran consumption causes lower cholesterol. The student says his supermarket will provide an opportunity to collect nonconfounded data. After the newspaper article on the CDC results came out, the supermarket announced a big sale on oat bran, and the student proposes to measure the cholesterol level in people who have never bought oat bran until this sale, and then to measure their cholesterol level again one month later — after they have consumed the oat bran they bought. How would you assess the value of his proposed experiment?

(a) The supermarket sale is an ideal intervention on oat bran use
(b) The supermarket sale is a randomized assignment of treatment
[✓ (c)] Even if 1st time oat bran buyers have lower cholesterol one month into the experiment, there is still a chance that the association is due to an unmeasured confounded.
(d) None of the above

44. Randomizing (as a method of experimental control) helps us *avoid*: (circle all that apply)

(a) ideal interventions
[✓ (b)] “fat hand” interventions
[✓ (c)] confounding
(d) observed causal paths

45. Suppose that smoking causes lung disease and heart disease, obesity causes heart disease, and working in coal mines causes lung disease. Smoking is not associated with working in coal mines or with obesity, and working in coal mines is not associated with obesity. Which of the following would you expect to be true? (Circle all that apply)

(a) Among obese people, heart disease is not associated with smoking
(b) Among coal miners, lung disease is not associated with smoking
[✓ (c)] Among people who smoke, the association of heart disease and lung cancer is small than in the general population
[✓ (d)] Among people who do not smoke, the association of heart disease and lung disease is smaller than in the general population
(e) None of the above

Section 2.

1. What kind of information does a causal claim give us that an associational claim does not?

Answer: The claim that X is associated with Y does not tell us what would happen if we intervened on the variable X or on the variable Y; on the other hand, the claim that X causes Y does give us information about what would happen if we intervened on X, of if we intervened on Y.

2. Give an example (in words, not just a graph) of a causal chain, ensuring that no variable is the direct cause of more than one other.

Answer: An example would be something like the refrigerator door: opening the door causes the light switch to turn on, which causes the light to come on. The door is only a direct cause of the switch and the switch is only a direct cause of the light.

3. What is the difference between a deterministic system and a pseudo-indeterministic system?

Answer: In a deterministic system, knowing the causal assignment tells you exactly whether the effect will occur. In an indeterministic system, there is at least one causal assignment for which you are unsure whether or not the effect will occur. A pseudo-indeterministic system appears indeterministic when we only consider some set of causal factors, but appears deterministic if we consider some larger set of causal factors.

4. The following contingency table shows the results of a study of a large number of individuals, some of whom were treated with an experimental drug. The study was looking for frequencies of Disease X among these individuals.

	Experimental Drug = Yes	Experimental Drug = No
Disease X = Yes	500	1,000
Disease X = No	2,000	4,000

What is the frequency of disease X in this population? What is the conditional frequency of having disease X on having taken the experimental drug in this population?

Answer:

$$\text{Fr(Disease } X = \text{Yes)} = \frac{\#\text{ with Disease } X}{\text{total }\#\text{ of people}} = \frac{1500}{7500} = \frac{1}{5}$$

$$\begin{aligned}\text{Fr(Disease } X &= \text{Yes} \mid \text{Experimental Drug} = \text{Yes)} \\ &= \frac{\#\text{ with Disease } X = \text{Yes and Experimental Drug} = \text{Yes}}{\#\text{ with Experimental Drug} = \text{Yes}} \\ &= \frac{500}{2500} = \frac{1}{5}\end{aligned}$$

5. Consider the variables Height and Eye Color. Height can take on the values { Short, Medium, Tall } and { Eye Color } can take on the values { Blue, Green, and Brown }. (1) List all of the independence relations that must hold if Height is independent of Eye Color. (2) For each independence relation, give two frequencies that could be checked to see if that independence relation were true.

Answer: (1) The independence relations are:

- Height = Short $\perp\!\!\!\perp$ Eye Color = Blue
- Height = Medium $\perp\!\!\!\perp$ Eye Color = Blue
- Height = Tall $\perp\!\!\!\perp$ Eye Color = Blue
- Height = Short $\perp\!\!\!\perp$ Eye Color = Green
- Height = Medium $\perp\!\!\!\perp$ Eye Color = Green
- Height = Tall $\perp\!\!\!\perp$ Eye Color = Green
- Height = Short $\perp\!\!\!\perp$ Eye Color = Brown
- Height = Medium $\perp\!\!\!\perp$ Eye Color = Brown
- Height = Tall $\perp\!\!\!\perp$ Eye Color = Brown

(2) The relative frequencies that could be checked are:

- Height = Short $\perp\!\!\!\perp$ Eye Color = Blue could be check with Fr(Height = Short) and Fr(Height = Short | Eye Color = Blue), or with Fr(Eye Color = Blue) and Fr(Eye Color = Blue | Eye Color = Blue).
- Height = Short $\perp\!\!\!\perp$ Eye Color = Green could be check with Fr(Height = Short) and Fr(Height = Short | Eye Color = Green), or with Fr(Eye Color = Green) and Fr(Eye Color = Green | Eye Color = Green).
- Height = Short $\perp\!\!\!\perp$ Eye Color = Brown could be check with Fr(Height = Short) and Fr(Height = Short | Eye Color = Brown), or with Fr(Eye Color = Brown) and Fr(Eye Color = Brown | Eye Color = Brown).
- Height = Medium $\perp\!\!\!\perp$ Eye Color = Blue could be check with Fr(Height = Medium) and Fr(Height = Medium | Eye Color = Blue), or with Fr(Eye Color = Blue) and Fr(Eye Color = Blue | Eye Color = Blue).
- Height = Medium $\perp\!\!\!\perp$ Eye Color = Green could be check with Fr(Height = Medium) and Fr(Height = Medium | Eye Color = Green), or with Fr(Eye Color = Green) and Fr(Eye Color = Green | Eye Color = Green).
- Height = Medium $\perp\!\!\!\perp$ Eye Color = Brown could be check with Fr(Height = Medium) and Fr(Height = Medium | Eye Color = Brown), or with Fr(Eye Color = Brown) and Fr(Eye Color = Brown | Eye Color = Brown).
- Height = Tall $\perp\!\!\!\perp$ Eye Color = Blue could be check with Fr(Height = Tall) and Fr(Height = Tall | Eye Color = Blue), or with Fr(Eye Color = Blue) and Fr(Eye Color = Blue | Eye Color = Blue).
- Height = Tall $\perp\!\!\!\perp$ Eye Color = Green could be check with Fr(Height = Tall) and Fr(Height = Tall | Eye Color = Green), or with Fr(Eye Color = Green) and Fr(Eye Color = Green | Eye Color = Green).

- Height = Tall $\perp\!\!\!\perp$ Eye Color = Brown could be check with Fr(Height = Tall) and Fr(Height = Tall | Eye Color = Brown), or with Fr(Eye Color = Brown) and Fr(Eye Color = Brown | Eye Color = Brown).

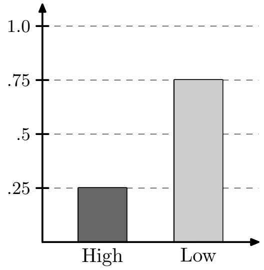
6. Consider the variables Reading Ability, Income, and Education. Reading Ability can take on the values { Excellent, Average }, Income can take on the values { High, Medium, Low }, and Education can take on the values { High School, College }. Draw fully labeled histograms showing that Reading Ability = Excellent is independent of Income = High given Education.

   Answer: Reading Ability = Excellent is independent of Income = High given Education means that:

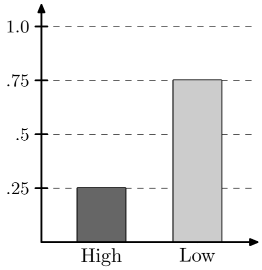
   - Reading Ability = Excellent is independent of Income = High given Education = High School, and
   - Reading Ability = Excellent is independent of Income = High given Education = College.

   Thus, for this question you will need two sets of histograms, one for each value of Education. First, for Education = High School:

Fr(Income | Education = High School)

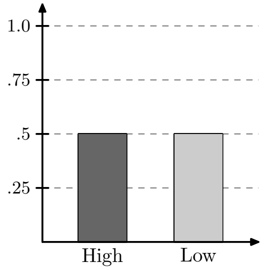


Fr(Income | Education = High School & Reading Ability = Excellent)

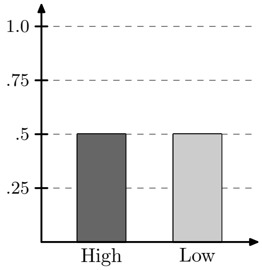


   Then for Education = College:

Fr(Income | Education = College)

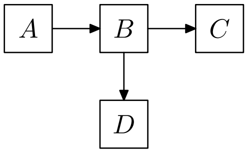


Fr(Income | Education = College & Reading Ability = Excellent)

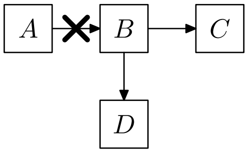


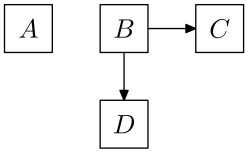
7. Assume that A is a direct cause of B, B is a direct cause of D, B is a direct cause of C, and no other variables are direct causes of any other variables. Draw the pre-manipulation and post-manipulation graphs for an ideal intervention on B.

   Answer: The pre-manipulation graph is:



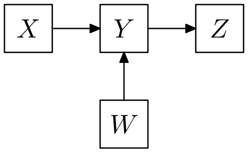
   The post-manipulation graph can be represented by either graph below:





8. Draw a single causal graph with the variables W, X, Y, and Z in which Y is both a collider and a non-collider.

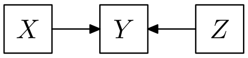
   Answer: Here is one example:

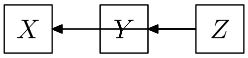
9. Draw two different causal graphs using X, Y, and Z such that *both* predict:

   - X is associated with Y, and
   - X and Y are independent conditional on Z

   Answer: Here are two examples:

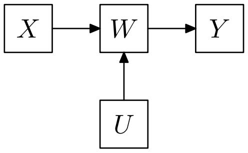
   

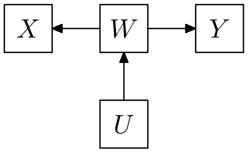
   and

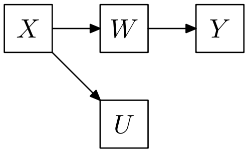
   

10. Suppose we observe that X and Y are associated, but independent conditional on W. Construct 3 distinct causal graphs among the variables $\{X, Y, W, U\}$ that predict these facts, and construct 3 more that don't predict it.

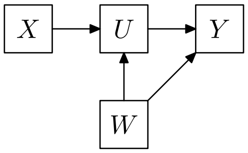
    Answer: Three graphs that do predict $X \not\perp\!\!\!\perp Y$ and $X \not\perp\!\!\!\perp Y \mid W$:

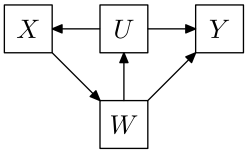
    

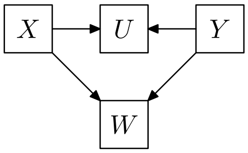
    

    Three graphs that do not predict $X \not\perp\!\!\!\perp Y$ and $X \not\perp\!\!\!\perp Y \mid W$:

11. What is an ideal intervention on a variable X, and how does it affect the relationships between X and other variables in the causal system?

    Answer: An ideal intervention on X is some specific action (whether practical or only hypothetical) which completely determines the value of X and does not directly cause any other variables. An ideal intervention breaks the causal relationship between X and its direct causes, and does not change any other causal relationships in the system.

12. Suppose we wanted to see if a gasoline additive actually improved gas mileage. Describe an experiment that would answer this causal question.

    Answer: We might perform an experiment similar to the one described in the lecture notes to see if red cars cause tickets. We might, for example, take 10 pairs of different makes of cars, making sure that each pair was exactly the same. We would then get pairs of drivers that are the same weight, and have the pairs of drivers drive together for a month. In each pair, one car would have gasoline with the additive, and the other would have gasoline without it. The drivers would drive on various different roads, at various speeds, with varying amounts of weight in the car, at varying times of the day and night, but within each pair, the drivers would always do the same thing. The drivers would always fill up with exactly the same gas, and only one of the pair would continually add the additive to his or her car. In this experiment, then, each pair of cars would have exactly the same experience except for the additive, and if there were any difference in gas mileage, we would know that it had to be due to the additive.

13. Explain the idea of randomized assignment of treatment and why we use it.

    Answer: In our terminology, "treatment" is the cause we're interested in and "response" is the effect we're interested in; specifically, we want to know if the treatment causes the response. To do this, we might want to perform an intervention on treatment. For example, have half the subjects get the treatment and half the subjects not. In order to do this, though, we have to decide how we are going to choose who gets the treatment. We do not want the selection criterion to introduce a confounder in our experiment (that is, we do not want the selection criterion to be a common cause of both the treatment and the response). Thus, we need a selection criterion which is completely independent of any of the causes of the response. If we randomly select subjects for the treatment, we can be sure that this method is independent of all other causes of the response and, hence, in the long run any association between treatment and response can only be due to the treatment causing the response.

14. In a random sample of ten thousand college students, family wealth was associated with student grade point average. Social scientist A claimed the result was evidence that studying harder would increase family wealth, and social scientist B claimed the result was evidence that family wealth causes improved school performance. Write a short critique of social scientist A and another of social scientist B.

    Answer: Both scientists are falling victim to the idea of thinking that association is the same as causation. What each scientist is doing is using associational evidence to make a claim about what would happen in the case of an intervention. To know what would happen in the case of an intervention, however, we need to know the correct causal graph. Unfortunately, the evidence "Family Wealth is associated with Grade Point Average" can justify many different causal graphs, only two of which are FAMILY WEALTH → GRADE POINT AVERAGE and GRADE POINT AVERAGE → FAMILY WEALTH. It could also be the case, for example, that FAMILY WEALTH and GRADE POINT AVERAGE have a common cause, or it could be a combination of these three possibilities. In any case, each scientist is making too strong a claim based on the evidence.