PH 250B Week 2, Tab 1 Ungraded quiz - ANSWERS
Topic: Causality

Problem 1. (Source: 250B Problem Set)
Using the counterfactual framework, describe the ideal experiment that you would perform if you wished to examine the causal relationship between growing up in a home with high levels of lead (compared to low levels of lead) and cognitive development score at age five.

In the ideal experiment, you would take a population of children, have them grow up in a home with lead from conception to age 5 and then measure their cognitive development score. You would then go back in time, take the exact same population of children, have them grow up in a home with low levels of lead from conception to age 5 and then measure their cognitive development score. You would also make sure that everything else was exactly the same between the two scenarios.

a. Explain why you could not conduct this experiment and what experiment you could conduct instead. What might differ between your realistic experiment and the ideal experiment?
   You cannot 'go back in time', and you cannot ethically expose children to higher levels of lead. It would also be very difficult and expensive to collect data on all children in your population (some will move, etc.). Instead, you could take a sample of children from the population, measure the lead in their homes at one or more time points between conception and age 5, and then compare cognitive development scores between children who had high and low lead exposures. However, children exposed to high and low lead are likely to differ in other ways as well.

b. Write a mathematical expression to represent the **individual**-level causal effect for exposure to high levels of lead and child behavioral problems (hint: see lecture slides). Interpret this equation.
   $Y_1 – Y_0$
   or
   $Y_{high\ lead} – Y_{low\ lead}$
   This expression represents the counterfactual (i.e. causal) difference in cognitive development score for an individual child if he/she were exposed to high lead compared to the same child's cognitive development score if he/she were exposed to low lead (going back in time and keeping everything else the same).

c. Write a mathematical expression to represent the **population** level causal effect for exposure to high levels of lead and child behavioral problems. Interpret this mathematical expression.

   $E[Y_1 - Y_0]$
   or
   $E[Y_{high\ lead} – Y_{low\ lead}]$

This expression represents the counterfactual (i.e. causal) difference in cognitive development score if *the whole population of children were* exposed to high lead compared to if *the whole population of children were* exposed to low lead (going back in time and keeping everything else the same).

Note that in the non-ideal experiment described above, you would only be able to estimate the observed difference in cognitive development score between *those children who were* exposed to high lead and *those children who were* exposed to low lead.

Problem 2. (Source: Fall 2017 250B Exam 1 Practice Exam Problem Set)
Imagine you are able to conduct a counterfactual experiment to identify the causal effect of exposure to stagnant water on malaria. You observe a population of 10 people who do not live near any stagnant water for 3 months. Over those 3 months, 2 people develop malaria. You then go back in time to the start of your study, put a pool of stagnant water outside each person's home, and observe them under the exact same conditions for 3 months. Over those 3 months, 6 people develop malaria.

a. Write a mathematical expression to represent the individual-level causal effect of stagnant water on malaria. Interpret this expression.
   Y1 – Y0
   Y1 represents an individual's outcome (malaria) when they are exposed to stagnant water. Y0 represents that same individual's outcome (malaria) when they are not exposed to stagnant water, with everything else remaining the same. The equation represents the difference in these outcomes. Remember that these are counterfactual outcomes; we do not observe each of these outcomes for each individual.

b. The counterfactual information for each of your participants is listed in the table below. (Remember that, in our counterfactual notation, "a" represents the exposure of interest and "Y" represents the outcome.) Calculate the individual-level causal effect of stagnant water on malaria for each person.

Participant	a	Y_0	a	Y_1	Individual-level causal effect
1	0	1	1	0	-1
2	0	0	1	1	1
3	0	0	1	0	0
4	0	0	1	1	1
5	0	0	1	0	0

6	0	1	1	1	0
7	0	0	1	1	1
8	0	0	1	0	0
9	0	0	1	1	1
10	0	0	1	1	1

c. Write a mathematical expression to represent the population-level causal effect of stagnant water on malaria. Calculate this population-level causal effect and interpret it.

E[Y1-Y0]
This represents the expected difference in the population's outcome if everyone were exposed to stagnant water, compared to the population's outcome if nobody were exposed to stagnant water, assuming all other conditions remain the same.
Note that this is the counterfactual parameter that the risk difference attempts to estimate.
In this case, the expectation is the mean of this difference:
[-1+1+0+1+0+0+1+0+1+1]/10 = 0.4
Since we can assume a causal interpretation, 0.4 of the malaria risk over 3 months in this population is due to stagnant water.

Problem 3. (Source: Aschengrau and Seage, Ch 15)
Indicate whether the following statements are true or false: [250A, 250B]

a. The presence of an association is indicative of a causal relationship.
   False
b. Time order is an essential attribute of a cause.
   True
c. Sir Bradford Hill's nine causal guidelines should be used as rigid criteria to establish causation.
   False
d. According to the sufficient component-causal model, blocking the action of a necessary cause will prevent all cases of a disease by all of its causal mechanisms.
   True