

## An Observational Study: Effects of Alcohol on LSAT Scores

The data below come from a study of the effect of alcohol on LSAT scores. We want to know the effect of drinking on test scores, but there is a confounding variable, grades. How does this confounder make our causal inference biased? What happens when we control for this variable?

Units	Observed LSAT Score	Drink Yes or No	Letter Grades
1	95	1	A
2	94	0	A
3	78	1	C
4	81	1	C
5	83	0	C
6	83	0	C
7	93	1	A
8	83	0	C
9	91	0	A
10	85	1	A

1. What are the IV and DV in this example? What is the mechanism?
2. What is the true causal effect? Can we know this from the above table?
3. What is the causal effect we estimate from our observations? [Hint: What is the average LSAT score for those who drink and the average LSAT score for those who don't drink. What is the difference of these two averages?]
4. What is the causal effect we estimate by 'controlling' for grades? [Hint: What is the average LSAT score for those who drink and are A students; average LSAT score for those who don't drink and are A students; the average LSAT score for those who drink and are C students; average LSAT score for those who don't drink and are C students?]

Now consider what happens if we could somehow observe both counterfactual worlds, i.e., we now know the effect of alcohol on LSAT scores *and* the effect of no alcohol on LSAT scores for all units.

Units	LSAT with alcohol	LSAT without alcohol	True Effects?
1	95	98	
2	89	94	
3	78	83	
4	81	86	
5	74	83	
6	80	83	
7	93	94	
8	74	83	
9	84	91	
10	85	90	

5. Can we know this from the above table? If yes, what is the average causal effect? [Hint: Subtract the LSAT score without alcohol from the LSAT score with alcohol for each unit. Average these differences.]

Let's randomly assign alcohol drinking the night before the test.

Units	LSAT with alcohol	LSAT without alcohol	Drinking, randomly assigned	Observed Drinking
1	95	<b>98</b>	0	
2	<b>89</b>	94	1	
3	<b>78</b>	83	1	
4	81	<b>86</b>	0	
5	74	<b>83</b>	0	
6	80	<b>83</b>	0	
7	<b>93</b>	94	1	
8	<b>74</b>	83	1	
9	84	<b>91</b>	0	
10	<b>85</b>	90	1	

6. From the above table, what is the observed effect of drinking on LSAT scores in this experiment? Is this closer to the 'true effect' from Question 5? Why is randomizing helpful here? [Hint: What is randomizing 'controlling for' here?]

