Instructions. Do this quiz in partnership with exactly one other student. Write both your names at the top of this page. Discuss the answer to the question with each other, and then write your joint answer below the question. Use the back of the page if necessary. It is fine if you overhear what other students say, because you still need to decide if they are right or wrong. You have seven minutes.

Question. Consider a linear regression model to predict the college GPA $y$ of students. The real-valued feature $x_8$ codes the binary feature “owns a car” as $x_8 = 1$ if yes and $x_8 = 0$ if no. Learning from a large training set yields the model

$$ y = \ldots - 0.4x_8 + \ldots $$

There are many other predictors in the model, including high school GPA, major, living on or off campus, etc. Dr. Webster concludes that the GPA of the average student who owns a car is 0.4 lower than that of the average student. Explain two important reasons why this conclusion is likely to be false.

Answers. (1) The correct contrast is with students who do not own cars, not with all students. (2) The average GPA difference equals the coefficient only if the two contrasted groups have the same average value for all other predictors. (3) The variable $x_8$ will be strongly correlated with the variable for living off campus. A negative coefficient for this variable will make the net impact of owning a car smaller.

Additional notes. Reason (3) is an important special case of reason (2). There may also be other valid reasons, or ways of expressing reasons. Whatever the average difference is, one cannot conclude that car ownership is the cause of the difference. However, the question says nothing about causation, so answers that mention causation are incorrect.