# STAT 113: CONFOUNDING, CAUSATION, AND EXPERIMENTS

Sports teams prefer to play in front of their own fans rather than at the opposing team's site. Having a sell-out crowd should provide even more excitement and lead to an even better performance, right? Well, consider the Oklahoma City Thunder, a National Basketball Association (NBA) team, in its second season (2008-09) after moving from Seattle. This team had a win-loss record that was actually worse for home games with a sell-out crowd (3 wins and 15 losses) than for home games without a sell-out crowd (12 wins and 11 losses). (These data were noted in the April 20, 2009 issue of Sports Illustrated in the Go Figure column.)

(1) Identify the cases and variables in this data, and whether each variable is categorical or quantitative.

(2) Calculate and compare the proportion of games won for sell-out vs non-sell-out games.

> Definition. Two variables are **associated** or related if knowing the value of one variable gives you information about the value of the other variable. When comparing groups this means that the proportions or means are different across groups.

(3) Do the two variables appear to be associated here?

Often when a study involves two associated variables, it is natural to consider one the **explanatory variable** and the other the **response variable**.

> Definition. The explanatory variable is the variable we think is "explaining" the change in the response variable and the response variable is the variable we think is being impacted or changed by the explanatory variable. The explanatory variable is sometimes called the independent variable and the response variable is sometimes called the dependent variable.

(4) Which would you consider the explanatory variable in this study? Which is the response? (That is, what are the roles of these variables in this study?)

There are three possible explanations for this odd finding that the team had a better winning percentage with a smaller crowd:

- The sell-out crowd caused the Thunder to play worse, perhaps because of pressure or nervousness.

- The sell-out crowd did not cause a worse performance, and some other issue (variable) explains why they had a worse winning percentage with a sell-out crowd. In other words, a third variable is at play, which is related to both the crowd size and the game outcome.

- The difference in this set of games is just due to random chance.

Using methods you will learn later, it will turn out that we can essentially rule out random chance in this case.

(5) Consider the second explanation. Suggest a plausible alternative variable that would explain why the team would be less likely to win in front of a sell-out crowd than in front of a smaller crowd. (Make sure it's clear not just that your explanation would affect the team's likelihood of winning, but that the team would be less likely to win in front of a sell-out crowd compared to a smaller crowd.)

> Definition. A **confounding variable** is a variable that is related both to the explanatory and to the response variable in such a way that its effects on the response variable cannot be separated from the effects of the explanatory variable.

(6) Identify the confounding variable based on your suggested explanation in (5). Explain how it is confounding – what is the link between this third variable and the response variable, and what is the link between this third variable and the explanatory variable? (Hint: Remember that this variable has to be recorded on the cases: home games for the Thunder.)

Another variable recorded for these data was whether or not the opponent had a winning record the previous season. Of the Thunder's 41 home games, 22 were against teams that won more than half of their games. Let's refer to those 22 teams as strong opponents. Of these 22 games, 13 were sell-outs. Of the 19 games against opponents that won less than half of their games that season (weak opponents), only 5 of those games were sell-outs.

(7) Was the Thunder more likely to have a sell-out crowd against a strong opponent or a weak opponent? Calculate the relevant proportions to support your answer.

When the Thunder played a strong opponent, they won only 4 of 22 games. When they played a weak opponent, the Thunder won 11 of 19 games.

(8) Was the Thunder less likely to win against a strong opponent than a weak one? Again calculate the relevant proportions to support your answer.

(9) Explain how your answers to (7) and (8) establish that strength of opponent is a confounding variable that prevents drawing a cause-and-effect conclusion between crowd size and game outcome.

(10) Summarize your conclusion about whether these data provide evidence that a sell-out crowd caused the Thunder to play worse. Write as if to a friend who has never studied statistics. Be sure to address the fact that the Thunder had a much smaller winning percentage in front of a sell-out crowd.