

STAT 113: CONFOUNDING AND CAUSATION (PART 1)

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 at home that was actually **worse for games with a sell-out crowd** (3 wins and 15 losses) **than for games without a sell-out crowd** (12 wins and 11 losses). This data was reported in the April 20, 2009 issue of Sports Illustrated in the Go Figure column.

1. Identify the **cases** in this data.
2. Identify the two **variables** referenced in this data. For each, identify whether it is **categorical or quantitative**.

Variable	Type (Categorical or Quantitative)

3. What **proportion of sell-out games** did the Thunder win?
4. What **proportion of non-sell-out games** did they win?
5. What is the **difference in the Thunder's win proportion** between the two types of game?

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, because knowing what group a case provides information about the other variable.

6. 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 **response variable** in a study is the variable whose variation we are interested in accounting for. An **explanatory variable** is a variable we can use to help us predict the response variable. Differences in the explanatory variable can contribute to differences in the response variable (though not all associations are causal). The explanatory variable is sometimes called the **independent variable** and the response variable is sometimes called the **dependent variable**, though these terms are usually used for experimental studies where the explanatory variable is being manipulated directly by the researchers.

7. What would you consider to be the **explanatory variable** in the description of the Thunder's record?
8. What would you consider to be the **response variable**?
9. Give a possible explanation for the association between the crowd-size variable and the win-loss variable in which crowd-size is **influencing** the outcome of games (that is, where the association is **causal** in nature)
10. Give an alternative explanation for the association in which crowd-size is **not influencing** the outcome, either directly or indirectly (that is, where there is not a causal chain linking the two)