STAT 113: CONFOUNDING AND CAUSATION (PART 2)

Here are three possible explanations for the potentially odd finding that the Thunder 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 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 for 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.

1. 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 that this variable would be associated with **both the team's likelihood of winning and the likelihood that the crowd is a sell-out**. (Hint: Remember that a variable is something that can be recorded for **each case**: home games for the Thunder.)

Definition. A **confounding variable** is a variable that is related to **both** the explanatory and response variables in such a way that the **explanatory and response would be associated even if the former were not directly influencing the latter**. The confound typically is something that may have an influence on the response and either also has an influence on the explanatory, or it has an association with the explanatory which is not easily described as causal in nature. If it is **influenced by** the explanatory variable and **influences** the response, it is not really a confounding variable, but rather a **mediating variable**: it **elaborates on an (indirect) causal relationship** between explanatory and response, rather than offering an explanation for the association which is an alternative to causality.

2. **Does the variable you identified** in 1 **meet the above definition** of a confounding variable? **If not, why not?** If so, **what is the nature of its proposed effect on the response variable, and what its proposed association with the explanatory variable**?