Causal Inference Project Proposal (097400)

Yoav Meiri - 212724462
Ziv Tamir - 324276054

January 2023

1 Background

The NBA draft is an annual event in which the teams from the American National Basketball Association (NBA) can draft players who are eligible and wish to join the league.

This process consists of 2 rounds, we will focus on the first one. In this round each team selects one of the nominees and offers him a 2 year contract with another optional 2 years. Usually teams that pick in later in the draft are the better teams, so the worse teams will have a chance to pick more valuable players.

2 Description of the causal question

It is well known that the early career of a NBA player is highly influenced by his placement in the draft picks. We wish to investigate the relationship between the draft placements and the average salary of a player in his first 5 years.

But, we cannot ignore the fact that the draft placement and the early career salary are both highly influenced by the personal talent of the player. Given that, it might be irrelevant to ask: what could have happened to a player that was picked 1st, if he would have been picked 30th. So, we will limit our treatment to the following:

Treated: player was picked between 13th-15th
Non-treated: player was picked between 16th-18th

We assume players who were picked in the middle picks have similar 'talent level' and were might as well picked in any of the other ranks of the same range. In practice, the placement is highly influenced by the current needs of the selecting teams (which are also assumed to be at a similar level in the middle ranks).

3 Data

For this project we will combine 2 datasets (taken from this source). One is describing NBA players and their personal statistics which include draft placement, draft year, college, picking draft team, multiple physical stats, position, birthplace and a lot more.

The other one contains player salaries throughout their career.

One challenge we face with this data is that there is little to no data about the teams. The resources and character of the selecting teams have big influence on both our treatment and outcome. We plan to apply inference tools on the existing data to take the teams characteristics into account, or even gather more data if necessary.

4 Scope and methods

We plan to apply 2 main analysis methods:

- Potential outcomes: we plan to estimate the treatment effect (ATE estimation) with counterfactual analysis using covariate adjustment (for example, we think the matching method will be appropriate here).

- Do-calculus: we will use our world knowledge of relationships between the different covariates to create a causal graph which we will use to estimate the treatment effect. We will do it using the do-calculus tools.

5 Possible challenges and issues

In unprofessional eyes it might be difficult to understand the connections between the different covariates. We can learn more about the relationships between the multiple covariates using statistical analysis tools.