Discovering the effect of hero choice on the outcome of a Dota 2 game

Noyan Eren Toksoy
n.e.toksoy@student.tudelft.nl
Supervisors: Richard Karlsson, Jesse Krijthe

1. Background

- Dota 2 is one of the most popular MOBA[1] games.
- A match is played between 2 teams of 5 players and the opposing team's Ancient tower must be destroyed to win the game.
- An essential part of a match is the hero selection phase, where players can take turns picking and banning heroes from a pool of 123 heroes.
- Causal inference[2] can be used to estimate the **causal** effect of a hero's selection on the outcome of a match.
- This effect could have been observed by calculating[3] [1]:
  $E[Y^{t=1}] - E[Y^{t=0}] = \text{causation measure}$
  But in real-world scenarios, this requires counterfactual outcomes. By analysing historical data, the following measurement can be made [1]:
  $E[Y|T = 1] - E[Y|T = 0] = \text{association measure}$
- However, in the presence of confounding factors[4], association != causation.

2. Research question

- *What is the effect of the Pudge hero being picked in a team on the outcome of a Dota 2 game?*

3. Methodology

- All confouding factors, needed for the causal analysis of this research, can be found in the causal diagram in Figure 1.

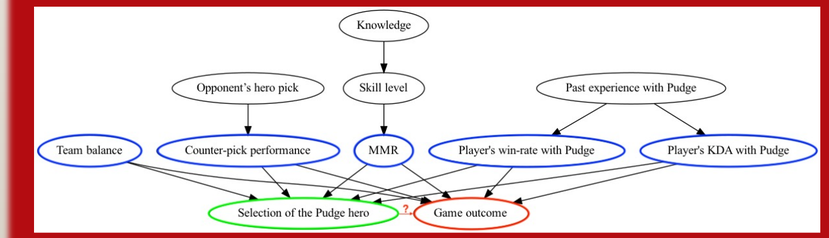

Figure 1: The g-formula used to estimate the counterfactual outcomes. L refers to the confounding factors and Pr(x) means probability of x.

- MMR (Matchmaking rating) is one of the confounding factors and it affects the selection of the Pudge hero because as the MMR of a player decreases, it is more likely to pick more popular and easier to play heroes. It also affects the outcome because a high MMR is associated with high skill level, which is known to affect the outcome.
- There are existing methods that can be used to correct for these confounding factors by estimating the counterfactual outcomes mentioned earlier, g-formula is one of them.
- The g-formula makes use of regression and standardisation, and its full form can be seen in Figure 2 [1]. With these estimations, the causal effect can be measured.

$$E[Y^t] = \sum_l E[Y|T = t, L = l] \times \Pr(L = l)$$

Figure 2: The g-formula used to estimate the counterfactual outcomes. L refers to the confounding factors and Pr(x) means probability of x.

4. Results

- The Average Treatmen Effect (ATE), a numerical measure of causal effect, was calculated as **-0.2848%** when corrected for all the confounding factors, as can be seen in Table 1.

	Average Treatment Effect (%)
Adjusted for all confounding factors	-0.2848
Without adjusting for the confounding factors	-5.4006

Table 1: A table with the obtained results from the g-formula.

- The collective, and individual, causal effect of all the confounding factors are displayed in Table 2. The individual effects were calculated by:

$$ATE(x) - ATE(y),\ \text{x} = \text{Adjusted for single confounding factor, y} = \text{No adjustment.}$$

- The 95% confidence interval for the ATE was calculated as [-3.3936, 2.2824], using bootstrapping with 2000 re-sampling iterations. Figure 3 illustrates the probability density function of the ATEs obtained from all the iterations.

	Confounding Effect (%)
All confounding factors	5.1158
Pudge win rate	0.8691
Pudge KDA	0.6127
MMR	4.5007
Counter-pick performance	0.0057
Team balance	0.0194

Table 2: A table with the confounding effect of each factor, as well as the combined effect of all confounding factors.

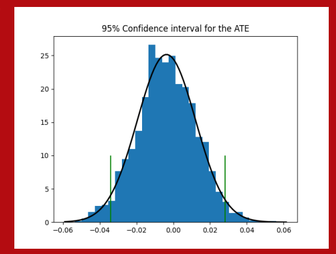

Figure 3: The probability density function of the ATE list obtained by bootstrapping. The green vertical lines indicate the start and end of the 95% confidence interval (-0.033936, 0.028240).

5. Conclusion

- On average, the teams in which the Pudge hero was selected, had a 0.2848% less chance of winning the game compared to the teams that did not select Pudge.
- This is a relatively small difference for the game of Dota 2, meaning that the hero does not have a significant advantage or disadvantage over others.
- A 5.1158% effect introduced by the confounding factors was corrected for using the g-formula.
- It was observed that the confounding factor with the highest effect is MMR.

6. Limitations & Future Work

- Most of the limitations encountered were from the data gathering process, related to the OpenDota API that is used.
- Compromises had to be made in the final dataset in terms of the number of games collected and their recency in order to assure accuracy.
- A possible extension of this research would be to discover the causal effect of other heroes, which can be used to make generalizations on the hero selection phase's effect on the game outcome.

[1] Multiplayer online battle arena
[2] Causal inference is a discipline that is concerned with discovering causal relationships between variables using data analysis under certain assumptions about the data [1].
[3] E[x] refers to the mean or the expected value of x. T refers to whether the treatment (selection of the Pudge hero) is applied or not. Y is the outcome. Y^(t=0) refers to the outcome that would have been observed, if the treatment had value t.
[4] Factors that affect both the treatment and the outcome

References

- [1] M A Hernán and J M Robins. Causal Inference: What If. Chapman & Hall/CRC, 2020.