Analysis of the difference between official appearance rate and player experience in Arknights

Tong Li

2020-12-22

Abstract

The game experience of players is a very important aspect for game companies, however, the uncertain factor of operator pulling probability will being affected game experience of players. Pulling special banner in the Arknights always has the player to report spending a lot of resources still cannot get the operator that wants. This analysis is based on the actual extraction experience from the video Publisher to analyze whether the feelings of players are true. Are this survivor bias or be it truly different from the officially reported probability? What is the factors that affect the probability? After script simulation and multiple Linear Regression, the problems will discuss based on the results.

Keywords

Observational study, multiple linear regression model, Causal Inference, "Gacha" system, Arknights

Introduction

Arknights is a free mobile game that contains banner pulling mechanism, players can get game resources through banner pulling (game resources will be referred to as the operator below). The number of operators increases over time, and there is a rarity rating. The more star of an operator has, the harder it is to gain (3 stars — 6 stars in the game). Banner can be divided into ordinary banner and special banner. All banners will have mechanism of increasing probability of special operators, and limit open time. A new operator added to an ordinary banner is still withdrawable after banner rotation*, but the limited banner contains a 6-star limited operator which can no longer be obtained after the banner rotation. This mechanism makes players want to gain the limit operator, but the actual experience is inconsistent with the officially announced pulling probability. Players spend a lot of resources but cannot get restricted operator, this gap caused dissatisfaction. And from the game open service to now, there are a totals of three limited card pool, every time will appear the same phenomenon. A number of players gave videos of themselves drawing operators to show that their investment was not commensurate with the return. The game experience of players affects the operation of the whole game company.

This analysis will be a two-sided analysis aim at Arknights, both from official appearance rate that game company given and player's experience, to see whether the official appearance rate is matching with player's experience or not. If yes, then survivorship bias occurs, players only noticed the voices of a small number of people who did not gain; if not, game company is publishing unrealistic appearance rate. It will calm the dissatisfaction of many players and help clear the name of the game company. In this study, a script will be used base on official banner pulling probability (by python). It is showing chance of getting at least one limited operator. And r/r studio will be used to find out the relation between account level and probability of gain limited operator. Besides, real data was obtained from the hands of the players for analysis, and different predictor (number of pull to get at least one limited operator, account level, pull from which banner,

total number of pulling 6 stars operators and number of pulling limited operator) will used for analysis using the multiple linear regression model, to see whether the players' comprehensive acquisition probability has linear relation between account level, since there is a statement that company will reduce the probability of high-level accounts. The specific method is to use same script but player's experiment probability to calculate at what time of pulls, player will get at least one limited operator.

Methodology section would contain process of data collection and explains on each variable that use in multiple linear regression model. Result would post on result section, further discussion about summary and conclusion of result, weakness and next step would conclude on Discussion part. Bibliography is in Reference part.

Methodology

Data

Data was collected manual from video publisher that specialized in video of game. I chose the video with more banner pulling times as the data, video with target of extract multiple times of limited operator is the best choice. Data contains 11 variables: _id: unique identifier. player_id: unique id of each account given by server. num_total: total time counts of pulled out from banner. num_total_6op: total number of 6 stars operator gain from banner. num_limit_op: total number of limit operator gain from banner. num_up6: total number of appearance rate up 6 stars operator gain from banner. banner: limit operator's code name, a differentiate of each special banner. server_region: the server at different region, CN means mainland China,WW means other region. Observed by language of the game interface. account_level: the level of account that accounted. date: date of video uploaded.

All data was collected either by eye watching each video or summary by publisher at the end of video.

On one hand, the official appearance rates for three banner are as follows:

3 stars Operators have a 40% chance to appear in a pull. 4 stars Operators have a 50% chance to appear in a pull. 5 stars Operators have a 8% chance to appear in a pull. 6 stars Operators have a 2% chance to appear in a pull. For special banner, appearance rate up operators (one is limited, one is not) accounts for 70% of the odds when pulling a 6 stars. If a 6 stars Operator does not appear after 50 pulls, each subsequent pull will increase the 6 stars Operators' rate by 2%, up to 100%.

On the other hand, in multiple linear regression model, total number of pulling, total number of 6 stars operator gain, total number of appearance rate up 6 stars operator, total number of limited operator gain and account level are predictor. After remove missing values, 81 remain observations are available to access analysis. During cleaning process, probability of getting a 6 stars operator and probability of getting a limited operator are added into data by doing simple division method. It is because the analysis is focusing on probability, not specific occurrence time number.

Model

By fitting multiple linear regression model and observing p-value, - | Estimate | P-Value | ——|:————:|:————:| Intercept | 135.7553 | 0.00239 | account_level | 0.3297 | 0.45024 | bannerRosmontis| -12.1299 | 0.49078 | bannerW | 3.2245 | 0.84271 | bannerW | 7.4650 | 0.90233 | num_total_6op | 15.9935 | <0.0001 | num_limit_op | -95.7584 | <0.0001 |

Only p-value of intercept, number of total gain 6 stars operators and number of limited operators get are smaller than 0.05, which is significant. Others has no evidence against null hypothesis, $\beta = 0$, so model would be:

$$\hat{y} = 135.76 + 15.99x_{total6stars} - 100.67x_{num_limit_op}$$

Holding other variable unchanged, as one 6 stars operator get, number of pulling to get at least one limited operator will increase around 16 times. Also holding other variable unchanged, as one limited operator get, number of pulling to get at least one limited operator will decrease around 100 times.

Results

Residual plot of pulling time of one limited and Account level

Fitted values
lm(pull_get_lim ~ account_level)
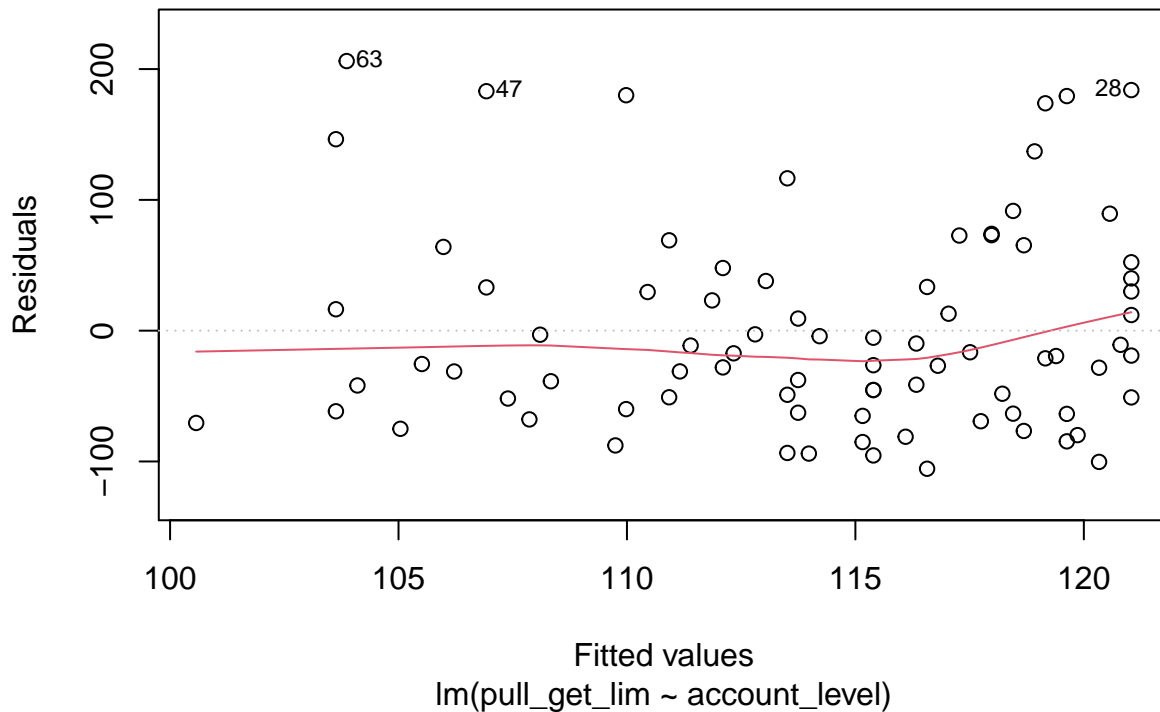
In Model section, account level is not significant with p-value that greater than 0.05. Residual vs. Fitted plot between account level and pulling time of getting at least one limited operator shows a almost constant mean (red line). Thus, there is no linear relationship between account level and pulling time of at least one limited operator. Also no casual relation between account level and rate of get limited operator, since pulling time has a strong negative linear relation between appearance rate. The lower rate, the more pulling time.

In script, scatter plot is showing the chance of getting at least one limited operator from 0% to 99% in range 0
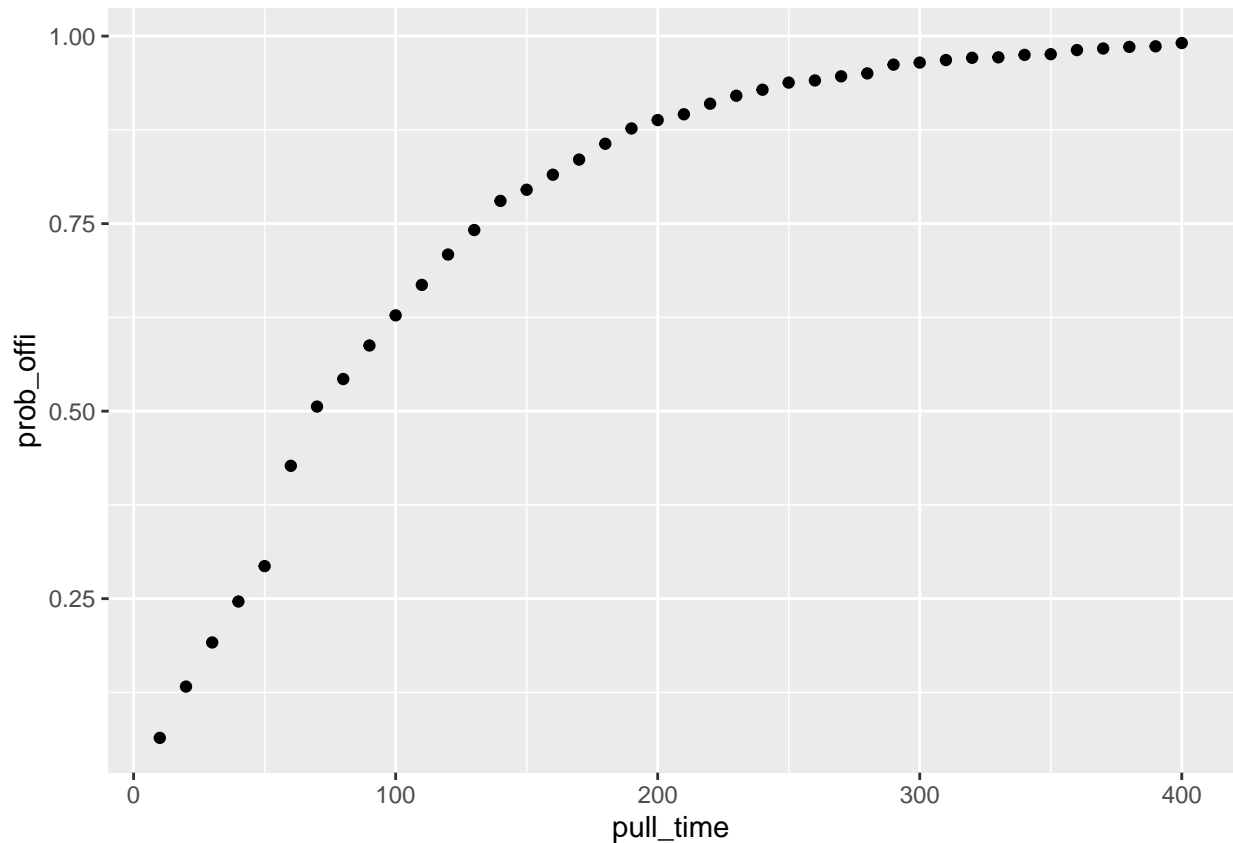
pull to 400 pull.

At the beginning, probability rises rapidly, and gradually flattens out after more than 75%. There is a gap between 50 - 60 times pull, and outstrip 50% at 70 times pulling.

Based on player's experiment, appearance rate up operators accounts for 65.7% of the odds when pulling a 6 stars. 6 stars Operators have a 3.4% chance to appear in a pull. After plug in the new rate found from player, scatter plot becomes below in range 0 pull to 330 pulls to reach 99% chance:

```
## Warning: Removed 7 rows containing missing values (geom_point).
```
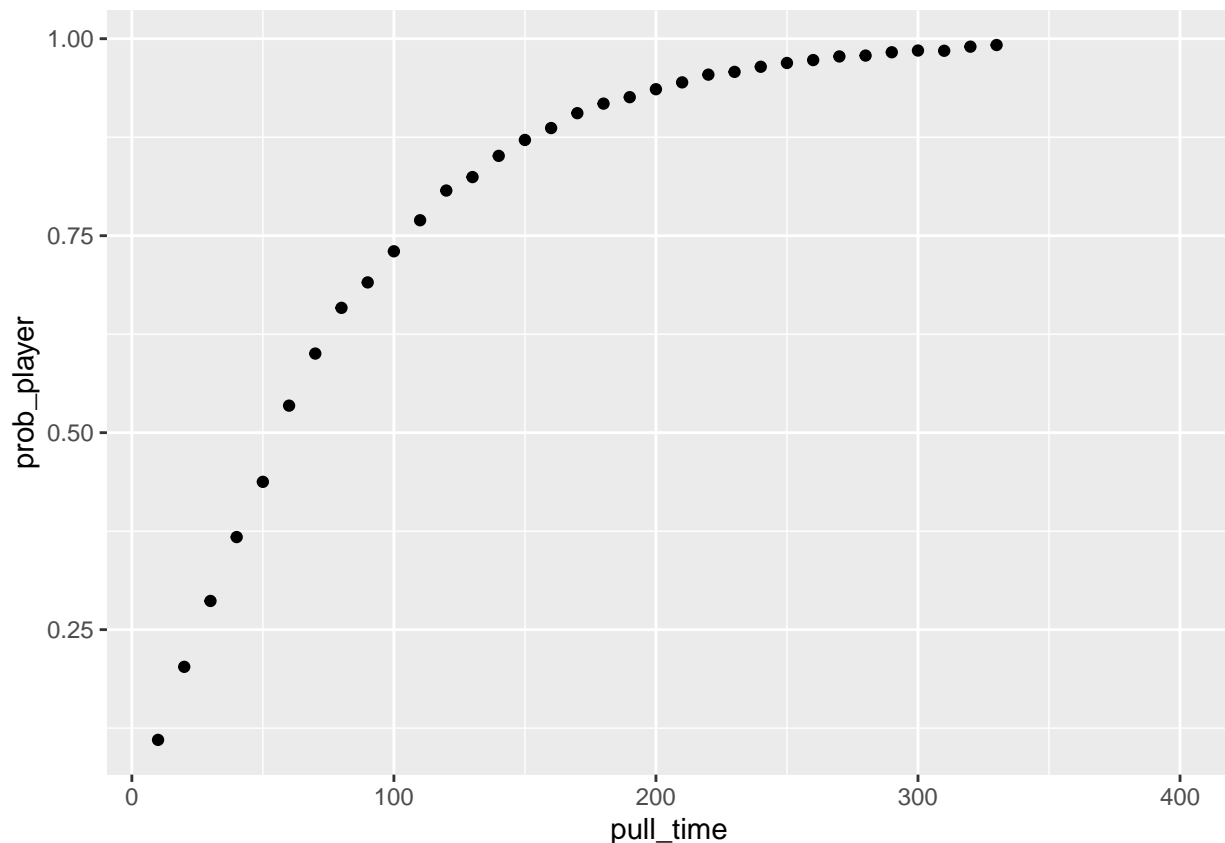
It has same distribution as above with smaller range, smoother distribution. Outstrip 50% at 60 times pulling.

Discussion

Overall, Conclusions can be drawn after collecting data, cleaning data, establishing multiple Linear Regression model, using script, and drawing plots. Now what we know is there is no casual inference between account level and pulling time for at least one limited operator, chance for player to get one limited operator has no huge difference between official announced.

There is indeed survivor bias, and players pay too much attention to the information that can't pull limited operators. Of course, only a small group of those who have operator will show that they have, but those who haven't will get together and discuss it under a post. Analysis shows player's data acquisition rates are even higher than the official rates. With the same 99% gain rate, the official rates pulling takes 400, and the player only needs 330. But the line of 50% has no huge difference, since both are 10 pulls for one count, 60 and 70 have not that big difference. Company will give 10 pulls card each time after special banner is opened to fill up the 10 pulls difference and make player's cost less.

Weaknesses and Next Steps

First, data size is small for a game probability analysis, there are only 81 data available after cleaning, which is far from enough compared to the huge number of players, that is why I have to get the probability and then use script to come to the conclusion. Normally, percentage of players after a fixed pulling number who get limited operator, and keep increase pulling number to find distribution. Second, 10 pulls for one count is rough rather than 1 pulls one count. Third, data is obtained entirely by watching video, and there may be unknown factors such as editing that influence the final result. This may also be the reason why player probability is higher than official. Fourth, game include 3 special banner in CN, but only one in WW. But I

put all them together to analysis, separate them could be a batter choice. Also for 3 banner in CN, game company might make some changes in the next two pools. Therefore, this analysis is very one-sided and can only be used for reference

In the future, the database can be gradually enriched through other channels, and the time factor can be added to explore whether the game's card pool mechanism will lead to an increase in the draw probability when the card pool is refreshed every day.

References

1. official appearance rates - https://mrfz.fandom.com/wiki/Headhunting
2. WishIHadAComputer, script of chance of getting limited operator. 2020. - https://www.reddit. com/r/arknights/comments/hz23mf/here_are_the_chances_of_you_ getting_nian_in_x/ 3.script by WishIHadAComputer, 2020.- https://www.onlinegdb.com/rkswEC2xv