Examining the Relationship Between the Use of Official Anime Viewing Platforms and Viewer Characteristics

Kexin Qin

12/21/2020

Code supporting this analysis is available at: https://github.com/qinkexinnn/Anime_viewer_characteristics

Abstract

In this report, I will explore the relationship between use of official anime viewing platforms and the characteristics of an anime viewer. The 2020 Anime Survey (Reysen et al., 2020) will be analyzed for the probability of a anime viewer using official anime viewing platforms based on their age, whether or not they collect anime related merchandise, and their preference in anime. The results showed that an increase in age and collecting anime related merchandise increases one's log likelihood of using an official anime viewing platform to watch anime. Identifying which type of anime fans will use official platforms is important because it is one of the ways production companies make revenue.

Keywords

Logistic regression, Observational Study, Viewer Characteristics, Anime Fandom, Streaming Services

Introduction

Along with the increasing popularity of streaming services, large amounts of copyrighted anime episodes are also being shared online via unoffcial downloads and unoffical streaming services. The former supports the production of anime, and the latter does not (Blair, 2017). Anime production studios make profit from the sales of DVDs, merchandise, and sales of licensing rights to distributors. Therefore, it is important to identify the type of viewers that will use an official platform to view anime, since the production companies make revenue through these offical platforms. Identifying the type of anime fans who will pay for content can help production companies better understand their audience, and possibly target them for future productions.

In this report I will model whether or not a anime viewer watches their anime through official platforms based on their age, whether or not they purchase offcial anime related merchandise, and the type of anime they watch most frequently. One way to model a binary dependant variable is through logistic regression. Logistic regression was first used in early twentieth century and employs a logistic function to model a binary dependent variable (Swaminathan, 2018). In this report, logistic regression will be used as the reponse only has two outcomes: one either uses offcial platforms or not.

Data from the 2020 Anime Survey will be used investigate the relationship between anime viewer characteristics and whether they use official platforms as their main source of anime consumption. In the Methodology section (Section 2), I describe the survey, the data, the cleaning process, and the model used in more details. Results of the logistic regression along with interpretations are in the Results section (Section 3). In the Discussion section (Section 4), I will address some weaknesses of the analysis and why ways to make casual inferences were not used. I will also address future steps of the analysis.

Section 2: Methodology

Data

The data comes from an observational study. The survey data is from Anime Survey 2020, a research conducted by the International Anime Research Project to study the anime fandom (Reysen et al., 2020). It was posted as a survey and solicited volunteers from various anime-related websites. The target population

of the survey is everyone who watches anime. The frame is people who uses anime-related websites and the sample is all the respondants of the survey. Participants who did not complete the majority of the survey were removed(Reysen et al., 2020).

The dataset has its own strength and weaknesses. The strengths also include having a large number of topics, allowing for a lot of flexibility with the analysis. With responses coming from 89 different countries, the dataset has a lot of respondants coming from different backgrounds. However, the data also has weaknesses. By the fact that the data was collected through a survey, it is prone to response bias. Since the survey was posted on anime-related websites, we are missing information on the anime viewers who do not use these websites. Lastly, there might also be communication failings as different respondants might interpret different questions in different ways which can affect their answers.

The survey had a list of consumption source such as TV programming, DVD, legal streaming services, unofficial downloads and unoffcial streaming services. The participant rated the frequency they used these platforms from 1 (never) to 7(often). I combined these variables and make a binary variable: whether or not the participant used official more frequently. If the respondent's response for using unoffical downloads/streaming services added up to be less than their use of legal streaming services/DVD/TV programming, they will be considered as someones who uses offcial platforms to watch anime. A similar cleaning process was also used to create the variable "purchase", a binary variable specifying whether or not the participant collects official anime related merchandise. If their response was 5 or greater (on a scale of 1 being never to 7 being often), they will be considered as someone who purchases anime related merchandise.

I also added new variables that determined the type of anime the participant prefered. The survey question asked the participant to rate the frequency they consume anime from different decades. If their frequency for watching 1960s anime and 1970s anime is greater than the frequency they watch 1980s-1990s and 2000s-2010s anime, they will be considered as someone who mostly watches 1960s and 1970s anime. The other variables for watching 1980s-1990s and 2000s-2010s anime are created using the same process. If the frequency is equal, then they will be considered as not having a prefence for anime from a specific time period.

This dataset has many other variables, but I will be mainly focusing on the variables that I cleaned above. A baseline characteristics table with the cleaned data is shown below with the mean of each group presented. For age, it is the average age of all survey respondants. For other variables, the numbers are shown as proportions. These variables will the discussed and explored further in the next sections.

Table 1: Baseline Characteristics of Survey Respondants

	Mean of each group (n=3385)
Age	23.298
Watches anime through official platforms	0.926
Knows alot about how anime is made	0.255
Collects offical anime-related merchandise	0.253
Mostly watches 1960s and 1970s anime	0.003
Mostly watches 1980s and 1990s anime	0.056
Mostly watches 2000s and 2010s anime	0.881
No preferece for anime from a specific time period	0.061

Model

For this report, we are interested in whether or not an anime fan will use official platforms to watch anime. A logistics regression model was selected because the our response variable is binary: one either uses official platforms, or they do not. The model is ran by Rstudio. We will be using the anime fan's age, whether or not they purchase offcial anime related merchanse, and their preference for anime for our model. Age is recorded as a numeric variable, the purchase of anime related merchandise is binary, and their preference for anime has been divided into 3 categories: preference for anime from 1960s-1970s, 1980s-1990s, 2000s-2010s.

$$\log(\frac{p_{offical}}{1 - p_{official}}) = \beta_0 + \beta_1 x_{age} + \beta_2 x_{purchase} + \beta_3 x_{decade1} + \beta_4 x_{decade2} + \beta_5 x_{decade3}$$

$p_{official}$ represents the probability that the anime fan watches anime on a official platform. β_0 is the intercept of the model, which is the probability that an anime fan who does not purchase anime related merchandise, no preference for anime from a specific time period with age 0 would watch anime of a official platform. This intercept has no practical interpretation as babies of age 0 do not watch anime.

β_1 represents the change in log odds of watching anime of a official platform for one year increase in age. Age was selected as a variable since millennials prefer online streaming over traditional television (Fitzgerald, 2018). Therefore we want to see how age plays a role in whether their online streaming is official, or unofficial. β_2 is the change in log odds for watching anime on a official platform if the anime fan purchases official anime related merchandise. Since official merchandise often include DVD and blue rays, this could be a good indicator of whether a person will prefer watching anime through official sources as they may already have purchased a DVD through their purchase of merchandise. Lastly, $\beta_3, \beta_4, \beta_5$ represents the change in log odds of watching anime on a official platform given one's preference for anime from 1960s-1970s, 1980s-1990s, or 2000s-2010s. These variables were selected because popular streaming platforms like Netflix do not carry old anime, and have a preference for newer anime (Donaldson, 2018).

There are weaknesses to this model. This model does not have a finite population correction incorporated, and methods of making causal inferences were not employed either. Weaknesses will be discussed in details in Section 4.

Section 3: Results

Table 2: Summary of Logistic Regression Results

	Estimate	Std. Error	Z-value	P-value
Intercept(Beta0)	2.58167	0.45479	5.677	1.37e-08
Age(Beta1)	0.0527	0.01356	3.887	0.000102
Collects offical anime-related merchandise(Beta2)	0.4921	0.17523	2.808	0.00498
Mostly watches 1960s-1970s anime(Beta3)	11.03855	292.18	0.038	0.969863
Mostly watches 1980s-1990s anime(Beta4)	-1.12144	0.49165	-2.281	0.02255
Mostly watches 2000s-2010s anime(Beta5)	-0.86889	0.42157	-2.061	0.039295

The table above shows summary of the logistic regression coefficients. Plugging in our estimates into our model described previously , our equation is

$$\log(\frac{p_{offical}}{1 - p_{official}}) = 2.25861 + 0.0527 x_{age} + 0.4921 x_{purchase} + 11.03855 x_{decade1} - 1.12144 x_{decade2} - 0.86889 x_{decade3}$$

As we see in our model results, age is a significant indicator of one's probability of watching anime through an official platform. It has a positive value, which means that the log likelihood of one using an official anime viewing platform increases as age increases. Whether or not one collects offial anime merchandise is also significant, meaning that the log likelihood of someone using offcial anime viewing platform increases if they collect offcial anime related merchandise. As the coefficients for β_4 and β_5 are negative, it means that a preference in anime from 1980s-1990s or 2000s-2010s will decrease the log likelihood of an anime fan using official anime viewing platforms.

Section 4: Discussion

Summary and Conclusions

In the interest of identifying traits of anime fans who will watch anime on official viewing platforms, we cleaned a survey data for our variables of interest, and built a logistic regression model with response variables

age, whether or not one collects anime related merchandise, and preference in anime.

According to the model, we can expect a older person who has collects anime related merchandise to have a higher probability of using official anime viewing platforms than someone who is younger and does not collect merchandise. Furthermore, we discovered that the purchase of anime related merchandise is the most influential indicator of an increase in the log likelihood of an anime fan using offcial anime viewing platform.

Weakness

There are weaknesses in both the data and the model. Being collected through a survey, the data is prone to response bias and response error. The sampling frame also had its flaws as it only included anime fans who used anime related websites and forums. This could potentially lead to bias as well.

There was also no data on the total population of anime fans, and the sample was collected based on individual who volunteered to participate, so it was not possible to incorporate finite population correction into our model. Another weakness of the model is that we were not able to make causal inferences. We were only able to talk about correlation between the reponse variables and the predictor varitions. This means that we cannot necessarily say that an increase in age is what caused the increase of the likelihood of using an official anime viewing platform, we can only say that they are correlated.

Next Steps

As mentioned in the weakness section, we were not able to make causal inferences with our model. Since we found that purchase of anime related merchandise was significant, a good next steps would be working towards making causual inferences potentially through propensity score matching. We can set purchase of anime related goods as our treatment, and the likelihood of using offcial anime viewing platform being our outcome of interest. Propensity score matching it will be for the purchase of merchandise propensity. This way, we will be able to make more meaningful analysis that include casuality.

References

Blair, G. (2017). How Netflix Is Disrupting, Empowering Japan's Anime Industry. Retrieved from https://www.hollywoodreporter.com/news/how-netflix-is-disrupting-empowering-japans-anime-industry-1050297

Donaldson, K. (2018). Every Anime Available to Watch on Netflix. Retrieved from https://screenrant.com/netflix-anime-best/

Fitzgerald, T. (2018). Wow: Millennials Watch More Online Video Than Traditional Television. Retrieved from https://www.forbes.com/sites/tonifitzgerald/2018/11/28/wow-millennials-watch-more-online-video-than-traditional-television/?sh=715b59d94138

Hao, Z. (2020). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.1. Retrieved from: https://CRAN.R-project.org/package=kableExtra

Reysen, S., Katzarska-Miller, I., Plante, C. N., Roberts, S. E., Gerbasi, K. C., Brooks, T. R., & Tague, A. (2020). Anime and global citizenship identification. The Phoenix Papers, 4(2), 48-61.

Swaminathan, S. (2018). Logistic Regression — Detailed Overview. Retrieved from https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686