

Statistics Project and Dissertation STATS5029P

**#34: Does playing Pokémon Go increase physical activity?**

Hoi Him Kwok (2640948k)

MSc in Data Analytics

06/10/2021

# Content

**1. Abstract1**

**2. Introduction2**

2.1 Background Information2

2.2 Research Objectives3

**3. Data Description and Processing4**

**4. Methodology5**

**5. Analysis Results6**

**6. Discussion and Future Works7**

**7. Conclusion8**

**8. References8**

# 1. Abstract

# 2. Introduction

## 2.1 Background Information

It is well known that sport is essential for individual’s physical health. With sport, we can train our muscles and respiratory, as well as enhancing immunity (Ornulf Seippel. 2006). In accordance with Buraimo, Jones and Millward (2011), approximately half of the people did not participate in any kinds of sports based on several reasons, including job character, change of entertainment way and long working period. In this way, the risk of obesity, depression, high blood pressures, as well as a series of cardiovascular diseases increased rapidly, being harmful to public health. For solving health issues, augmented reality (AR) videos games are in consideration. Augmented reality is an extended version of VR. It combines virtual elements with real world, through the assist of visual devices, including eyeglasses, monitor, as well as smart devices (Tim Fisher. 2021). Overlapping on and tracking in real world objects, AR objects seems occupying the same space. Besides visualization, AR system can also contain sound and tactile, for providing an new form of world. AR technology can be applied in various types of applications, such as map and games. In this study, we focus on Pokémon Go, a popular AR mobile games.

Pokémon Go is a famous mobile game developed by Niantic Inc., as well as released at 2016 on both Apple Store and Google Play (Luke Reilly. 2017). Pokémon Go players using GPS signals to locate, catch (also obtain ingredients for training Pokémon), hatch (players walk around 2 to 10 km to obtain a Pokémon) and train virtual creatures, Pokémon. Those Pokémon can be used for battling and Gym controlling (Andrew Webster. 2015). Pokémon Go use a map and camera to display the virtual spots, like Pokestop, Gym and activity location points. (Smith. 2017) Pokémon Go is a celebrated application, with 632 million times of downloading and 147 millions of monthly active players. Due to the popularity, Pokémon Go plausibly facilitating an obvious behavior change in public health (Dillet. 2016). If relationships between Pokémon Go and amount of physical activity is confirmed, a new method for increasing the sport rate and improving the public health was discovered.

According to multiple researches, Pokemon Go obviously and positively affect the amount of physical activities. The amount of physical activities increased approximately 25%, comparing with previous activity level. Pokemon Go, additionally, rising the physical activity level across gender, ages, as well as weight status. The physical activityk level of player, inactive originally, increase sharply in general (Gunther Eysenbach. 2016). Some studies, however, suggested that Pokemon Go cannot directly advance the public physical health. Despite the best effects in first period, players’ physical activity level drop sharply, meaning that the positive effects mentioned is not sustainable (Allana LeBlanc et al., 2016). This effect can be related to the motivation of players. There are three kinds of motivation for players, including health, social and immersion. Although players with health motivation presented the significant increase of the amount of physical activities, the effects caused by social and immersion motivation is limited. Merely the time spending outdoor increased rapidly (Lukas Dominik et al., 2017). In this situation, a research, studying the relationships between Pokemon Go and physical activity, was conducted. If Pokemon Go can increase amount of physical activity, we can develop more AR mobile games for public health improvement. The aspects of the entire study were indicated in “Research Objectives” session.

## 2.2 Research Objectives

It is commonly knowing that Pokémon Go was not designed for public health improvement. That is the reason of the confusion about the relation between Pokémon Go and physical activity. For truth discovering, the entire study was conducted based on four aspects. The relation between frequency of app usage and amount of app usage, firstly, is a main focus. Theoretically, the more the app usage, the higher the opportunity for players doing physical activity (e.g. catching Pokémon, or turning Pokestop). Following the application usage aspect, we also discuss the problem in players’ characteristic. Since some researches proved that Pokémon Go players tend to join game-related physical activity, instead of physical activity in general (Alessandro Gabbiadini, 2017). The aim is discovering the existence of relations between Pokémon Go players and amount of general physical activity. In fact, the level of physical activity can be affected by various factors, including motivation, education level, and gender. We, thus, want to locate variables associating with the amount of physical activities. Last but not least, we want to examine the effects of the attitude towards physical activity caused by gender or educational level?

# 3. Data Description and Processing

The data was obtained from a study, carried out following the code of ethics of the world medical association (Declaration of Helsinki) for studies using human as data. Amazon Mechanical Turk (MTurk), an internet-based platform offering online participant pool, was applied for data collection (Buhrmester, Kwang, & Gosling, 2011; Paolacci & Chandler, 2014). The original data contains 999 records, described by 31 variables. Prior to processing data, we examine the number of missing values (which is 0), as well as filtering out records in accordance with the variable “ATTENTION\_filter1”.

“ATTENTION\_filter1” is a variable for filtering out non-focus participants as Mechanical Turk experiment was applied. In spite of the convenience as well as limitless of time and location, Mechanical Turk experiment cannot guarantee that participants are paying attention as the survey was completed online (Jennifer Jacquet, 2011). Ensuring only data from focusing-on-survey participants were collected, the item symboled by “ATTENTION\_filter1” was used. If failed choosing “Disagree” in this question, the records will be removed due to being classified as non-focus records. The remaining records inside the dataset is 981. After primary data cleaning, we transform all columns into integer score, according to the scales above-listed. This subjective assigning method is plausible for applying interval scale and the concept of distance (Chaowei Yang, 2014). Both ***id***, ***submitdate***, ***ipaddr*** and ***ATTENTION\_filter*** is not in consideration due to irreverent of objectives. Thus, Twenty-eight variables, for further grouping and analysis, will be discussed below.

Despite Age and gender, discrete number and nominal data respectively, all the others are ordinal variables. Both “Frequency of App Usage” and “How often sharing on social media” are ordinal data, anchored with the scale from 1 = “never” to 7 = “very often”. Former accessing the extent of players using Pokemon Go per month, while the latter demonstrate the frequency players sharing their achievements on social media. Those variables mentioned above were treated as independent variables, while the remaining variables will be grouped by row mean, in accordance with the result of Cronbach’s alpha. Cronbach’s alpha, also known as alpha reliability, is a measure for assessing strength of internal consistency, of several items or variables. The alpha score was calculated by correlating the score for every items with the total score for related observations, following the comparison of the variance of individual item scores (Cronbach Lee, 1951). The formula is:

Where is number of scale items, is to the variance associated with item , and denoted as the variance associated with the observed total scores (Chelsea Goforth, 2015). In accordance with Rule of Thumb, if the alpha score is between 0.7 and 0.8, the grouping process is plausible (Stephanie Glen, 2021). Proving by alpha score, We group variables having strong internal consistency by mean of each instances. Cortina (1992) mentioned that grouping is acceptable if score is larger than 0.7 as this score is unterpreted, recommend in many researches additionally. In supplement, group-by-mean method (Underhill L.G, 1998) was applied as we prefer grouping variables, without missing much information (like median)(Akhihesh Ganti, 2021) or altering the scale(Daniel McNeish & Melissa Gordon Wolf, 2020). The details were mentioned below.

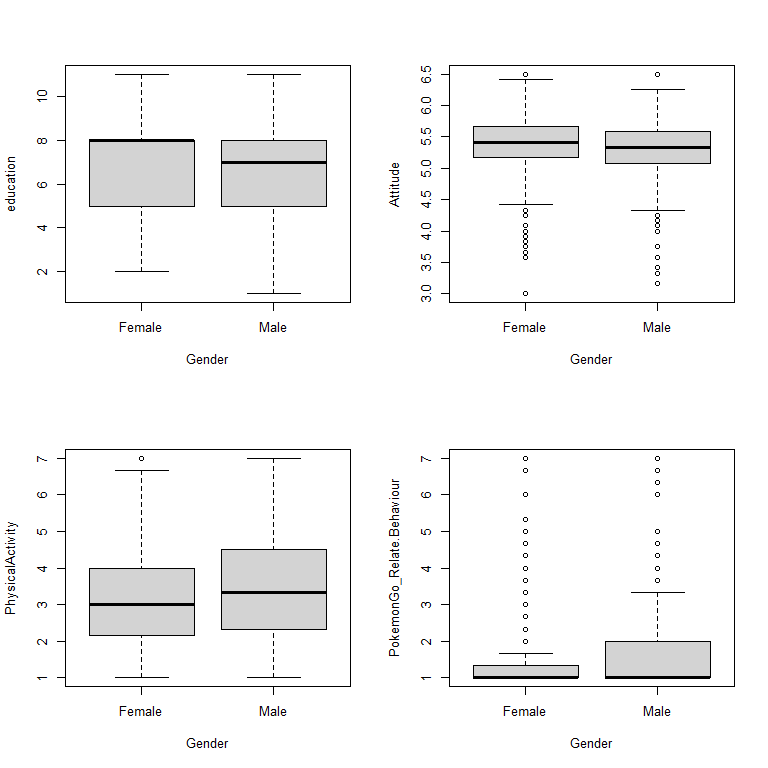
There are 12 variables, grouped as , measuring participants’ attitude towards physical activities (). (Scale of all questions were from 1 = “completely disagree” to 7 = “completely agree”). Players’ physically behavior was assessed in two aspects, recency and frequency (). The first three items for measuring recency of participants’ physical activity were “When was the last time you had (1) a walk for more than 30 min/(2) had a run/(3) had a bike ride to get some exercise?”. (The scale for those questions is 1=“more than one month ago”, 2=“about four weeks ago”, 3=“about three weeks ago”, 4=“about two weeks ago”, 5=“about one week ago”, 6=“during the last week” and 7=“yesterday”.) For measuring frequency, the following three questions were adopted: “How many times have you had (1) a walk for more than 30 min/ (2) had a run/ (3) had a bike ride to get some exercise during the last month? with the scale from 1=” never" to 7=“every day”. Both former and latter were transformed as one variable, . The remaining three variables, names starting with “PokemonPastBehaviour”, were used for assessing participants’ physical behavior relating to Pokémon Go (). Questions represented by these three variables were “How many times have you walked more than 30 min/ had a run/ had a bike ride with the intent of searching for Pokémon Go during the last month?” (anchored with 1= “never”, 2= “two times”,3= “from three to five times”,4= “from six to eight times”, 5= “from nine to eleven times”, 6= “from twelve to fourteen times” and 7= “every day”). Those variables were grouped as . Back of grouping, the new data set contains eight variables and 981 records. Prior to model selection, we look at the summary:

## age education Gender Attitude   
## Min. :18.00 Min. : 1.000 Min. :1.000 Min. :3.000   
## 1st Qu.:25.00 1st Qu.: 5.000 1st Qu.:1.000 1st Qu.:5.167   
## Median :30.00 Median : 8.000 Median :1.000 Median :5.417   
## Mean :32.55 Mean : 6.889 Mean :1.378 Mean :5.362   
## 3rd Qu.:37.00 3rd Qu.: 8.000 3rd Qu.:2.000 3rd Qu.:5.667   
## Max. :74.00 Max. :11.000 Max. :2.000 Max. :6.500   
## PhysicalActivity PokemonGo\_AppUsage social\_sharing PokemonGo\_Relate.Behaviour  
## Min. :1.000 Min. :1.00 Min. :1.000 Min. :1.000   
## 1st Qu.:2.333 1st Qu.:1.00 1st Qu.:1.000 1st Qu.:1.000   
## Median :3.000 Median :1.00 Median :1.000 Median :1.000   
## Mean :3.244 Mean :2.45 Mean :1.611 Mean :1.492   
## 3rd Qu.:4.167 3rd Qu.:4.00 3rd Qu.:1.000 3rd Qu.:1.667   
## Max. :7.000 Max. :7.00 Max. :7.000 Max. :7.000

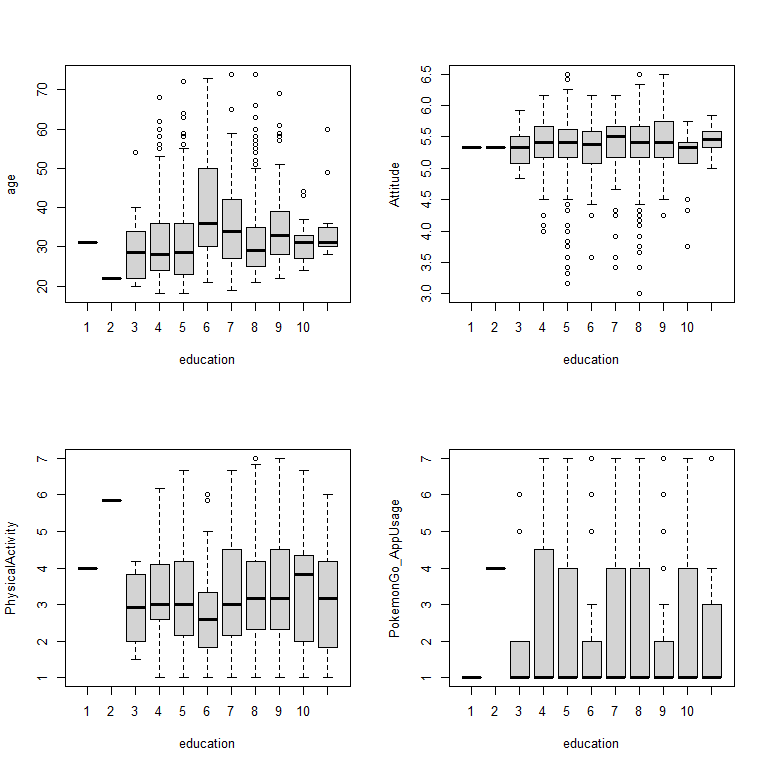
According to the previous summary, there are 981 records and 8 variables. In accordance with the pattern demonstrated in plot 3.2, we can observed that “PokemonGo\_Relate.Behaviour” have unusual relations with three variables, including , and . This is possible that there are curve-linear relations between variables, with itself or others. Plot 3.3, additionally, showed that relations exists between and three variables, like , and . Plot 3.4 also proved that “education level” has positive relations with and . For modeling the interactions between variables, we apply polinomial regression model, mentioning details in the following session.



Plot 3.2: Scatter plot of data distribution



Plot 3.3: Boxplots of Gender vs 4 other variables



Plot 3.4: Boxplots of Education level vs 4 other variables

# 4. Methodology

Polynomial linear regression model, a linear regression model with degree of coefficient more than one, is a model combining interactions between variables (Abhigyan, 2020). Polynomial regression can model the non-linear relationship between dependent and independent variables by adding polynomial terms to linear regression, such as the square of a variable (Agrawal. 2021). Polynomial regression was selected as the most suitable approximation of relations between variables can be provided. Some variables, in reality, have correlations with others or itself. Simple linear regression, however, cannot model these interactions, leading large error and inaccurate relationship estimation. With polynomial terms, we built up a better model for relations observations. A vast range of model, also, can be applied in polynomial regression model, including linear, Gamma and Poisson. Curvature of polynomial regression model, moreover, is flexible (Pant. 2019). We can fit wide range of positive values. Due to accuracy and flexibility, we select polynomial regression model. Linear model was applied as it is easy for interpretation, as well as explanation. There are four assumptions for polynomial linear model. The behavior of a response variable, to begin with, can be explained by an additive relationship (both linear and curvilinear are plausible) between a response variable and several explanatory variables. It is essential that the relations between response variable and all explanatory variables are linear or curvilinear. The independent variables, additionally, should be independent with each other. Last but not least, the errors must be independent and normally distributed, following a mean zero and constant variance (Abhigyan. 2020). We examine the following assumption after obtaining the best model by stepwise selection method.

After developing full linear model, stepwise selection, using AIC as criteria, was applied to select the best model. Stepwise regression is a step-by-step iterative and automatic model selection approach (Adam Hayes, 2021), based on backward regression and combining with forward. With this approach, we can re-examine the importance of variables, as well as correct the misleading caused by backward selection. For instance, it is plausible that a variable, removed in backward selection, is included in first stage of forward selection method. Applying stepwise selection method, we can include that variable again, for obtaining the best model (R. R. Hocking, 1967). McElreath (2016) revealed that Akaike Information Criteria (AIC) is a well-known information criterion, for evaluating the data-fitting performance of a model. The model with smallest AIC is the best as describing greatest amount of information with smallest amount of variables (Bevans, 2021). It is necessary for checking the model assumption, to guarantee that model can be applied. The assumptions were examined by all four plots in **Plot 4.2**, the explanation and examination of assumption plots were written below.

According to the plot, Residuals vs Fitted in **Plot 4.2**, the pattern of residuals is not obvious, suggesting that the assumption of linear or curvilinear is acceptable. The residuals spread equally around the zero line, proved that the error terms have same variance. Outliers, additionally are not exists as no residual standing away from the pattern (Department of Statistics Online Programs. 2018); Although having a light tail, Normal Q-Q plot suggested that the dependent variables, inside the model, are normally distributed (Ford. 2015). In Scale-Location plot, Since the red line is approximately horizontal across the plot, with no clear pattern. In this way, spreading of the residuals is random, as well as in the neighborhood of equal for all fitted values. (Zach. 2020). Observing Residuals Vs Leverage, the last plot in **Plot 4.2**, there is no points affecting the trend much. In this way, there is no outliers. Based on the graphs in **Plot 4.2**, the polynomial linear regression model, demonstrated in the summary, was applied for studying the relations between Pokémon Go and amount of physical activity. The details final model was demonstrated in **Table 4.1**.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Formula** | | | | |
| PhysicalActivity = age + education + Gender + Attitude + PokemonGo\_AppUsage + PokemonGo\_Relate.Behaviour + Attitude\* Attitude + PokemonGo\_Relate.Behaviour\* PokemonGo\_Relate.Behaviour + age\*education + education\*Attitude | | | | |
| Coefficients | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 6.136942 | 3.010388 | 2.039 | 0.04176 |
| age | 0.012821 | 0.015683 | 0.818 | 0.41382 |
| education | -0.279905 | 0.284092 | -0.985 | 0.32474 |
| Gender | 0.278323 | 0.084595 | 3.290 | 0.00104 |
| Attitude | -1.761775 | 1.013901 | -1.738 | 0.08260 |
| PokemonGo\_AppUsage | -0.198716 | 0.035205 | -5.645 | 2.17e-08 |
| PokemonGo\_Relate.Behaviour | 0.973577 | 0.207318 | 4.696 | 3.03e-06 |
| I(Attitude^2) | 0.155076 | 0.097228 | 1.595 | 0.11104 |
| I(PokemonGo\_Relate.Behaviour^2) | -0.046845 | 0.029368 | -1.595 | 0.11101 |
| age\*education | -0.003554 | 0.002289 | -1.553 | 0.12079 |
| education\*Attitude | 0.085761 | 0.051815 | 1.655 | 0.09822 |

*Table 4.1: Formula and Summary of Final Model*



Plot 4.2: model assumption plots

# 5. Result Analysis

With the polynomial regression model constructed before, we answer the questions mentioned in the “research objectives” session. In accordance with the summary (**Table 4.1)**, the estimate of PokemonGo\_AppUsage is -0.198716, with a variation approximately 0.0002460, meaning that 1% increase of App usage lower -0.198716 amount of general physical activity. In contrast with the expectation, playing Pokémon Go frequently negatively affect the amount of physical activity. This phenomenon can be related to variable “PokemonGo\_Relate.Behaviour”.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Formula** | | | | |
| PhysicalActivity = age + education + Gender + Attitude + PokemonGo\_AppUsage + Attitude\* Attitude + age\*education + education\*Attitude | | | | |
| Coefficients | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 7.070780 | 3.010388 | 3.164619 | 0.0257 |
| age | -0.001712 | 0.016454 | -0.104 | 0.9171 |
| education | -0.016170 | 0.298011 | -0.054 | 0.9567 |
| Gender | 0.392655 | 0.088170 | 4.453 | 9.43e-06 |
| Attitude | -2.156115 | 1.060194 | -2.034 | 0.0423 |
| PokemonGo\_AppUsage | 0.051826 | 0.021254 | 2.438 | 0.0149 |
| I(Attitude^2) | 0.230975 | 0.101574 | 2.274 | 0.0232 |
| age\*education | -0.001736 | 0.002289 | -1.553 | 0.12079 |
| education\*Attitude | 0.085761 | 0.051815 | 1.655 | 0.09822 |

*Table 5.1: Formula and Summary of Model without “****PokemonGo\_Relate.Behaviour***”

# Discussion and Future Works

**Conclusion**

# Reference

# Agrawal Raghav. (2021). “All you need to know about Polynomial Regression”. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/07/all-you-need-to-know-about-polynomial-regression/

# Bevans Rebecca. (2021). "An introduction to the Akaike information criterion". Scribbr. https://www.scribbr.com/statistics/akaike-information-criterion/

# Bradfield & Underhill. (2004). "IntroSTAT". 2nd. Department of Statistical Sciences, University of Cape Town. p.181-183 https://books.google.co.uk/books?id=f6TlVjrSAsgC&lpg=PP1&pg=PA181&redir\_esc=y#v=onepage&q&f=false

# Buraimo Babatunde, Jones Helen & Millward Peter. (2011). "Adult participation in sport: Analysis of the Taking Part Survey". department for culture, media and sport. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/137986/tp-adult-participation-sport-analysis.pdf

# Cortina Jose. (1992). "What Is Coefficient Alpha? An Examination of Theory and Applications". American Psychological Association, Inc. p.101

# Cronbach, Lee J. (1951). "Coefficient alpha and the internal structure of tests". Psychometrika. Springer Science and Business Media LLC. 16 (3): 297-334

# Department of Statistics Online Programs. (2018). “4.2 - Residuals vs. Fits Plot”. The Pennsylvania State University. https://online.stat.psu.edu/stat462/node/117/

# Eysenbach Gunther. (2016). "Influence of PokÃ©mon Go on Physical Activity: Study and Implications". Journal of Medical Internet Research. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5174727/

# Fisher Tim. (2021). "What Is Augmented Reality?". Lifewire. https://www.lifewire.com/augmented-reality-ar-definition-4155104

# Ford Clay. (2015). “Understanding Q-Q Plots”. University of Virginia. https://data.library.virginia.edu/understanding-q-q-plots/

# Ganti Akhilesh. (2021). "Median". Investopedia. https://www.investopedia.com/terms/m/median.asp

# Glen Stephanie. (2021). "Cronbachâs Alpha: Simple Definition, Use and Interpretation". Statistics How To. https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/cronbachs-alpha-spss/

# Goforth Chelsea. (2015). "Using and Interpreting Cronbachâs Alpha". University pf Virginia Library. https://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/

# Hayes Adam. (2021). "Stepwise Regression". Investopedia. https://www.investopedia.com/terms/s/stepwise-regression.asp

# Hocking, R. R. (1976). A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. Biometrics, 32(1), 1â49. https://doi.org/10.2307/2529336

# Jacquet Jennifer. (2011). "The pros & cons of Amazon Mechanical Turk for scientific surveys". Scientific American. https://blogs.scientificamerican.com/guilty-planet/httpblogsscientificamericancomguilty-planet20110707the-pros-cons-of-amazon-mechanical-turk-for-scientific-surveys/

# Kostoulas Achilleas. (2013). "On Likert scales, ordinal data and mean values". Achilleas Kostoulas. https://achilleaskostoulas.com/2013/02/13/on-likert-scales-ordinal-data-and-mean-values/

# McElreath Richard. (2016). "Statistical Rethinking: A Bayesian Course with Examples in R and Stan". CRC Press. p.189

# McNeish, D., Wolf, M.G. (2020). Thinking twice about sum scores. Behav Res 52, p.2287â2305. https://doi.org/10.3758/s13428-020-01398-0

# Pardoe Iain. (2017).

# Reilly Luke. (2017). "PokÃ©mon GO Coming to Smartphones". IGN Entertainment. https://www.ign.com/articles/2015/09/10/pokemon-go-coming-to-smartphones

# Seippel Ornulf. (2006). "The Meanings of Sport: Fun, Health, Beauty or Community?". Tandfonline https://www.tandfonline.com/doi/full/10.1080/17430430500355790?scroll=top&needAccess=true

# Sengupta Somak. (2020). "Gamma Distribution Explained | What is Gamma Distribution?". Great Learning. https://www.mygreatlearning.com/blog/gamma-distribution/

# Smith, C. (2017, November 12).80 amazing Pokemon Go statistics. RetrievedNovember 17, 2017, from https://expandedramblings.com/index.php/pokemon-go-statistics/.

# Stacy, E. W. (1962). A Generalization of the Gamma Distribution. The Annals of Mathematical Statistics, 33(3), 1187â1192. http://www.jstor.org/stable/2237889

* Webster Andrew. (2015). "With PokÃ©mon Go, Nintendo is showing that it takes mobile seriously". The Verge. https://www.theverge.com/2015/9/10/9300101/pokemon-go-nintendo-mobile-games
* Zach. (2020). “How to Interpret a Scale-Location Plot (With Examples)”. Statology. https://www.statology.org/scale-location-plot/