**Analysis of Video Game Sales**

**Name : Dinesh S**

**Email id :** [**sethudinesh1998@gmail.com**](mailto:sethudinesh1998@gmail.com)

**College : Sri Sivasubramaniya Nadar College of Engineering**

**Mobile No : 7358167166**

1. **Introduction**

The ultimate aim of this project is to analyze the video game sales across countries like North America, Europe, Japan, etc., considering many factors and how they are affecting them.

In this project, the analysis is done with insights from graphs, histograms, box plots, bar plots drawn between the global sales versus different factors such as sales in North America, Europe, Japan.

The data is collected from vgchartz.com in October 26, 2016 which investigates the video game sales across three or more countries during the time period, 1980 . It includes various attributes such as Rank, Name, Platform, Year, Genre, Publisher, North America Sales, Europe Sales, Japan Sales, Other Sales, Global Sales. These attributes will be analyzed as to which extent it will affect the global sales.

Here we will look for various correlations amongst various variables in our dataset. Many genre’s such as sports, platform, racing, role-play, puzzle, misc, action flourish under the umbrella of video games.

In this paper, we investigate whether which factors will affect the global sales and also predicts in which year the video game was released, the publisher name, platform and genre.

1. **Overview**

Our field study concerns sales of video games across three or more countries. Here we try to use tests such as t-test, correlation tests (pearson or spearman), regression analysis and histograms, box plots, bar plots to gather insights about the data. We also have a year wise comparison of global sales across all countries. Our analysis reveals that attributes such as North America Sales, European sales, Japan sales affect the global sales but attributes such as Platform, Year, genre does not affect the global sales.

1. **An Empirical Field Study of Video game Sales across North America, Europe and Japan**
   1. **Correlation**

We had done correlation to test global sales with attributes and found three most important attributes. They are :

1. Sales in North America
2. Sales in Europe
3. Sales in Japan

We have also found there are independent attributes which global sales does not depend upon. They are :

1. Publisher
2. Genre
3. Year
4. Name
   1. **Hypothesis**

We study how the global sales is affected by sales in north America, Japan, Europe. We assume that out of 11 attributes, there are 5 important attributes. They are sales in North America, Japan, Europe, Other sales and Global sales on the basis of strong correlation.

The hypothesis were :

1. The sales in North America is greater than that of Europe
2. The sales in Japan is less than that of Europe
   1. **Data**

The purpose of this project is to identify the factors that matter the most. The dataset tracks global sales across three or more countries. My dataset is based on genre’s such as action, adventure, fighting, misc, platform, puzzle, racing, role-playing, shooter, simulation, sports and strategy. It is indeed probable that many factors govern the global sales.

Any meaningful empirical analysis will need to control for attributes such as whether the global sales is affected by sales in North America, Europe, Japan and the genre in which the video game is published and also includes the publisher.

Platform – It indicates in which platform the video game was released

Year - It indicates in which year the video game was released

Publisher – Denotes the publisher name

Genre - Denotes in which genre the video game was released

NA\_Sales - Indicates the Sales in North America

EU\_Sales - Indicates the sales in Europe

JP\_Sales - Indicates the sales in Japan

Other\_Sales – Sales in rest of the world

Global\_Sales – Indicates the total world wide sales

* 1. **Model**

The research question was analyzed. We established the effect of sales in North America, Europe, Japan and Other sales with the simplest model we could come up with.

***Model 1: Global sales vs. North America Sales***

```{r}

fit <- lm(Global\_Sales ~ NA\_Sales, data=video.df)

summary(fit)

```

Call:

lm(formula = Global\_Sales ~ NA\_Sales, data = video.df)

Residuals:

Min 1Q Median 3Q Max

-20.0071 -0.1086 -0.0528 0.0172 11.6456

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.063202 0.004292 14.72 <2e-16 \*\*\*

NA\_Sales 1.791827 0.005000 358.38 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.526 on 16596 degrees of freedom

Multiple R-squared: 0.8856, Adjusted R-squared: 0.8856

F-statistic: 1.284e+05 on 1 and 16596 DF, p-value: < 2.2e-16

**From the above model,**

**p-value = less than 0.001**

**Multiple r-squared and adjusted r-squared = 0.8856**

***Model 2: Global sales vs. European sales***

```{r}

fit <- lm(Global\_Sales ~ EU\_Sales, data=video.df)

summary(fit)

```

Call:

lm(formula = Global\_Sales ~ EU\_Sales, data = video.df)

Residuals:

Min 1Q Median 3Q Max

-11.1023 -0.1378 -0.0856 0.0274 30.1642

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.130021 0.005404 24.06 <2e-16 \*\*\*

EU\_Sales 2.778137 0.010271 270.49 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6687 on 16596 degrees of freedom

Multiple R-squared: 0.8151, Adjusted R-squared: 0.8151

F-statistic: 7.317e+04 on 1 and 16596 DF, p-value: < 2.2e-16

**From the above model,**

**p-value = less than 0.001**

**Multiple r-squared and adjusted r-squared = 0.8151**

***Model 3: Global sales vs. Japan sales***

```{r}

fit <- lm(Global\_Sales ~ JP\_Sales, data=video.df)

summary(fit)

```

Call:

lm(formula = Global\_Sales ~ JP\_Sales, data = video.df)

Residuals:

Min 1Q Median 3Q Max

-10.408 -0.319 -0.198 0.062 70.845

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.298181 0.009845 30.29 <2e-16 \*\*\*

JP\_Sales 3.076039 0.030871 99.64 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.23 on 16596 degrees of freedom

Multiple R-squared: 0.3743, Adjusted R-squared: 0.3743

F-statistic: 9929 on 1 and 16596 DF, p-value: < 2.2e-16

**From the above model,**

**p-value = less than 0.01**

**Multiple r-squared and adjusted r-squared = 0.3743**

**The best model is chosen based upon two conditions:**

**1. The model must have lesser p-value than all.**

**2. It must also have higher multiple r-squared or adjusted r-squared value.**

**The model which satisfies the above two conditions is Model 3. Therefore, this is the best model of all.**

**The benefits of having the three regressors outlined in the models was that it helped us rule out some alternate explanations for the variation in global sales.**

1. **Results Output**

The coefficients and linear model of the above three models are mentioned below :

***Model 1: Global sales vs. North America Sales***

```{r}

fit <- lm(Global\_Sales ~ NA\_Sales, data=video.df)

fit$coefficients

```

(Intercept) NA\_Sales

0.06320234 1.79182728

Global Sales(Y) = b0 + NorthAmerica Sales(b1)

b0=-1, b1=1.7918

Global Sales = -1 + NorthAmerica Sales\*1.7918

***Model 2: Global sales vs. European sales***

```{r}

fit <- lm(Global\_Sales ~ EU\_Sales, data=video.df)

fit$coefficients

```

(Intercept) EU\_Sales

0.1300213 2.7781369

Global Sales(Y) = b0 + European Sales(b1)

b0=-1, b1=2.7781

Global Sales = -1 + European Sales\*2.7781

***Model 3: Global sales vs. Japan sales***

```{r}

fit <- lm(Global\_Sales ~ JP\_Sales, data=video.df)

fit$coefficients

```

(Intercept) JP\_Sales

0.2981812 3.0760394

Global Sales(Y) = b0 + Japan Sales(b1)

b0=-1, b1=3.0760

Global Sales = -1 + Japan Sales\*3.0760

***Model 4: Global Sales vs. other sales***

```{r}

fit <- lm(Global\_Sales ~ Other\_Sales + NA\_Sales + EU\_Sales + JP\_Sales, data=video.df)

fit$coefficients

```

(Intercept) Other\_Sales NA\_Sales EU\_Sales JP\_Sales

0.0003229497 0.9995874903 0.9999405824 0.9999875831 0.9998838156

Global Sales(Y) = b0 + Other Sales(b1) + NorthAmerica Sales(b2) + European Sales(b3) + Japan sales(b4)

b0=-1, b1=0.9995, b2=0.99994, b3=0.99998, b4=0.99988

Global Sales = -1 + Other Sales(0.9995) + NorthAmerica Sales(0.99994) + European Sales(0.99998) + Japan Sales(0.99988).

1. **Conclusions**

This paper was motivated by the need for research that could improve our understanding of how external and internal factors influences the worldwide sales of video games

The unique contribution in this paper is that we investigated the year wise global sales comparison, according to genre’s – the sales in different countries and the worldwide sales.

This research has also some important managerial relevances. When the buyer sees good rating or review about the video game it urges him to see the description about the video game and an willingness to pay.

1. **Closing Note**

From the above outputs, we have found out that

1. In which year the global sales was high,low.

2. Comparisons between sales in northamerica, europe, japan and others

3. Genre which had the highest and lowest global sales.