# SCHOOL OF COMPUTING

# COURSE: SECI 2143-06

PROBABILITY AND STATISTICAL DATA ANALYSIS

PROJECT REPORT

INDIVIDUAL PROJECT-2

DR CHAN WENG HOWE

# RELATIONSHIP OF A FORTNITE PLAYER IN-GAME STATISTICS

| NAME | MATRIC ID |
|---|---|
| HUSNY MUSHARRAF BIN SHAMSUL KAMAL | A19EC0051 |

# TABLE OF CONTENTS

## Contents

# INTRODUCTION

In early 2017, the first battle royale game was released and it changed the world of gaming. It is called PlayerUnknown Battleground (PUBG). When this game is released, it inspires many game developers to try and make a game that is based on this type of mode which is Battle Royale mode. In September 2017, Epic Games finally released their own battle royale game and it is called Fortnite. However, Epic Games seems to come on a different approach with making it a free to play game and achieved millions of players downloaded the game, it also won a Ultimate Game of the Year award in 2018 at the 36th Golden Joystick Awards. (Hoggins, 2018) Furthermore, this battle royale mode is not like any other type of game where you can win and perform a good gameplay every game. Therefore, there is one player who make his own statistics of every game he played. It is very interesting to see how random he performs in each game and this is what I am going to study about.

The purpose of this project is to study on the statistics of a Fortnite player in 87 endgame. The variables for this project is the mental state of the player while playing, the number placed in a game, damage done to other players, damage taken from other players, and the distance traveled in a single game. The first test is to investigate whether the damage to other players deals higher in a high mental state or sober mental state. The second test is to test out the correlation between the distance traveled with the number placed in the game. The third test is to observe the relationship between damage taken from other players with damage done to other players. Finally, the last test is to determine whether the player mental state effect the number placed in the game.

# METHODOLOGY

The data set that is used in this project is called "Fortnite Statistics" and it is obtained from a website called data.world. The data set is a secondary data. This data set has over 80 games statistic of a player with 16 different variables. In order to study the statistics, only 5 variables are taken which is the mental state of the player while playing, the number placed in a game, damage done to other players, damage taken from other players, and the distance traveled in a single game. On the other hand, the types of analysis test that is used in this project is Hypothesis Testing 2 Samples, Correlation, Regression and Chi-Square Test for Independence. All the analysis for this project is conducted using Rstudio.

# DATA ANALYSIS AND RESULTS

## 1. Hypothesis Testing 2-Samples

This test is to analysis whether the mean damage to players with the player on a high mental state deals lower mean damage with the player on a sober mental state. These two samples are independent with the sample size of high = 42, sober = 45. The population variances are assumed to be unequal. Hence, this test is conducted at $\alpha = 0.05$, $\alpha$ = level of significance.

Null Hypothesis : The mean of damage deals to players on high mental state is the same as the mean damage deals to players on sober mental state.

Alternate Hypothesis : The mean of damage deals to players on high mental state is lower than the mean damage deals to players on sober mental state.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$

where,

$\mu_1$ is the mean of damage deals to players on high mental state and $\mu_2$ is the mean of damage deals to players on sober mental state.

```
> highdmgtoplyr = dmgtoplyr[`Mental State` == "high"]
> soberdmgtoplyr = dmgtoplyr[`Mental State` == "sober"]
> t.test(highdmgtoplyr, soberdmgtoplyr, alt = "less")

        Welch Two Sample t-test

data:  highdmgtoplyr and soberdmgtoplyr
t = -0.61283, df = 83.661, p-value = 0.2708
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 80.27003
sample estimates:
mean of x mean of y
 557.5476  604.3778

> qt(0.05, 84)
[1] -1.663197
```

Figure 1 : Test statistic calculated using Rstudio

Based on figure 1, the sample mean of x(high) is 557.5476 while the sample mean of y(sober) is 604.3778. The degree of freedom is df = 83.661 rounded up to 84. The test statistic, $T_0$ = -0.61283. The P-value of the test statistic is 0.2708.

At $\alpha = 0.05$, the critical value is $t_{0.05,84} = -1.663197$.

Since $T_0$ = -0.61283>-1.663197 or P-value = 0.2708>0.05, we fail to reject $H_0$, null hypothesis. This is because there is not enough evidence to prove that the mean of damage deals to players on high mental state is lower than the mean damage deals to players on sober mental state at significance level of 0.05.

## 2. Correlation

For correlation, the variables that is used to test is the distance traveled in a game with the number placed in a game. To determine the strength of the relationship between the variables, Pearson's product-moment correlation coefficient is used to find the *r, correlation coefficient.*

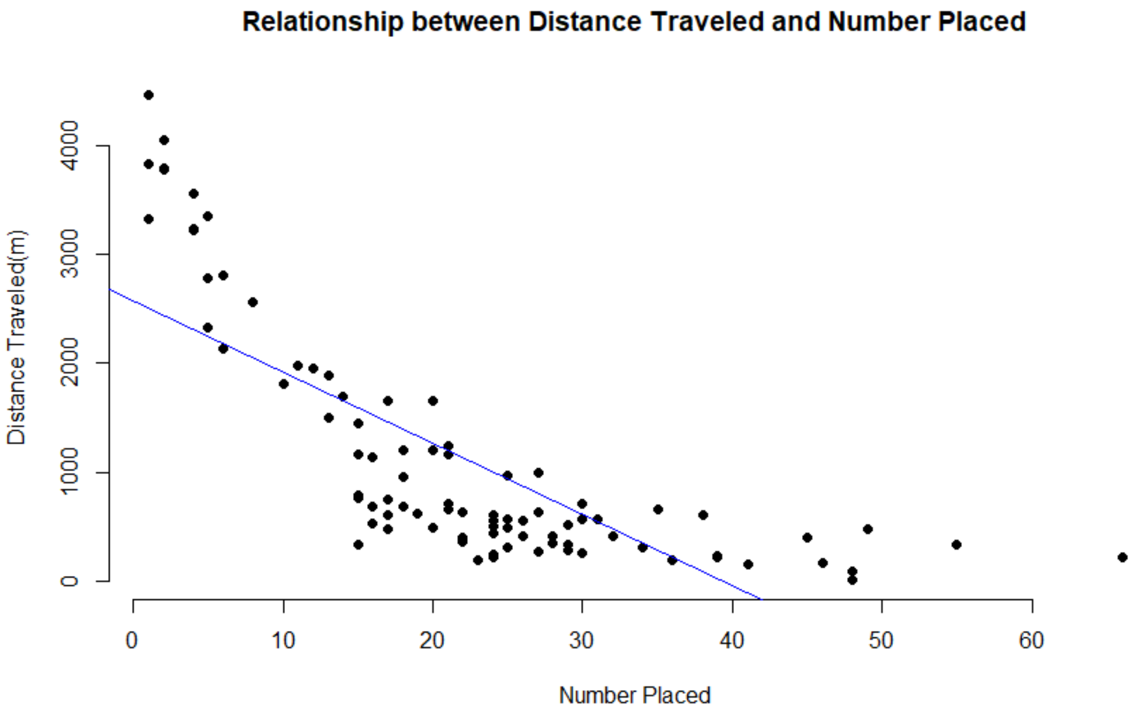


Figure 2 : Relationship between Distance Traveled and Number Placed

By observing the Figure 2, we can see that the Distance Traveled increases as Number Placed decreases. This shows that there is negative relationship between the variables.

```
> plot(x, y, main = "Relationship between Distance Traveled and Number Placed",xlab = "Number Placed", ylab = "Distance Trave
led(m)", pch = 19, frame = FALSE)
> abline(lm(y~x), col = "blue")
> cor.test(x,y)

	Pearson's product-moment correlation

data:  x and y
t = -11.149, df = 85, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.8442390 -0.6685133
sample estimates:
       cor
-0.7706393

> qt(0.025, 85)
[1] -1.988268
```

Figure 3 : Correlation coefficient and test statistic calculated using Rstudio

Based on Figure 3, the *r, coefficient correlation* and *t, test statistic* is calculated using Rstudio. The variable x represents the Number Placed while variable y represents the Distance Traveled. From the figure above, $r$ = -0.7706393, from this value, we can say that there is a strong relationship between x and y. To further investigate, we test the significant test for the correlation with level of significance, $\alpha = 0.05$, since this is two tailed test, $\alpha/2$ = 0.025.

Null Hypothesis : There is no linear correlation between the Distance Traveled and Number Placed.

Alternate Hypothesis : There exist a linear correlation between the Distance Traveled and Number Placed.

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Critical value : $t_{0.025, 85}$ = -1.988268, 1.988268

Test statistic, t = -11.149

P-value = $2.2\text{x}10^{-16}$

Since t = -11.149<-1.988268 or P-value = $2.2\text{x}10^{-16} < 0.05$, we reject $H_0$, null hypothesis. This is because there is sufficient evidence to prove that there exists a linear correlation between the Distance Traveled and Number Placed at significance level of 0.05.

## 3. Regression

In regression test, we will test the relationship between the Damage Taken and Damage to Other Players. The independent variable is the Damage to Other Players and the dependent variable is the Damage Taken.

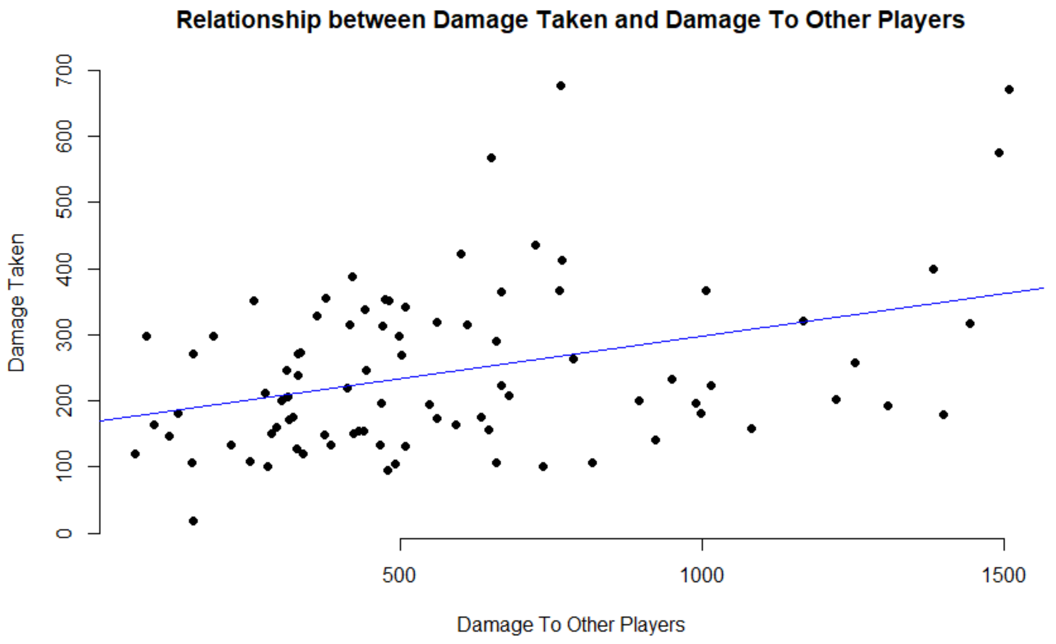


Figure 4 : Relationship between Damage Taken and Damage to Other Players

By referring to the figure above, we can say that the relationship between Damage Taken and Damage to Other Players appears to be positively linear. To further investigate the relationship, we use the linear regression model.

```
> plot(dmgtoplyr, dmgtaken, main = "Relationship between Damage Taken and Damage To Other Players",xlab = "Damage To Other Pl
ayers", ylab = "Damage Taken", pch = 19, frame = FALSE)
> abline(lm(dmgtaken~dmgtoplyr), col = "blue")
> mxc <- lm(dmgtaken~dmgtoplyr)
> mxc

Call:
lm(formula = dmgtaken ~ dmgtoplyr)

Coefficients:
(Intercept)    dmgtoplyr  
   169.2014       0.1289  

> summary(mxc)

Call:
lm(formula = dmgtaken ~ dmgtoplyr)

Residuals:
    Min      1Q  Median      3Q     Max 
-170.43  -83.19  -33.51   80.10  409.09 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 169.20139   24.22133   6.986 5.93e-10 ***
dmgtoplyr     0.12887    0.03562   3.618 0.000503 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 117 on 85 degrees of freedom
Multiple R-squared:  0.1335,	Adjusted R-squared:  0.1233 
F-statistic: 13.09 on 1 and 85 DF,  p-value: 0.0005029

> qt(0.025, 85)
[1] -1.988268
```

Figure 5 : Summary of linear regression model using Rstudio

Based on Figure 5, we can get the estimated regression model as follows,

$$\hat{y} = 169.2014 + 0.1289x$$

where,

$\hat{y}$ = Damage Taken

x = Damage to Other Players

$b_0$ = 169.2014

$b_1$ = 0.1289

From this equation, we can conclude that the average value of y = 169.2014 at x = 0. We can also conclude that as x increases by 1, the change in average of y value is 0.1289. Besides, the $R^2$, coefficient of determination is also calculated using Rstudio. So, $R^2 = 0.1335$. Since $0 < R^2 < 1$, weaker linear relationship between x and y. To test for regression, level of significance, $\alpha = 0.05$ is used, since this is two tailed test, $\alpha/2 = 0.025$.

Null Hypothesis : There is no linear relationship between the Damage Taken and the Damage to Other Players

Alternate Hypothesis : There exists linear relationship between the Damage Taken and the Damage to Other Players

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

Critical value : $t_{0.025, 85} = -1.988268, 1.988268$

Test statistic, t = 3.618

P-value = 0.0005029

Since $t = 3.618 > 1.988268$ or P-value $= 0.0005029 < 0.05$, we reject $H_0$, null hypothesis. This is because there is enough evidence to prove that there exists linear relationship between the Damage Taken and the Damage to Other Players at significance level of 0.05.

## 4. Chi-Square Test for Independence

Chi-Square Test for Independence is a test to find out whether there is a relationship between two nominal variables which is the mental state of the player and the number placed. For this test, the variable Number Placed is divided into 3 groups which is Top10 (1 to 10), Top30 (11 to 30) and Top70 (31 to 70).

```
> tbl = table(FS$Placed,`Mental State`)
> tbl
    Mental State
     high sober
  1     2     1
  2     2     1
  4     0     3
  5     0     3
  6     2     0
  8     0     1
  10    1     0
  11    0     1
  12    1     0
  13    1     1
  14    0     1
  15    3     2
  16    3     0
  17    2     2
  18    1     2
  19    0     1
  20    2     1
  21    1     3
  22    3     0
  23    1     0
  24    3     3
  25    2     2
  26    0     2
  27    2     1
  28    0     2
  29    1     2
  30    2     1
  31    0     1
  32    0     1
  34    0     1
  35    1     0
  36    1     0
  38    1     0
  39    1     1
  41    0     1
  45    0     1
  46    0     1
  48    0     2
  49    1     0
  55    1     0
  66    1     0
> tbl2 = cbind(tbl['1',]+tbl['2',]+tbl['4',]+tbl['5',]+tbl['6',]+tbl['8',]+tbl['10',],tbl['11',]+tbl['12',]+tbl
['13',]+tbl['14',]+tbl['15',]+tbl['16',]+tbl['17',]+tbl['18',]+tbl['19',]+tbl['20',]+tbl['21',]+tbl['22',]+tbl
['23',]+tbl['24',]+tbl['25',]+tbl['26',]+tbl['27',]+tbl['28',]+tbl['29',]+tbl['30',],tbl['31',]+tbl['32',]+tbl
['34',]+tbl['35',]+tbl['36',]+tbl['38',]+tbl['39',]+tbl['41',]+tbl['45',]+tbl['46',]+tbl['48',]+tbl['49',]+tbl
['55',]+tbl['66',])
```

```
> chiqme<-chisq.test(tbl2, correct = FALSE)
> chiqme

        Pearson's Chi-squared test

data:  tbl2
X-squared = 0.41523, df = 2, p-value = 0.8125

> qchisq(0.05, 2)
[1] 0.1025866
> chiqme$observed
      [,1] [,2] [,3]
high     7   28    7
sober    9   27    9
> chiqme$expected
          [,1]     [,2]     [,3]
high  7.724138 26.55172 7.724138
sober 8.275862 28.44828 8.275862
```

Figure 6 : Calculations and Contingency Table using Rstudio

Based on Figure 6, the [,1] represents the Top10, the [,2] represents the Top30 and lastly the [,3] represents the Top70. Test is conducted with a significance level of 0.05.

$H_{0,}$ Null Hypothesis : The Mental State of the player and the Number Placed are independent

$H_{1,}$ Alternate Hypothesis : The Mental State of the player and the Number Placed are dependent

Critical value : $\chi^2_{0.05,\ 2}$ = 0.1025866

Test statistic, $\chi^2$ = 0.41523

P-value = 0.8125

Since $\chi^2$ = 0.41523>0.1025866, we reject $H_0$, null hypothesis. This is because there is sufficient evidence to prove that the Mental State of the player and the Number Placed are dependent at significance level of 0.05.

## CONCLUSION

To put in a conclusion, for hypothesis testing 2-samples, we can conclude that the mean of damage deals to players on high mental state is higher than the mean damage deals to players on sober mental state. This can be true because it might be that the player on high mental state is increasing his own performance in gaming. After that, for the correlation test, we can conclude that there exists a linear correlation between the Distance Traveled and Number Placed. This result is true because the more you traveled in the game means you survive longer which leads to a higher number placed in the game. Furthermore, for regression test, we can conclude that there exists linear relationship between the Damage Taken and the Damage to Other Players. For this result, it is also true because the more damage taken means you encountered more enemies in the game which will makes the damage to other players higher. Lastly, for the Chi-Square Test for Independence, we can conclude that the Mental State of the player and the Number Placed in the game are dependent. This explains that the condition of a human mental state always effects everything we do.

# REFERENCES

Hoggins, T. (2018, November 19). *Golden Joystick Awards 2018: Fortnite wins game of the year while God of War takes four*. Retrieved from The Telegraph: https://www.telegraph.co.uk/gaming/news/golden-joystick-awards-2018-fortnite-wins-game-year-god-war/

Reynolds, K. (2018). *Fortnite Statistics_80 Games*. Retrieved from data.world: https://data.world/kreynol3/fortnite-statistics80-games