Universitat Autònoma de Barcelona

Science Faculty

What is the favourite weapon in Valorant?

Author:

Ona Sánchez: 1601181

08/06/2022

Índex

1 Introduction **3**

1.1 What is Valorant? . . . 3

1.2 Weapons in Valorant . . . 3

1.3 Goal of the project . . . 4

2 Description of the data set **5**

3 Initial approach to the data **6**

4 Non-parametric bootstrap **7**

4.1 Goal . . . 7

4.2 Estimation of the mean . . . 8

4.2.1 Analysis . . . 8

4.3 Filtered data . . . 9

4.3.1 Analysis . . . 10

4.4 Additional information . . . 11

4.4.1 Analysis . . . 11

5 Parametric bootstrap **12**

5.1 Goal . . . 12

5.2 Estimation of the probability p . . . 13

5.2.1 Analysis . . . 14

6 Conclusions **15**

7 Bibliography **16**

8 R scripts **17**

8.1 Initial approach to the data set . . . 17

8.2 Non-parametric bootstrap . . . 18

8.2.1 Initial mean . . . 18

8.2.2 Mean filtered data set . . . 18

8.2.3 Median . . . 18

8.3 Parametric bootstrap . . . 19

1 Introduction

1.1 What is Valorant?

Valorant is a first-person shooter with an objective-based focus that revolves around planting and diffusing bombs. Players play as one of a set of Agents, characters designed based on several countries and cultures around the world. In the main game mode, players are split between attackers and defenders, protecting or assaulting bomb spots, and while their shooting skill is key to success, they'll also be able to use different character abilities.

Every player starts each round with a classic pistol and one or more signature ability. Other weapons and abilites can be purchased using an in-game economic system.

1.2 Weapons in Valorant

The different weapons to choose from are:

1.3 Goal of the project

In my personal experience playing Valorant, I have a go-to gun which is also my favourite, the Vandal.

When talking to other players, I have noticed that a lot of them also had the Vandal as their favourite, while other prefered snipers or shotguns.

My goal is to determine which is the most liked gun of all of the ones shown in the picture above, using a non-parametric bootstrap.

For the most liked gun, and assuming that the players of the game always choose their favourite gun, the probability of at least someone of the enemy team to choose it, using a parametric bootstrap.

2 Description of the data set

The data set used for this project was taken by myself.

I created a google form and asked valorant players to write their name and choose their favourite gun.

The google form looked like this (but with all the weapons):

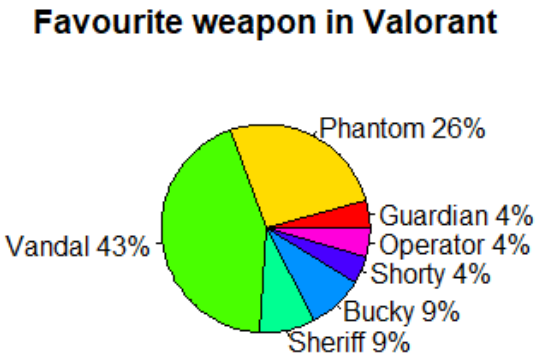
The google form generates automatically an excel with the data, which was the Nickname of the player and their favourite weapon. After collecting the data, I gave every weapon a number from 0 to 6, so the final data set was:

Nickname	Favourite weapon	Number associated
Jeke	Vandal	2
Ibra	Sheriff	3
Dori	Vandal	2
TaurumTheHunter	Guardian	0
Duna	Vandal	2
Pp	Vandal	2
Marufa	Bucky	4
Tontolhaba77	Phantom	1
I3LAC13	Vandal	2
aruizc02	Shorty	5
Xuga	Vandal	2
Peruanopremium69	Vandal	2
LordTachMain	Bucky	4
David136502	Vandal	2
Badre	Vandal	2
Tomas	Operator	6
SAISU	Phantom	1
Pepe	Phantom	1
TwiS bylegend86	Vandal	2
ADC19	Phantom	1
guille73	Phantom	1
Se7en	Phantom	1
Mr Quake	Sheriff	3

The final data set consists of a sample of 23 players with their nickname, favourite weapon and the number associated to it.

3 Initial approach to the data

When seeing the data collected, it is pretty obvious that the Vandal and Phantom rifles are the more repeated chosen in this specific data set. To make it more visual and have a clearer idea of the data set, a circular graph was generated using *R*.
In the graph, attacked below, it is clear that the Vandal was the most repeated one, followed by Phantom and, with a lot of percentage difference, by Bucky and Sheriff.



Though for this sample the vandal wins by so much difference, what if we repeated the experiment 1000 times? Would it change? Can this help us to predict what our enemies will choose?
The code used to create this graph is can be found in section 8.

4 Non-parametric bootstrap

The method consists in:

- Using the frequency distribution of the n data values as our best guess of the population or probability distribution.
- Simulating the sampling from the population distribution obtaining the new bootstrap samples.
- For each sampling, calculating the sample statistic of interest.

4.1 Goal

Determine which is the most liked gun of all. To do so, we will generate 1000 random samples of 20 observations each, using the data collected.
The favourite one will correspond to the mean of the samples generated.

4.2 Estimation of the mean

Using the R codes in section 8, 1000 samples of 23 observations each have been generated only taking into account the guns that appear in the data set (the classic gun, for example, has not been taken into account, as it doesn't appear as anyone's favourite).

A fraction of the output is:

```
1.826087 2.434783 1.565217 2.521739 1.782609 2.5217397 2.043478 1.652174 2.043478
```

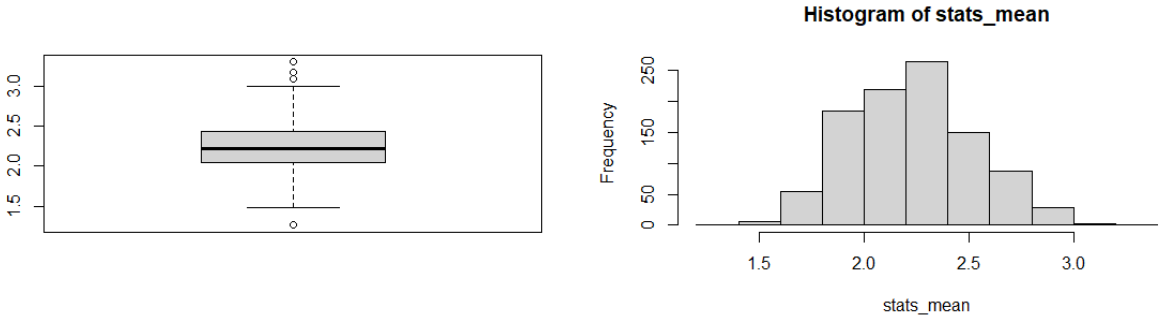
The mean of all the samples is: 2.22413.

The standard error is: 0.29743.

The confidence interval is: $[1.652174, 2.826087]$, which can also be expressed as: 2.2 ± 0.57.

4.2.1 Analysis

The graphics generated to analyse the data are the following:



(a) boxplot of the mean

(b) histogram of the mean

The boxplot shows that there are a few outliers in the mean, are caused by the values 5 and 6 that appear only 1 time each in the data set. On the other hand, the mean of all the datasets (the line in black) and the majority of the means land between 2.0 and 2.4. The histogram generated shows that que highest frequencies are, as in the boxplot, the ones between 2.0 and 2.4 (from 2.5 to 3.0 the frequency gets lower).

With the information above, we see that the standard error, 0.29, is small, indicating that the results are very close in value to the mean.

With all the data and graphics above, and also with the conficence interval 2.2 ± 0.57, we can confirm that the favourite gun in general is the one corresponding to the value 2 of the dataset, the Vandal, but in some cases it can be Sheriff, corresponding to the number 3.

4.3 Filtered data

When looking at the dataset, it is pretty clear that there are some weapons which are not usually picked (the ones only 1 person chose), those are: Guardian, Shorty and Operator. Since Shorty and Operator have the numbers 5 and 6 respectively, I ask: Would the mean significantly change if those values which appear only 1 time were removed?

To answer that question, those values are removed from the data set, which ends up looking like this:

Nickname	Favourite wea	Number associated
Jeke	Vandal	2
Ibra	Sheriff	3
Dori	Vandal	2
Duna	Vandal	2
Pp	Vandal	2
Marufa	Bucky	4
Tontolhaba77	Phantom	1
I3LAC13	Vandal	2
Xuga	Vandal	2
Peruanopremi	Vandal	2
LordTachMair	Bucky	4
David136502	Vandal	2
Badre	Vandal	2
SAISU	Phantom	1
Pepe	Phantom	1
TwiS bylegend	Vandal	2
ADC19	Phantom	1
guille73	Phantom	1
Se7en	Phantom	1
Mr Quake	Sheriff	3

Using the R codes in section 8, 1000 samples of 20 observations each have been generated, using the new dataset.

A fraction of the output is:

```
2.15 2.00 2.35 1.80 1.70 2.00 1.95 1.90 2.65 2.20
```

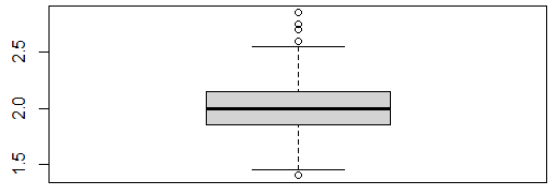
The mean of all the samples is: 2.004

The standard error is: 0.2020268.

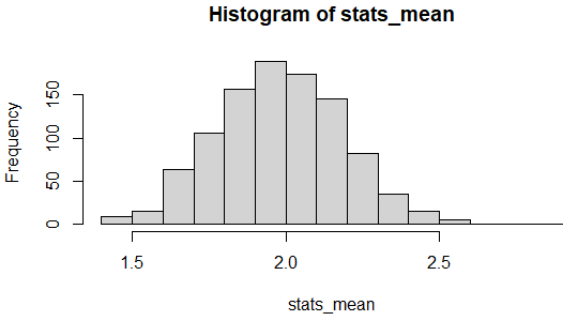
The confidence interval is: $[1.65, 2.40]$, which can also be expressed as: 2.025 ± 0.375.

4.3.1 Analysis

The graphics generated to analyse the data are the following:



(a) boxplot of the mean



(b) histogram of the mean

In the analysis of the previous section, 4.2.1, we saw that the mean was 2.2, but by getting rid of the outliers (Guardian, Shorty and Operator), the mean is now 2.004, and the standard error is even smaller, being 0.20. The confidence interval is now reduced to $[1.65, 2.40]$, the lower bound staying the same and the upper bound being 0.42 points lower. That means that the possibility of the Sheriff to be the favourite weapon (the number 3 is very unlikely, as for any value of the confidence interval the favourite one is the Vandal, 2.

The graphics shown above also prove that Vandal is the favourite for most people, as the histogram has the highest frequency is around 2.0, and the boxplot is also centered in 2.0, with only 4 outliers above the value 2.5 and 1 below 1.5.

4.4 Additional information

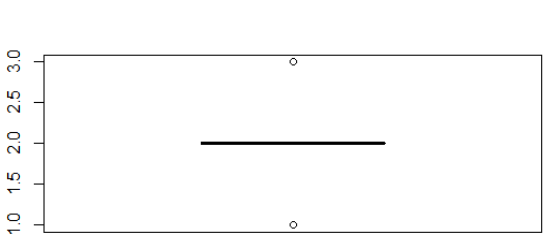
To see if the median was also around 2 as expected, I did a non-parametric bootstrap to estimate de median of 1000 samples of 23 observations each of the original dataset (containing the weapons that only appeared 1 time). The results were the following:
A fraction of the output is:

```
2 2 2 3 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2
```

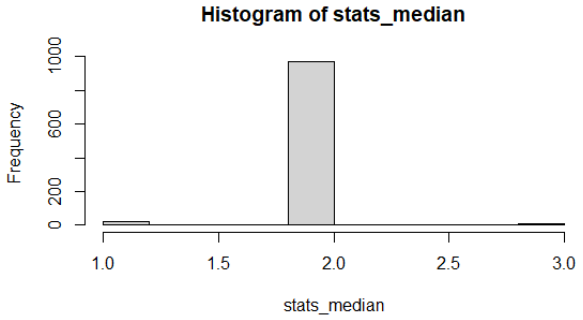
The mean of all the medians is: 1.987
The median of the medians is: 2
The standard error is: 0.1638837.
The confidence interval is: $[2, 2]$, which can also be expressed as: 2.

4.4.1 Analysis

The graphics generated to analyse the data are the following:



(a) boxplot of the median



(b) histogram of the median

Taking into account that below 2 there's only the Guardian (0) and the Phantom (1), whereas above it there are the Sheriff (3), the Bucky (4), the Shorty (5) and the Operator (6), we can assume that people choose as favourite weapons the Guardian, Phantom or Vandal more than they choose the other ones, as the median of the 1000 samples generated is 2 most of the time and when it's not, it is 1.
We confirm that by looking at the mean of all the medians, which is 1.987. The standard error is the lowest of the 3 analysis that have been done, being 0.16. That also supports that Vandal is the most picked out of all the weapons as favourite, with the Phantom following.
The graphics show that the frequency of 2 being the median is the highest by far, followed by 1.

5 Parametric bootstrap

Parametric bootstrapping assumes that the data comes from a known distribution with unknown parameters. The method consists in:

- Assume data come from random variables from a particular distribution (E.g., log-normal, Poisson, Normal, etc).
- Use data to estimate the parameters of the distribution (E.g., mean, variance, etc).
- Use a random number generator to create samples from the pattern distribution with the same size as original.
- Calculate the sample mean or the statistics of interest. Confidence intervals can be calculated from the quantiles.

5.1 Goal

In the section 4, we have estimated that the favourite gun for most people is the Vandal.
For this weapon, and assuming that in game the players always choose their favourite gun, the goal is to estimate the probability of at least one member of the enemy team to choose the Vandal.
Knowing this can be useful in game, because it helps you predict what the enemy team is going to do and what to get yourself in order to win, for more information, see `https://www.polygon.com/2020/4/6/21207403/list-of-all-weapons-prices-rifles-snipers-op-awp`.

5.2 Estimation of the probability p

As we consider the enemy team getting a Vandal a success, we assume that the distribution of our data is a binomial.

To start the method, we need an estimation of the probability p, where:

P = Probability of at least one player of the enemy team to choose Vandal.

P = 1 - Probability of no one of the enemy team choosing Vandal.

P(choosing Vandal) = $\frac{Vandal}{Total} = \frac{10}{23} = 0.4347826$

P(not choosing Vandal) = 1 - P(choosing Vandal) = 0.5652174

p = 1 - (P(not choosing Vandal))^5, as there are 5 members in the team.

p = 0.942313

By generating 1000 random binomial samples with 20 observations each and p = 0.942313, we get to the following results:

Fraction of the probabilities of each sample:

```
0.85 0.95 1.00 1.00 0.95 0.85 0.95 0.90 0.90 0.95 1.00 0.95 0.85
```

Mean of the probabilities: 0.94365.

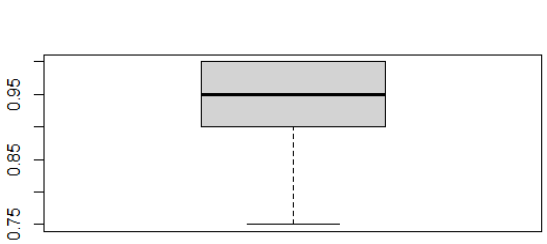
Bias: 0.00133695.

Estimated Standard Error: 0.05044524.

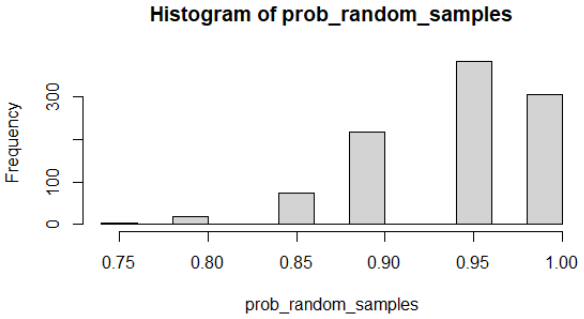
Confidence Interval: $[0.85, 1]$, which can also be expressed as: 0.925 ± 0.075.

5.2.1 Analysis

The graphics generated to analyse the data are the following:



(a) boxplot of p



(b) histogram of p

As we can see by calculating the Bias, which is 0.00133695, the value of the parameter p resulting of the bootstrap is really similar to the first estimation. The standard error is really small, which indicates that the probabilities of the 1000 samples are, in general, really close to the mean found, as we can see in both the boxplot and the histogram shown above.

The confidence interval given is really small, and the probability very high, meaning that we can most likely assume that at least one person of the enemy team is going to have a Vandal.

It is interesting to mention that the case with the lowest probability is 0.75, which is still really high.

6 Conclusions

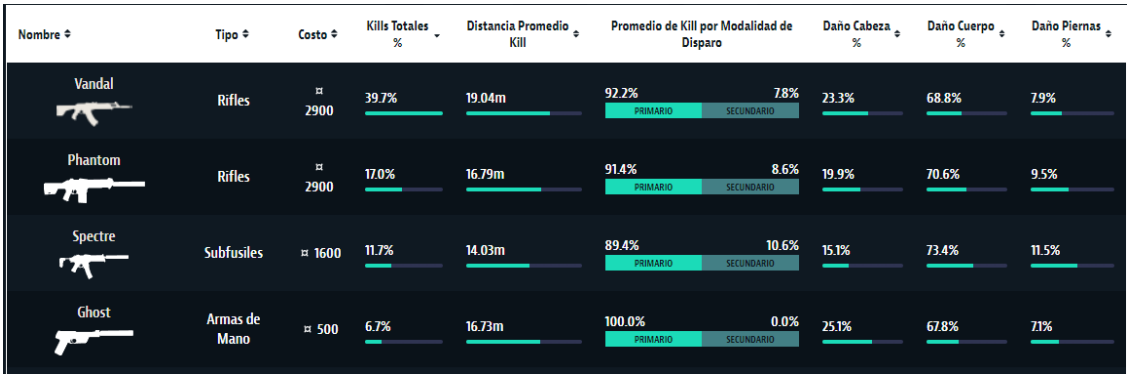
Based on the analysis of the first section, 4, I can affirm that Vandal is the most liked gun of the game.

The most reliable analysis of the three done with the non-parametric plot is the first one, done by estimating the mean of all the original data set, because it uses all the data and represents the reality better.

Assuming that people use their favourite weapon, based on the analysis done in the section 5, I can affirm that, with a probability between 0.85 and 1, at least one member of the enemy team will use a Vandal in the game.

When professional players were asked to rank all the guns in 4 different categories, S, A, B, C, Vandal appeared in the first class, S, as it has one of the highest cost/effectiveness ratio, so it makes sense that the probability of a team to have at least one is so high.

When looking at Valorant Statistics, we see that 39% of the total kills come from Vandals, being the second one the Phantom with 17%. This relates to the non-parametric bootstrap estimating the median, which was Vandal and, in second place, came the Phantom. We also can conclude that the most used weapons are rifles since both Vandal and Phantom are so.



This information can help you win the game, since you can buy more things that are effective against a Vandal, or pick a different Agent than if you did not know it. For example, since the Vandal is one of the weapons that produce more damage, it would be useful to have a healer in the team.

7 Bibliography

Parametric Bootstrap presentation from UAB: https://e-aules.uab.cat/2021-22/pluginfile.php/911701/mod_resource/content/1/4-%20Parametric%20Bootstrap_22.pdf

Non-Parametric Bootstrap presentation from UAB: https://e-aules.uab.cat/2021-22/pluginfile.php/937806/mod_resource/content/1/6-%20Non-Parametric%20Bootstrap_2022.pdf

List of weapons from valorant:

https://www.polygon.com/2020/4/6/21207403/list-of-all-weapons-prices-rifles-snipers-op-awp

Weapons statistics:

https://www.valorbuff.com/weapons

Weapon Tier List:

https://mobalytics.gg/blog/valorant-weapon-tier-list-best-guns/

8 R scripts

8.1 Initial approach to the data set

```
library("readxl")

path="C:/Users/onasa/Documents/2n2nsemestre/Dades Complexes/valorant.xlsx"
data = read_excel(path)
data

guardian = 0
phantom = 0
vandal = 0
sheriff = 0
bucky = 0
shorty = 0
operator = 0
others = 0

for (i in data$`Number associated`){
  if (i==0){
    guardian = guardian + 1
  } else if (i==1){
    phantom = phantom + 1
  } else if (i==2){
    vandal = vandal + 1
  } else if (i==3){
    sheriff = sheriff + 1
  } else if (i==4){
    bucky = bucky + 1
  } else if (i==5){
    shorty = shorty + 1
  } else if (i==6){
    operator = operator + 1
  } else{
    others = others + 1
  }

}
#Others is not shown as it has 0% probability
slices <- c(guardian, phantom, vandal, sheriff, bucky, shorty, operator)
lbls <- c("Guardian", "Phantom", "Vandal", "Sheriff", "Bucky", "Shorty", "Operator")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices,labels = lbls, col=rainbow(length(lbls)),
    main="Favourite weapon in Valorant")
```

8.2 Non-parametric bootstrap

8.2.1 Initial mean

```
non_param__mean = function(x){
x = mean(sample(x, size = length(x), replace = TRUE))
return (x)
}

stats_mean = replicate(1000, non_param__mean(data$`Number associated`))
stats_mean
final_mean = mean(stats_mean)
sd_mean = sd(stats_mean)
sd_mean
CImean = quantile(stats_mean, probs = c(0.025, 0.975))
CImean

boxplot (stats_mean)
hist(stats_mean)
```

8.2.2 Mean filtered data set

```
path_relevant="C:/Users/onasa/Documents/2n2nsemestre/Dades Complexes/valorant_filtrado.xlsx"
data_relevant = read_excel(path_relevant)
data_relevant

non_param__mean = function(x){
  x = mean(sample(x, size = length(x), replace = TRUE))
  return (x)
}

stats_mean = replicate(1000, non_param__mean(data_relevant$`Number associated`))
stats_mean
mean(stats_mean)
sd_mean = sd(stats_mean)
sd_mean
CImean = quantile(stats_mean, probs = c(0.025, 0.975))
CImean
boxplot (stats_mean)
hist(stats_mean)
```

8.2.3 Median

```
non_param_median = function(x){
x = median(sample(x, size = length(x), replace = TRUE))
return (x)
}

stats_median = replicate(1000, non_param_median(data$`Number associated`))
stats_median
mean(stats_median)
median(stats_median)
sd_median = sd(stats_median)
sd_median
CImedian = quantile(stats_median, probs = c(0.025, 0.975))
CImedian
hist(stats_median)
boxplot(stats_median)
```

8.3 Parametric bootstrap

```
p = vandal/length(data$Nickname)
p = 1-(1-p)^5

prob_random_samples = rbinom(1000,20,p)/20
prob_random_samples
mitjana = mean(prob_random_samples)
mitjana

Bias = mitjana-p
Bias

sd(prob_random_samples)

quantile(prob_random_samples, probs = c(0.025, 0.975))

boxplot (prob_random_samples)
hist(prob_random_samples)
```