DIAPO 1 (INTRODUCTION)

Hello, I am Cinta Arnau and I am going to analyze the dataset I have been working with regarding multiple linear regression.

DIAPO 2 (LoL)

First of all, what is this dataset about? It's about the popular videogame "League of Legends". LoL is a game where 5 players team up against another team of 5 players in an arena, where the objective is to destroy the opponent's nexus. During the game, it is essential that you kill the opponent's minions that are coming, which make you earn you some gold, as well as performing kills on your enemies, which will grant you gold as well so you can buy items that will help you and your team win.

For this dataset, I have decided to choose as our regressors the number of kills performed and the total number of minions killed and, as the response, I have decided to choose the total gold earned in that same game.

Therefore, what I was expecting from this dataset was to check if performing kills on your enemies and minions is what makes you earn gold in the game, which would confirm my initial assumptions, or if there is actually no relationship between those regressors I have chosen and the gold earned.

DIAPO 3 (ANALYSIS OF PLOTS)

Analyzing the results it can be seen how, regarding the regressor x1, the more kills performed, the more gold we get. The same happens with the regressor x2, which shows that the more minions killed, the more total gold obtained. Therefore, my initial assumptions have been confirmed, as I have obtained what I expected, this being that the number of kills that we get from both enemies and minions directly influence how much total gold we get, as both regressors and response are directly proportional.

This is also confirmed by the values computed for b1 and b2, which happen to be b1 = 451.6605 and b2 = 22.1336, which just by sight seem to be enough to influence the response.

It can also be seen in all these plots how the data seems to be well defined by the given plane I have obtained, which is confirmed by the coefficient of determination that has been computed.

DIAPO 4 (EXAMPLE OF LoL)

This is an example of the results of a game. The last column represents the gold that someone has earned, before that the minions that someone has killed and before that, on the first column we can see the enemy kills that someone has performed.
By this image it can be concluded that enemy kills give more gold than minion kills. For example, there's this user (bkardi liri) that has 21 kills and 179 minion kills, and then this other (klausibuff05) that has 10 kills and 227 minion kills. Despite the fact that the second

one has killed many more minions than the first, we can see that bkardi has earned more gold than klausibuff05, confirming that enemy kills are more relevant.

DIAPO 5 (ANALYSIS OF RESULTS)

As we know from theory, the general formula for obtaining the response based on the obtained regression plane is the shown shown in this slide, in which the values of the betas are the following: b0 = 8903,5; b1 = 451.6605 and b2 = 22.1336. As it can be seen, both b1 and b2 are positive and as it has been previously observed from the plot, it can be said that both of them have a positive impact on the output y. That is, both b1 and b2 are directly proportional to y. However, we still have to prove if the two regressors do actually influence our response, which I have done by doing both the F-test and the T-test.

I have also computed the coefficient of determination, which is the one that concludes if the data is well fitted by the plane. As it can be seen, the value of the coefficient of determination is 0.8906. Since it is greater than 0.8, it means that the plane we have obtained fits very well the data from our dataset, which makes sense as this is what I was expecting when observing the graphic representation of the data.

For the F-test, since I want to check if the model accounts significantly for the variability of the data, I will have as the null hypothesis that all of our slopes have a value of 0, that is, none of the regressors influence the response, against the alternative hypothesis that there exists a value of beta which is greater than 0, that is that at least one of the regressors influences the response.

In this case, when performing the F-test I have obtained that the value of F that is observed from the data is 3046, while the critical value of F is 3.0078. Therefore, since the observed value of F is greater than the critical value of F, we'll reject the null hypothesis of the test with a significance level of alpha = 0.05, that is, there exists at least one regressor influencing the response.

I have also computed the p-value in order to know the probability of obtaining a more extreme value of F observed than the one I have obtained. Since the F observed is very large and it falls very far from the critical value of F, it makes sense that the p-value is 0, as the probability of obtaining a more extreme value is extremely low.

I have also computed two T-tests, one for each slope value, that is, b1 and b2, in order to know which regressor is the one influencing the response, as the T-test tests each slope individually.

First of all, I have computed the acceptance region for the tests, which as it can be seen is the interval between -1,9631 and 1.9631.

For the first regressor x1, which is the number of kills performed, I have obtained a value of t_observed of 73,8028, which as it falls outside the acceptance interval, I'll reject the null hypothesis.

Same happens with the regressor x2 which represents the number of minions killed, for which I have obtained a value of t_observed of 26,8801, which also falls outside the acceptance region and therefore I can reject the null hypothesis.

Same as when I computed the p-value for the F-test, I have also obtained a p-value of 0 for each of our individual T-tests, which makes sense as the values I obtained are already quite extreme and the probability of obtaining a more extreme one is very low.

Since in both T-tests performed I have rejected the null hypothesis, I can therefore conclude that both the regressors influence the response. For this reason, once again my initial assumptions have been confirmed and what I expected from the very beginning is right, which is that both the number of kills performed and the number of minions killed influence the number of gold obtained positively, as they are directly proportional.

DIAPO 6 (ANALYSIS OF RESULTS) + DIAPO 7 (EXAMPLE OF EXTENSIONS)

Should I have gotten more samples to arrive at this conclusion? I have considered that there is no need of getting more samples than the ones we already have as the total data we have is already quite large (it is 750 records) and as it has been seen in our model, the data seems to be quite predictable using the plane I have obtained, as the data is very well fitted, which has been seen by observing the value of the coefficient of determination.

However, are these two the only methods of obtaining gold in League of Legends? No, they are not. Therefore, I believe it would have been useful to take into consideration more regressors such as the following represented by this image, which show the towers that have been destroyed, the inhibitors that have been destroyed (a type of structure), and three types of special NPC creatures that the team has killed.

Slide 7

These bonuses significantly impact the economy of the players that benefit from them and that can easily be seen by the general difference of gold that both teams have. Also, I should take into consideration other regressors such as the number of assists performed or which type of items have been bought, as all of these other extensions influence the number of gold obtained as performing each of them gives you an extra number of gold.

Another factor that influences gold gain are the assists that players have and also a steady amount of gold per second that every player in the match earns, despite any action they do.