DSC 530

Qua Tran

2022-02-24

Does demographical factors (age, gender, health score, test preparedness and family relationship) effect overall test scores?

The dataset contains 649 students with multiple variables in relation to multiple test scores (G1, G2 and G3). I ran a function to add up all these scores with a maximum perfect score of 60, with 20 points maximum score for each test. The purpose of this project was to see if age + other factors effect students total scores in general across the board. For starters, this data set was very easy to migrate because the factors are already classified and weighted in numerical figures. I did not have to convert anything into binary form. When I looked at the raw dataset, it was clear that total scores are widely ranged from the low to the top portion which is to be expected. In theory, we expect that some students will do very well, and some students will do poorly. This was confirmed through the total test scores vs student count distribution. There was a small shift of the 50% cumulative percentage towards the right a bit, because often, students are not shooting for the 30 out of 60-point spread. Overall, this was a bell-shaped curve.

I looked deeper into the grades and noticed that students who tends to score low, score low across all grades. (G1, G2 and G3). This was confirmed through the scatterplot.

The one key factor that can impact this analysis is my *n* value. I only looked at < 650 students and if this figure were to increase, I should see a more centered distribution. This specific dataset only consisted of students with ages 15 to 22, but if we really wanted to see if total grades correlates to age, we would have to expand this variable. For example, age 15 had over 100 students but age 22 had only 1 student. It is not significant enough to make any assumptions if age really correlates to total score. However, what I did look at was the following:

StudyTime: 1 Average Total Score: 32

StudyTime: 2 Average Total Score: 35

StudyTime: 3 Average Total Score: 38

StudyTime: 4 Average Total Score: 36

Absences

Less than 10 Average Total score: 35

>10 and <20 Total score: 32

>20 and <30 Total score: 28.5

> 30 Not valid due to only 2 students

Students who were rated with a study time of 3 and 4 scores higher than students who was rated at 1. Students who had fewer absences, will always do better on tests to achieve higher scores.

The understanding and purpose of these analysis I do fully understand. I tend to struggle on applying the right coding to reflect the true result. I believe with more practice in future classes, I can further implement these analyses.