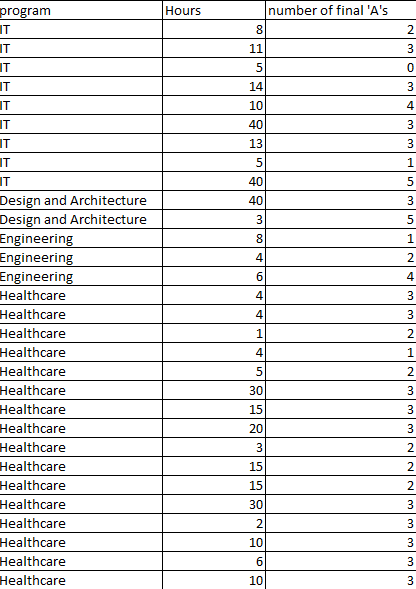
**Introduction:** the topic is to determine is there relationship between time spent studying and number of final 'A's for students of some programs.

For the survey was made blanks and proposed to the college students on Thursday (04/18) between 11:00 AM and 2:00 PM at the entrance of the Parr Center.

For the survey, a voluntary response sample was chosen, with one exception: people started answering the questions without prior knowledge of the questions and the topic, and no one stopped during the process. Thus, positive or negative opinions about the topic and the questions did not influence the survey. The sample is biased because students were selected who happened to be on the campus at that day and time, who went to the Parr Center, and who agreed to take the survey. A better approach would be to negotiate with the college administration to have professors distribute survey forms to students in class or through Brightspace. In this case, fewer people would refuse to take the survey, and the results would be more random.

## The Survey Results:



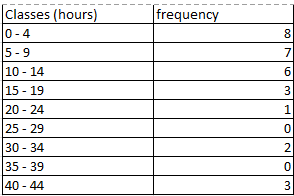
# Part A

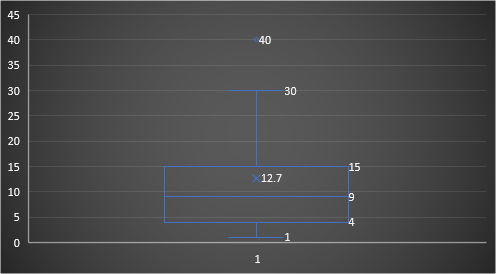
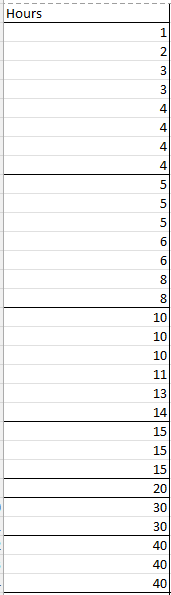
# 

The mode: “Healthcare”

**Conclusion:** As the pie chart shows, more than half (53.33%) of the students study in the "Healthcare" program, so it is the mode because the probability of selecting a student from this program from the sample is the highest. There are no students in the "Business and Accounting" and "Construction and Trades" programs. As Pareto chart shows the most popular programs are "Healthcare" and "IT" (it means that 33.33% of programs have 83.33% of surveyed students, it is close to Pareto principle (80/20)), while the least popular ones are "Business and Accounting" and "Construction and Trades".

# Part B





The Mean: 12.7

The median: 9

The range: 39

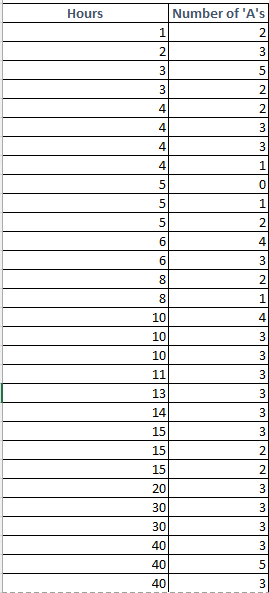
The standard deviation: 11.72133334

The outliers: 40

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| The five number summary: | | | | |
| min | Q1 | Q2 | Q3 | max |
| 1 | 4 | 9 | 15 | 40 |

**Conclusion:** The mean is 12.7, the median is 9, based on this data, the graph is skewed to the right. 40 is an outlier and it has a very strong influence on the mean. I was surprised by the outlier because outliers have unusual values. The interesting thing I noticed is that the number of class elements is the largest at the beginning of the class, then decreases (the minimum at 20- 24 class), and then increases unevenly (with breaks).

# Part C



The "Hours" is an explanatory variable and "Number of 'A's" is response variable.

The correlation coefficient: 0.347102429

regression equation: 2.25547775+0.03237708\*x (Slope: 0.03237708, y-intercept: 2.25547775)

**Conclusion:** The "Hours" is an explanatory variable and "number of 'A's" is response variable. There is a linear correlation with a correlation coefficient 0.347102429, this doesn't look like what I expected, as I expected a stronger relationship between the variables. A least squares equation (regression line) shows the relationship between an explanatory variable (independent variable) and a response variable (dependent variable). In the situation the slope is positive, so the variables have direct relationship, the y-intercept in the situation shows that if a student study 0 hours a week he/she would have about 2.26 final 'A's for the previous semester, however a student can not have 2.26 'A's because the amount of 'A's can only be an integer, so the y-intercept can not be interpreted. The most interesting thing is that the relationships between the variables are so weak, and the conclusion that the amount of study hours have little effect on the number of final A's. Based on the results, I can predict that students who do not spend much time studying will likely have high grades, almost the same as students who study a lot.

# Part D

1. The explanatory variable is a variable that, when it changes, causes the value of the response variable to change. The response variable is a value that changes depending on the value of the explanatory variable. In this case "weight" is an explanatory variable, and "MPG" is a response variable.
2. The equation of the regression line is the equation that shows the relationship between an explanatory variable (independent variable) and a response variable (dependent variable). In this case it is y = -0.006x + 47.942.
3. The correlation coefficient (r) describes the degree of linear relationship between two variables. In this case it is -0.85369.
4. The slope is negative (-0.006), so the variables have inverse relationship, the y-intercept shows that with weight 0 the MPG is 47.942, the y-intercept has not valid interpretation because a vehicle can not weight 0.