

Toledo Test Analysis: Josh, Jefferson, Jose

Introduction

The Toledo Test is a chemistry placement exam that is designed to evaluate how well a student is prepared for college level introductory chemistry.

It includes 20 math/chemistry questions with content up to College Algebra.

We were given 4 quantitative variables which include score, gpa, math grade, and chemistry grade. Additional information was also provided for some students.

Our Goals

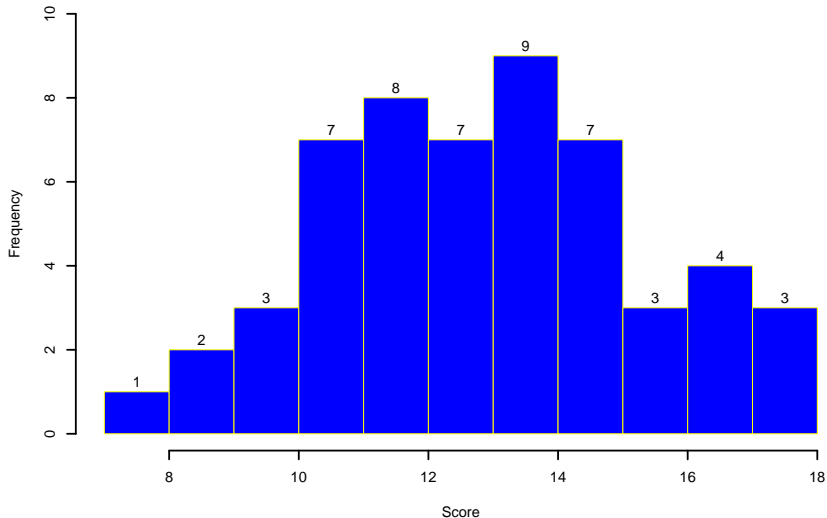
We wanted to analyze the data and see if we came up with any significant correlations.

We also wanted to see what will happen if we change the cut-off and find out what other information would be useful.

We wanted to learn new techniques when analyzing data in R.

Visualizing the Distribution

Distribution of All Exam Scores



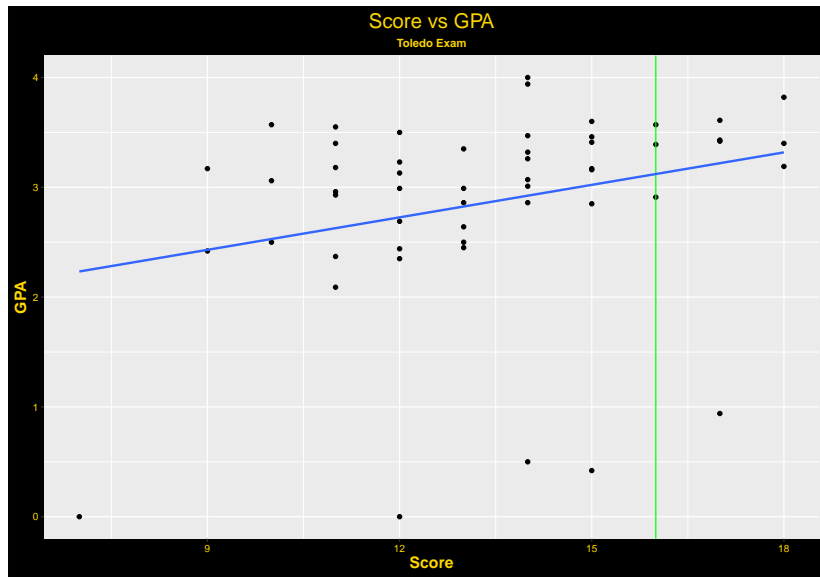
Correlation Coefficient

Our correlation coefficient is a number between -1 and 1. A correlation of -1 would be a perfect negative correlation and 1 would be a perfect positive correlation. The closer the coefficient is to 0, the less correlated the two variables are. We used the Pearson method of correlation test. The formula is below.

$$r = \frac{\sum (x-m_x)(y-m_y)}{\sqrt{\sum (x-m_x)^2 \sum (y-m_y)^2}}$$

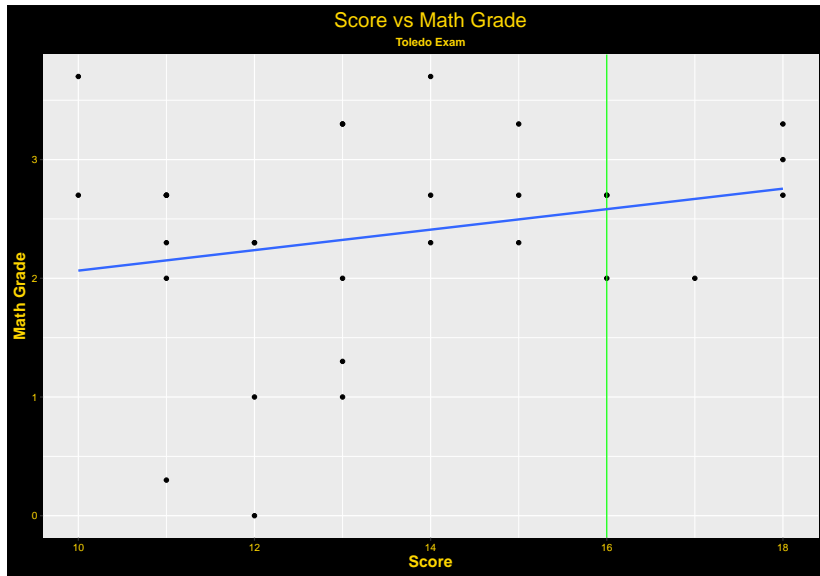
m_x and m_y are the means of x and y variables

Correlation between Test Score and GPA



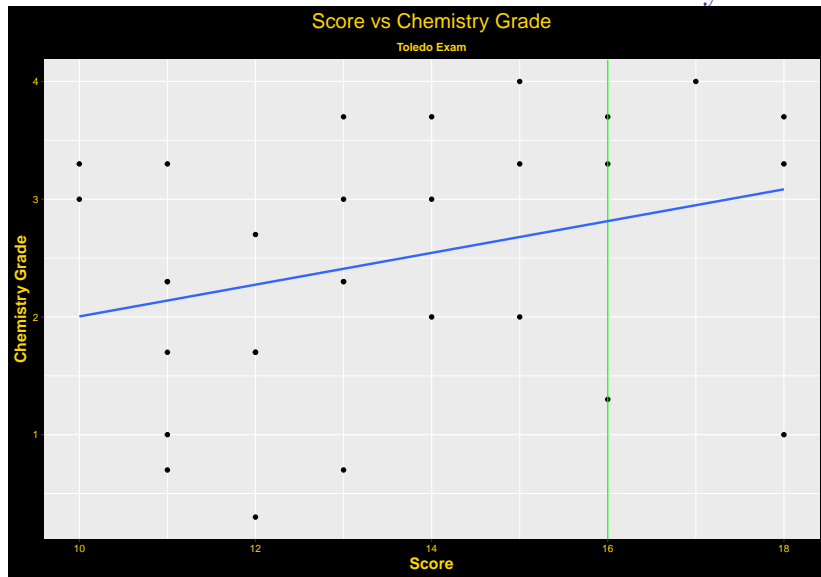
Correlation Coefficient = .27

Correlation Between Test Score and Math Grade?



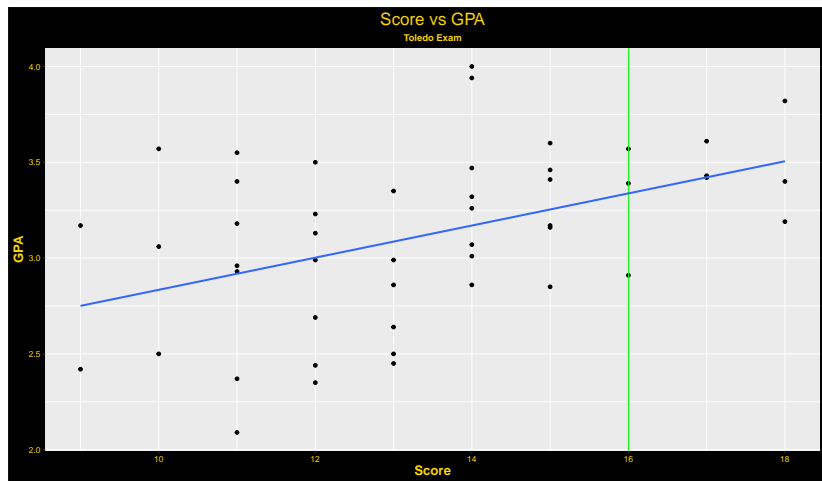
$$\text{cor} = .23$$

Correlation between Test Score and Chemistry Grade



$$\text{cor} = .30$$

What if we took out the outliers? How would this affect correlation?



$$\text{cor} = .46$$

What if- Analysis

The cutoff for passing is 16. Filtering out the passing grades, we can find out what our passing rate is.

```
pass <- filter(toledo, Score > 15)
```

```
passpct <- nrow(pass)/nrow(toledo)  
passpct
```

```
## [1] 0.1851852
```

We can see that the current passing rate for the Toledo Exam is pretty low at approx 18.5%

What if we wanted a certain passing rate? Here we wanted to see what the passing rates would be if the cutoff of the exam was changed.

```
pass1 <- filter(toledo, Score > 14)
```

```
passpct <- nrow(pass1)/nrow(toledo)  
passpct
```

```
## [1] 0.3148148
```

```
pass2 <- filter(toledo, Score > 13)
```

```
passpct <- nrow(pass2)/nrow(toledo)  
passpct
```

```
## [1] 0.4814815
```

Why is the passing rate so low?

We will look at all 3 ways to get into Chemistry. This data is from the App State website.

SAT scores - must be 550 or greater

25th percentile - 540

75th percentile - 630

Around 70 percent of people have an SAT score high enough to get into chemistry

ACT Scores - must be 22 or greater

25th percentile - 22

75th percentile - 26

75 % of people have an ACT score high enough to get into chemistry

Because of this information above, we can see that the people that are taking the exam are already in the bottom 30% of exam scores.

Is the Toledo Exam doing what we want it to?

The Toledo Exam could be thought of as a way to weed out people from taking chemistry. Underprepared people could be signing up for chemistry. The Toledo Exam prevents this for 82% of people that take it and makes them take remedial math. We will look and see what the students did after the Toledo Exam.

```
names(toledo)[10]<-paste("fall")
notenr <- filter(toledo, fall == "not enrolled")
enr <- filter(toledo, fall == "enrolled")
nrow(notenr)/(nrow(notenr) + nrow(enr))
```

```
## [1] 0.3333333
```

From the App State website, we see that they retain 88% of Freshamn which means 12% drop out or transfer. Above, we can see that 33% of Toledo participants are not enrolled after a year.

Another variable we can look at is how many people went to take chemistry.

Only about 60% of people who failed the Toledo went on to take chemistry. This means the other 40% either switched their major or dropped out. We think that Chemistry grades may improve because of barriers that are put up to enter chemistry.

Final Thoughts

The Toledo's passing rate may be so low because of the group that has to take it.

The Toledo exam may be causing people to change their majors at a more significant rate than normal.

Additional data that could of helped us woud be an analysis on each question of the Toledo Test to see if any questions are missed more than others. Also, additional years of data could help us reach more solid conclusions.

The toledo exam may not be a good indicator for student's performance in chemistry.

Next Semester...

- Work with big datasets
- Work more with Toledo Test data
- Work with different methods such as multiregression
- Work with Time Series
- Work with Shiny