Rafael Martins Guimarães

Foundations of Statistics

0ctober 27th, 2016

Analyzing writing scores

**Overview**

We are looking on how the writing ability from an individual is affected by different variables, this is a good opportunity to study if there is a significant difference between different group like gender, race and other factors. After analyzing and seeing how each variable is related or affects each other, we will run a regression model in this data set to see what answer we can get and also this is an opportunity to compare regression models.

**Some analysis with the writing score variable**

First we compared the scores of writing with other academic studies, math, science and social studies. The result showed us that although that there is a significant correlation between the variables, none of them is highly correlated. That means that the scores affect each other, but there might be other factors that contribute more for the differences between writing score. Also an interesting fact is that the means of the grades are very similar, between 52 and 51.

Second, is to see how are the writing scores between genders and race. The test showed us (t = -3.7341, df = 198, p < 0.0001) that in this data set, the female individual has a higher score than the male. The female score is 54.99083 while the male score is 50.12088, a difference of 4.869947. It’s seems that the female population is more concern or dedicates more to their writing abilities. That is really interesting, because we have a dominated male academic population around the world, and only over a few decades of changes in the role of the woman in the society, we see a positive result.

Third, is to compare different races. The result (t = -3.1711, df = 198, p < 0.0009) showed us that a white person has a higher score than people from other races. A white person has an average score of 54.05 while people from other races have an average score of 49.4. This happens because usually a white population has a higher income, meaning that they are able to provide themselves with a higher education.

Now we compare the writing score between different Socio-Economic Status level. This result (f = 4.97, df = 199, p < 0.0078) is really interesting, because it shows us that the difference between individuals of lower and middle classes is not that big, only 1.31, that is because they might frequent the same public school, but if you compare these scores with the upper class, we see a higher score, 55. That is possible, because people from higher classes can afford high quality education.

We can also see the score of writing between different programs. The outcome (f = 21.27, df = 199, p < 0.0000) shows us that people with a higher academic level have a better writing abilities, which makes sense, because they usually are used to write academic papers, essays, reports and many others. The average mean score of an academic individual is 56.25, the highest mean so far.

**How other variables affect each other**

We are analyzing the writing score, but it is also interesting to see how the other variables affect each other. First, let see how people from different races are distributed in the Socio-Economic Status level. By doing this (chi2(2) = 14.79, df = 200, p < 0.001) we can see there is a higher concentration of white people in the upper classes than the other races. It also shows us that there is a balance of between lower and middle classes in the other races, and only a small portion in the higher class.

Second, let’s see how people from different school type fit themselves in each ty of program. We see (chi2(2) = 9.27, df = 200, p < 0.01) that although there is a higher number of people from the public school with an academic achievement, the individuals from the private school’s proportion is way higher, 75%, compare to the public school’s individuals, 48%.

**Regression model**

To have a better notion of how writing score is affect by different variables, let’s do a regression model. This regression model will have the writing score as the dependent variable and race, gender, social level, school type, program type and the other scores as the independent variables.

The first model we generate shows us interesting numbers, but after analyzing it, some key values from the model were an issue (see table 1). To overcome this problem, we have to execute some tests to get a better fitting model. Some variables were removed from the model, social level, race and others. The only significant variables were the other scores and gender. The last model tells us that in every 1 point of increase in your writing score would also increase your other grades, also that if you are a woman, your score would increase by more than 5 points. It is interesting to see how the other variables that we thought were important will have a high p-value.

**Conclusion**

All in all, the hypothesis we can get by analyzing this data set, is that before, when we were studying the effects of all the variables in the data set with the writing score, we thought that all the variable would have a significant impact on the writing score, but after doing the regression model, we see that only the other academic studies and your gender had an impact in the writing score.

First the other academic studies, we can say that if an individual is dedicating himself to get a better writing score, he will also have dedicated his time for the other academic studies. And for the gender we can say that even today, with the changes of the role a woman has in the modern society, if she wants to achieve success in a career or studies, she will have to outcome the man. That may be one of the reason of why women has a better score that men. But as I said before this is only a hypothesis based on a small population data. To have a better idea of how writing scores are affected by different types of variables, we need more data and a bigger population set, to finally have a better hypothesis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table1: Logistic Regression predicting writing score | | | | | |
|  |  |  |  |  |  |
|  | MODEL 1 | |  | MODEL 2 | |
|  | Coef |  |  | Coef |  |
| math | 0.24 | \*\* |  | 0.28 | \*\*\* |
| science | 0.29 | \*\*\* |  | 0.28 | \*\*\* |
| socst | 0.25 | \*\*\* |  | 0.27 | \*\*\* |
| female | 5.31 | \*\*\* |  | 5.43 | \*\*\* |
| white01 | 0.10 |  |  |  |  |
| lowses | 0.84 |  |  |  |  |
| midses | -0.23 |  |  |  |  |
| privateschool01 | 1.06 |  |  |  |  |
| general01 | -1.33 |  |  |  |  |
| vocational01 | -1.90 |  |  |  |  |
| \_cons | 9.33 | \*\* |  | \*\* |  |
|  |  |  |  |  |  |
| AIC | 1302.94 |  |  | 1295.91 |  |
| BIC | 1339.22 |  |  | 1312.40 |  |
|  |  |  |  |  |  |
| Key: \*p < .05, \*\*p < .01, \*p < .001 | |  |  |  |  |