Fight SAT with statistical weapons

Sei Masuoka

Introduction

Have you wavered on those two answers on a multiple choice test? I certainly have, and its not a fun experience. I keep on thinking about that one question, and I can not concentrate on the other question. So, I have decided to research some methods for guessing on five-choice multiple choice questions. I looked at some multiple choice questions from the SAT practice books, and used them as a sample for data to find some pattern in the answers.

I first looked for any pattern within the sample data I took. Then, I created a couple of method of guessing, which were supported by the data I took, to be better than a completely random guessing. Then I tested the method out on actual SAT tests taken from the book, "10 Real SATs."

Figure 1. Photograph of students taking multiple choice tests.

Materials and methods

• Multiple choice question books (Cited in "Literature Cited")

• I used chi-square test to prove that there was a difference in the probability of some choices and patterns. Chi-square-tests are used to prove that the proportion between two data sets are statistically different. In a chi-square test, you state that a **null hypothesis**, which is a hypothesis you are trying to disprove, is true. You also state the alternative hypothesis, which is what you want to prove. Then, you look at your sample data, and see by how much chance it can occur if the null hypothesis was true. If the percentage is low enough, you can reject the null hypothesis, and prove yourself right.

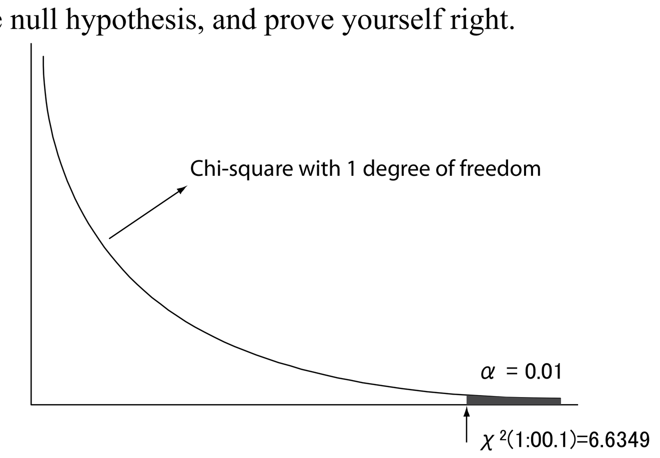

Figure 2a. This graph shows a chi-square curve with 1 degree of freedom. This curve is an approximation to the statistic of chi-squared (χ-squared) and the p-value is the probability that the chi-squared is larger than what you got.

• I used Excel extensively for this project. I created many Excel macros for easy and accurate data entry and statistical processing.

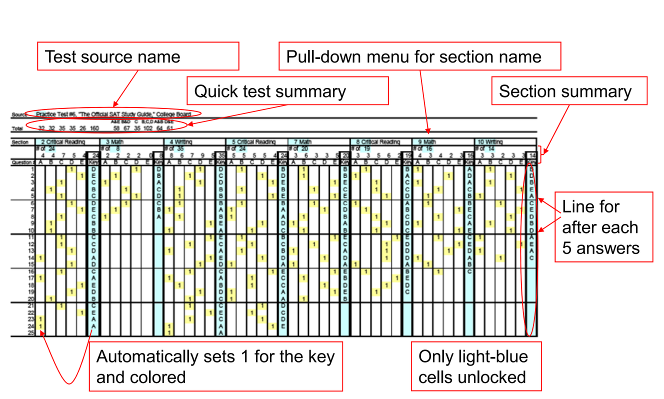

Figure 2b. Excel sheet for data entry. For example, I made macros for Excel so that whenever I type in an answer, the correct space is highlighted and gets the value 1.

Results

Analysis

I took 3520 multiple choice questions, which is equivalent to 22 SAT tests, as samples. When I first looked at the data, there were few things that hit me. They are:

- Both A and E had low scores compared to others
- Both B and D had high scores compared to others
- C had a pretty high score, but not as high as B and D

This led me to try a Chi-square-test on B, C, and D versus A and E. My null hypothesis was that probability for B, C, and D was 3/5, and the probability of A and E was 2/5. Then, I stated the alternative hypothesis, which said that the probability for B, C, and D was higher than 3/5, and the probability of A and E was lower than 2/5. So, I did the math and the Chi-square-test gave me a p-value of less than .0001. This means that the probability of a sample my size, 3520, yielding me a ratio of 1285 to 2235 when the true ratio of the data is 2:3 is less than .0001. **This is significant at the 1% level. Therefore, my data supports that the B, C, and D have a higher probability than 3/5 and A and E have a lower probability than 2/5.**

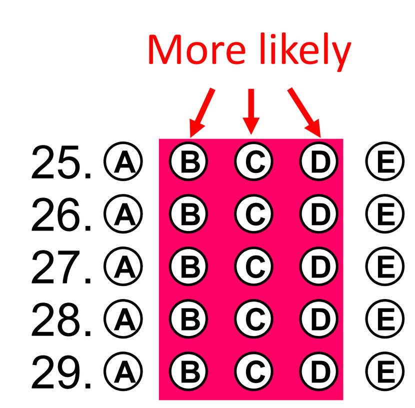

Figure 3a. B, C, and D have a higher chance than A and E, so when guessing, use B, C, or D for a better chance!

Then, I decided to look at answers when the answer before it is already known (**Fig. 3b**). If the probability of getting any letter after a certain letter is the same, then the probability of getting the same letter as the answer before should be 1/5. So, I set up my null hypothesis again, which is that the P(Same letters) = 1/5 and P(Not same letters) = 4/5. Then I set up the alternative hypothesis, which states P(Same letters) < 1/5, and P(Not same letters) > 4/5. I used the chi-square test again, and I found that the χ^2 = 43100.48. With one degree of freedom, this means that the p-value is less than .0001. This means that if the probability of getting two same letters in a row really is 1/5 and the probability of not getting two same letter in a row is 4/5, then the probability that I get a ratio of 522 : 2822 is less than .0001. **This is significant at the 1% level. Therefore, the probability of getting two of the same letters in a row is less than 1/5 and the chance of not getting two of the same letters in a row is more than 4/5.**

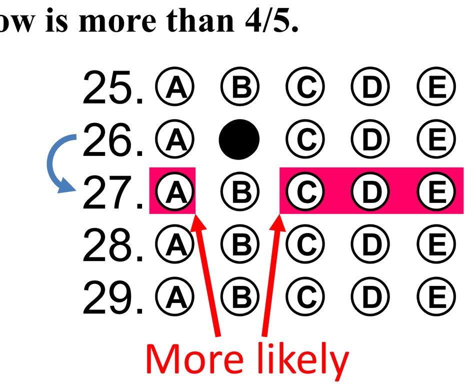

Figure 3b. The chance of two same answers occurring in a row is less than 1/5, so pick the other ones for a better chance!

```
=CHAR(65+INT(RAND()*5))
=CHAR(66+INT(RAND()*3))
=IF(EC9=0, "X", IF(ROW(ED10)=EVEN(ROW(ED10)),
  EC9,CHAR(MOD(CODE(EC8)+1+INT(RAND()*4),5)+65)))
```

Figure 4. Excel macros for test taking simulation. The top one is the Excel macro to produce completely random guesses, the second one is for the B, C, D only method, and the last one is for the no two same answers in a row method.

Test Taking Simulation

So, I used Ten Real SATs with three methods: the **completely random** method, **B, C, D only** method, and **no two same answers in a row** method. For the methods, I picked 8 real SAT tests.

For the **B, C, D only** method, I randomly picked either B, C, or D as an answer to all the questions. Then, I compared the number of answers that I got correct using this technique as to when I used a completely random method.

For the **no two same answers in a row** method, I gave an answer to every odd numbered problem. I used the total random method for every even numbered problem. Then I made another copy, except I did not let the answer for even numbered problems to be the same as the answer before it, and compared the two.

	Average After 81 Trials	Standard Deviation of Sample (S)	95% Interval (t*=1.99, df=80, Answers correct per SAT)
B,C,D Technique – Completely Random	2.741	6.407	1.324 – 4.158
Alternatively Not Same – Alternatively Random	1.173	4.473	0.1840 – 2.162

Figure 5. After 81 trials, there was enough evidence to believe that the B, C, D method was better than the completely random guessing method, and the alternatively not same method was better than the alternatively random method.

I performed 81 trials using Excel. I put the resulting scores into my calculator, and drew the normal quantile plot for it. Since it looked like it was a straight line, meaning it is normally distributed, and there was no reason to doubt that it was not normally distributed, I used the t-distribution in order to find the 95% range for the distribution. The data shows that the probability that the mean is in between **1.324 – 4.158** is 95%. Since it is above zero, it is likely that the **B, C, D only** method is better than the **complete random** method. I did the same thing for the alternatively not same (i.e. **no two same answers in a row**) method, and found that the middle 95% range for its distribution was **0.1840 – 2.162**. Since the difference is above zero, it is statistically likely that the **no two same answers in a row** method is better than the alternatively random method.

Conclusions

I concluded from my results that it is better to guess one of B, C, or D for a problem you do not know what the answer is to. This means that people can likely improve their scores if they guess B, C, or D for a problem they do not know the answer to. Also, the **no two same answers in a row** is also effective. This indicates that people can likely improve their scores if they guess an answer different from the last question.

However, this does not necessary mean you can just guess any problem and get the question right. First of all, for **the no two same answers in a row method, you need to be sure that one of the answer is actually correct**. Then, if you need to guess on the next question, it might be better to not guess the letter of the answer you wrote on the previous question. Also, in the 81 trials, the percentage of questions that were actually correct out of all the questions was a mere 21.33%. You still need to study and get many of the questions correct to get an impressive score. However, if you studied hard and still need to guess, then these special weapons can possibly give you the edge needed to win.

Literature cited

College Board. The Official SAT Study Guide. Ed. Ruth Mills. New York: College Board, 2007.

Harvard Students. 10 Practice Tests for the New SAT. New York: Spark, 2006.

Kaplan. SAT 2008 Edition Premier Program. Ed. Annette Riffle, et al. New York: Kaplan, 2007.

The College Board. 10 Real SATs. Ed. Cathy Claman. 3rd ed. New York: College Board, 2003.

Yates, Daniel, David Moore, and George McCabe. The Practice of Statistics. 1999. Ed. Patrick Farace and Diane Cimino Maass. N.p.: Michelle Russel Julet, 2000.

Acknowledgments

I thank my dad for being my statistical adviser and a great mentor. I thank my mother for her encouragements during the entire project. I thank my friends and my little brother for giving me ideas about the project. And I thank you all for having read my poster.

For further information

Please contact *sei@masuoka.net*. More information on this and related projects can be obtained at *http://masuoka.net/Sei-MTS*. You can find a PDF-version of the poster on the site.