Simulation Method in Teaching Statistics — T Statistic as Example

Zhizhong Yang *

Department of Mathematics, Qinghai Normal University, Qinghai Xining 810008, P.R. China

Yzz7017@163.com

Keywords: Hypothesis Test; Simulation; Test; Interval Estimation; Statistics

Abstract. This paper takes the single-sample T statistics as an example to describe how to teach hypothesis test and interval estimation problems via simulation method. The simulated empirical sizes, empirical powers, empirical confidence level and average interval length of T statistic for different sample size under the normal distribution and t distribution are reported. These results can explain intuitively why one should assume the data and follow the normal distribution, what are type I and type II error probabilities, and what factors may influence them.

Introduction

For many science and engineering professional undergraduates and professional students who major in mathematics, statistics, finance, and so on, probability theory and mathematical statistics is their basic and compulsory course, and its theoretical method has been widely used in the natural sciences, social sciences, agriculture and other fields. The teaching effect of the statistical part plays an important role in training the students' application ability, and it is the theoretical basis for students to learn more other applied statistic courses and to apply statistical methods to their respective expertise areas. However, most students think this course is full of theory in the learning process, its content is abstract, and the method is jerk. It is difficult for them to bring learning interest, and even makes them feel confused and afraid.

The main reason for this phenomenon is the traditional way of textbooks and out-of-date teaching methods. Currently, most domestic and foreign probability theory and mathematical statistics textbooks put emphasis on theoretical derivation and the step by step method described, and then matched with suitable examples. For example, explanation of hypothesis testing methods in textbooks can be summarized into four basic steps: The first step is to put the null hypothesis and the alternative hypothesis. The second step is to choose the appropriate test statistics, and use the sample observations to calculate the statistic value when the original hypothesis established. The third step is to raise the level of significance, and find out the critical value based on the distribution of the test statistic to determine the critical region and the rejection region. The fourth step is to make a judgment, if the value of the test statistic falls into the rejection region, then rejects the null assumption that the null hypothesis is not true, or else accepts the null hypothesis that the current sampling data can not explain the null hypothesis does not hold. Such textbooks arrangement allows students easily be able to remember the basic steps of hypothesis testing, but the essence of the principle of hypothesis tests can not be well understood, resulting in problems occur when doing a lot of problems and puzzles: How to choose the null hypothesis and the alternative hypothesis? Why choose this statistic? Why assume that the samples taken to obey the specified distribution? If the sample does not obey this distribution, what problems will arise? Why the significance level is the probability of committing a type I error? What is the probability of committing a type II error? Why accept the null hypothesis cannot be simply interpreted as the null hypothesis must be set up and so on.

In the teaching process, teachers are also easy to follow textbooks' written order step by step, resulting in that it is difficult to vividly explain to students these types of problems one by one, but also not be a good idea to train students in statistics.

In order to improve the quality of teaching mathematical statistics courses, many literatures concentrated on how to teach this course [1-5], and proposed a number of ways to improve teaching effectiveness, such as full analogy pedagogy [6], a progressive system of teaching methods [7], experimental teaching [8-9] and so on. However, these studies mainly explore the perspective of how teachers teaching. To improve the quality of teaching, and to foster and train students' statistical thinking, a teacher must pay attention to play the initiative of students, and improve students' interest of study, and solve problems and puzzles independently under the guidance of teachers.

It is conducive to better mobilize the students' interest [10] to increase the intensity of the experimental teaching, and combine the teaching theory and computer simulation experiment teaching on the basis of the theory of teaching. Monte Carlo simulation study of mathematical statistics course is an ideal experimental teaching method [11]. This paper based on the single-sample T statistics as example to describe how to make students understand the principles of hypothesis testing as soon as possible through computer simulation methods, understand why one must assume that the observed sample to be normal distribution, what is the relationship between committing of type I and type II error probabilities and sample size, what contact between interval estimation and hypothesis testing and other issues.

Simulation experiment method

A single-sample T statistic is defined as

$$\mathbf{T} = \frac{\bar{\mathrm{X}} - \mu_0}{S_n^*}\sqrt{n} \text{ ,}$$

Where, $\bar{\mathrm{X}}$ is the sample mean, μ_0 is the theoretical mean, S_n^* is the sample standard deviation, n is the sample size. This statistic obeys t(n-1) distribution when samples follow normal distribution.

In the hypothesis testing problem, the single-sample T test statistic is used for testing whether the mean of data which obeys normal distribution is the same as the theoretical mean. For a given significance level α, check the critical values table of the t-distribution to determine the appropriate threshold $t_{1-\alpha/2}(n-1)$. Researchers say that the mean of the sample distribution is different from the theoretical mean when $|T|>t_{1-\alpha/2}(n-1)$, otherwise researchers accept the null hypothesis, and conclude that the current observed data are not enough to explain the mean of sample distribution is unequal to the theoretical mean. Most textbooks in introducing single-sample T test method will generally provide direct observation of samples, and assume that the data obey the normal distribution, to make students test whether the mean is equal to the experience mean directly. Although teachers in the lecture of the course can tell student that assuming normal distribution because only when observational data is from normal distribution, T statistic obeys t(n-1) distribution, in order to find the exact threshold and the probability of significant level committing type I error. But students can not see intuitively what problems will occur if the data are not obey normal distribution, also prone to understand a test result as probability $1-\alpha$ is correct misconceptions, but do not see where reflect type II error probabilities. To solve these problems, researchers use the following simulation experiment methods:

Step 1*:* Use Matlab functions to generate a set of random samples, the mean and size of the sample are μ_0 and n respectively.

Step 2: Calculate the value of statistic T via the generated random data, and write $k=k+1$ if $|T|>C$. The critical value C given here is the case of two-sided test, in the one-sided test it should be changed accordingly. The initial value of k is 0.

Step 3: Repeat the step1 and 2 M times, and calculate the value of k/M which named as the empirical size is the frequency of type I error.

Researchers suggest the teacher ask students write their own programming when calculating the value of the statistic T, rather than using the existing command function. This is useful to help students remember the structure of test statistic. According to above steps, students can easily understand that the test data following the normal distribution with mean μ_0, and will make type I error if $|T|>C$. Since once repeat is equivalent to doing one time test, student can find that the empirical size and the significant level will be very close when the samples size n and the repeat number M is sufficiently large.

The statistics T in the interval estimation problem is used to estimate the parameter μ_0, and the $1-\alpha$ upper and lower confidence bound for μ_0 is $\bar{X}-t_{1-\alpha/2}S_n^*/\sqrt{n}$ and $\bar{X}+t_{1-\alpha/2}S_n^*/\sqrt{n}$ respectively when the sample is from normal distribution. When studying the interval estimation, most students can calculate the correct confidence interval via sample value, but are easy to mistakenly believe that the possibilities of the true value of parameter in the confidence interval is $1-\alpha$, and can not intuitively understand the true meaning of the confidence interval is: if you do n times repeated sampling and n times parameter interval estimation, then among them about $(1-\alpha)n$ intervals contain the true value of parameter, and αn intervals do not contain the true value of parameter. Simulation method makes it easy for students to understand the true meaning of interval estimation. The simulation process can be constructed as the following steps:

Step 1: Same as the above Step 1. .

Step 2: Calculate $C_1=\bar{X}-t_{1-\alpha/2}S_n^*/\sqrt{n}$ and $C_2=\bar{X}+t_{1-\alpha/2}S_n^*/\sqrt{n}$, and write $k=k+1$ if $C_1\le\mu_0\le C_2$. The initial value of k is 0.

Step 3: Repeat the step1 and 2 M times, and calculate the value of k/M named as the empirical confidence level, which describes the proportion of those intervals containing the true value of parameter.

Through this simulation, students can intuitively understand the meaning of the confidence level, and clarify the relationship between significant level in the hypothesis testing and confidence level of interval estimation by comparing the value of relationship between empirical confidence level and a given confidence level. Students also can analyze the relationship between the interval length and sample size by calculating the average value of C_2-C_1.

To enable students to understand the impact of sample size n and the distribution that samples obeying for test and estimate results, in the first step students can separately change the values of n and generate random data via other distribution and re-do simulation to observe the changes of the empirical size and empirical confidence level.

Simulation

This section conducts the simulation by generating N(1,1) and t-distribution with degree of freedom (DOF) 2,5,20 data. Researchers fix the significant level $\alpha=0.05$, and the repeat number M = 100000. The sample size was set to be n=10, 50 and 100 respectively.

Table 1 reports the empirical sizes. From Table 1 researchers can see that the empirical sizes near to the test level (i.e. level of significance) under the normal data. However, for the t-distribution data different DOF gives different empirical size, and small degree of freedom will lead to serious size distortion. This indicates that the T test will make sever type I error if the test data does not follow normal distribution but follow the t distribution. By observing and analyzing these results, students can easily understand the meaning of the significant level and understand the reasons about why researchers must require the test data following normal distribution. This will help students explore the characteristics of t distribution under different DOF. Furthermore, researchers should letting students clear that the probability of committing type I error is not as small as possible, but should be closer to the significant level.

Table 1 Empirical sizes

distribution	n=10	n=50	n=100
N(1,1)	0.0494	0.0503	0.0501
t(2)	0.0343	0.0397	0.0378
t(5)	0.0464	0.0476	0.0466
t(20)	0.0485	0.0478	0.0510

To enable students to understand what is the probability of committing type II error in the hypothesis testing, and the relationship between the probability type II error and the sample size. In the aforementioned first step researchers generate random data with mean $\mu_1 \neq \mu_0 = 1$, and test the null hypothesis of $\mu_0 = 1$. Since the null hypothesis is false, if the ratio (called empirical power) closer to 1 indicating that this test method is better, i.e, the probability of committing Type II error 1-k/M is smaller. Table 2 shows the simulated results when μ_1 equal to 1.5 and 2. From this table researchers can see that the larger sample size gives higher empirical power, i.e. the lower probability of committing Type II error, and the test data is closer to a normal distribution and the higher the test power. This also explains why the test data need to satisfy the requirements of normality from the other side. Comparing the results of for μ_1=1.5 and μ_1=2 researchers can conclude that the larger the difference of true mean of the test data and the mean of null hypothesis, the higher the empirical power, the lower the probability of committing a Type II error. By observing and analyzing these results, students can be very intuitively understanding of what is the probability of committing a Type II error, and some of the factors that affect the size.

Table 2 Empirical Powers

Distribution	μ_1=1.5			μ_1=2		
	n=10	n=50	n=100	n=10	n=50	n=100
N(1,1)	0.2894	0.9287	0.9984	0.8039	1	1
t(2)	0.1423	0.3934	0.5547	0.4112	0.8138	0.9177
t(5)	0.2392	0.7719	0.9641	0.6417	0.9976	0.9999
t(20)	0.2776	0.9063	0.9964	0.7637	1	1

Table 3 shows the results of empirical confidence level and the average interval length. As can be seen, when the samples follow a normal distribution, the empirical confidence level is closest to the confidence level, and the length of the interval is shortest. When samples from t distribution with DOF increase, the empirical confidence level is closer to the confidence level. This indicates that when the samples from a t distribution with large DOF, T statistic can still be used to estimate the population mean. Sample the size has little influence to the empirical confidence level, but has large influence to the interval length what is the larger the sample size, the shorter the interval length. These simulation results will not only allow students to understand what is the confidence level, but also to grasp the content of the impact on the range of the estimated sample size and other traditional textbooks that do not discuss, and help students explore the learning ability and capacity through these simulations. It should be noted that when the sample size is 100, the empirical confidence level is far from the true confidence size because of using the critical value of normal distribution to approximate the critical value of t distribution.

Table 3 Empirical Confidence Level and Average Interval Length

Distribution	Empirical Confidence Level			Average Interval Length		
	n=10	n=50	n=100	n=10	n=50	n=100
N(1,1)	0.9495	0.9502	0.9454	1.3913	0.5669	0.3909
t(2)	0.9664	0.9603	0.9566	3.1221	1.4884	1.0831
t(5)	0.9555	0.9526	0.9472	1.7467	0.7251	0.5019
t(20)	0.9517	0.9511	0.9481	1.4605	0.5965	0.4121

In addition, one can also change the size of the variance of normal distribution to observe the variation of empirical size and empirical power. This will be helpful to deepen students' understanding of the principles of hypothesis testing. By simulating, the one-sided hypothesis testing problems, it can help students understand how to choose the alternative hypothesis and the null hypothesis. These simulation results are no longer listed here.

Summary

In this paper, based on the single-sample T statistics, researchers have described how to teach the topic of hypothesis testing and interval estimation in mathematic statistics through computer simulation. From simulation, a student can easily conclude the change rule of empirical size, empirical power, empirical confidence level and the average length of interval estimation under the normal distribution and t distribution with the change of DOF and the sample size. These change rules can explain intuitively about why the test data are normal distribution, and reveal the relationships among type I error, type II error, sample size and other factors. In this paper, simulation just conducted based on single-sample T statistic. A teacher can use a similar method to simulate other hypothesis testing and interval estimation problem in teaching, and further study the performance of our proposed simulation based teaching method when teach point estimates, variance analysis and other related topics in textbook.

Acknowledgements

This work was supported by the Education Foundation of Qinghai Normal University (2013).

References

[1] L. Yin, Discussion of Teaching Methods in Mathematical Statistics, Mathematics teaching research, 32 (2013) 65-67.

[2] H. Wang, S. Li, Discussion on Statistical Hypothesis Testing, Advanced Mathematics Research, 15 (2012) 50-52.

[3] X. Ma, M. Li, Teaching Reform of Probability Theory and Mathematical Statistics, Statistics and Decision, 13 (2011) 2-3.

[4] X. Tan, D, Xu, Experience in Probability and Statistics Teaching Method, Advanced Mathematics Research, 14 (2011) 97-98.

[5] X. Shen, J. Zhou, Innovative Mechanisms to Probability Theory and Mathematical Statistics Curriculum Reform, Advanced Mathematics Research, 14(2011) 114-115.

[6] H. Peng, Y. Linag, The Whole Analogy of Probability Theory and Mathematical Statistics Courses Teaching and Research, Journal of Mathematics Education, 1 (2012) 95-97.

[7] X. Cheng, X. Wnag, Progressive System of Teaching Methods to Explore – to Probability Theory and Mathematical Statistics Teaching Case , University Education, 17 (2013) 95-97.

[8] Y. Xu, D. Xie, S. Qu, et.at. al., The Classroom Teaching of Using EXCEL Reform Probability Theory and Mathematical Statistics, University Mathematics, 4 (2013) 4-8.

[9] X, Sha, J. Xin, Teaching and Research of Probability Theory and Mathematical Statistics, University Mathematics, 4 (2013) 9-12.

[10] C. Peng, Explore Mathematics Majors Experimental Curriculum System Reform, China University Teaching, 9 (2013) 79-82.

[11] J. Zhang, Y. Ju, Hypothesis Testing the Efficacy of a Monte CarloSimulation, Statistics and Decision, 4 (2012) 83-84.