Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/38/20240554

The core idea of hypothesis testing and its application in examples

Xinyan Xu

School of Mathematics and Statistics, Shanghai Lixin University of Accounting and Finance, Shanghai, 201209, China

211490222@stu.lixin.edu.cn

Abstract. With the development of social science and technology, probability and mathematical statistics have been widely used in many fields. It also includes significant knowledge, such as parameter estimation and hypothesis testing. Hypothesis testing plays an important role in plenty of fields. This article introduces the basic theory of hypothesis testing, such as null hypothesis and alternate hypothesis. Apply test statistics and the level of significance α to select the rejection rejoin. There are also several advantages to using p-value in hypothesis testing. In head up display (HUD) display accuracy test, the accuracy test decision process is a crucial process. Applying hypothesis testing to the accuracy test decision process greatly improves HUD testing solutions. Meanwhile, the risks of both producers and users are also taken into consideration. In the field of automotive engineering, the detection of Max-torque and Max-power also requires hypothesis testing to ensure whether the productions are qualified. Thus, hypothesis testing improves the way of products are tested in the industry.

Keywords: Hypothesis testing, Null hypothesis, Alternate hypothesis, P-value.

1. Introduction

Probability and mathematical statistics have always been very basic and important subject in statistics. Probability embodies the ingenuity of mathematical thinking and the innovation of conclusions. It is also necessary for scholars to master the basic theory of probability. Apply probability to mathematical statistics flexibly and it can make many problems very simple. Mathematical statistics includes parameter estimation and hypothesis testing. Fisher constructed a special method in the women's tea tasting experiment, which called hypothesis testing.

Hypothesis testing uses the idea of establishing a null hypothesis to test events with small probability. Then select the rejection region by using test statistics and the level of significance α. But sometimes p-value makes more contribution to hypothesis testing. It can help scholars select the rejection region with the level of significance α directly. In practice, people can calculate p-value by using statistical software, which makes calculating p-value much easier.

The rest of this paper is organized as follows. Section 2 introduces the core idea of hypothesis testing. Section 3 further illustrates the core idea of hypothesis testing by two examples of applications in production tests. Improve the original production tests can decrease the risk of the manufacturer and users. In display engineering, head up display (HUD) is a display that can show the external environment. The accuracy test of HUD is necessary and significant in the production tests. However, many existing

DOI: 10.54254/2753-8818/38/20240554

HUD accuracy detection methods have some shortcomings. Even the test results will contain substandard products. Applying hypothesis testing to the process of HUD detection can make up for the loopholes in the current process of accuracy tests. In automotive engineering, the maximum torque and maximum power affect vehicle performance. It is necessary to have a complete detection process for the vehicle performance. Since the sample size of torques is sometimes small, choosing the test statistics becomes important. Understanding the application of hypothesis testing in many problems can help people comprehend the idea of hypothesis testing.

2. The concept and basic steps of hypothesis test

2.1. The fundamental concept of hypothesis test

Firstly, the logical reasoning method of hypothesis test is reductio and absurdum. To test whether a hypothesis is true, it is generally assumed that it is correct initially. Then based on the sample information, one can observe whether the result from this assumption is reasonable. This determines whether to accept the null hypothesis. Secondly, based on the principle that the small probability events are not easy to occur, it can be used to judge whether the hypothesis is reasonable. In a single trial, the small probability event is almost impossible to occur. But if a small probability event occurs in the null hypothesis, and the null hypothesis is deemed unreasonable. Contrarily, if the small probability event does not occur, and the null hypothesis is reasonable [1].

2.2. The steps of hypothesis testing.

The First step is to establish the *null hypothesis*, which is denoted by the symbol H_0. It is typically to gather evidence against null hypothesis in hypothesis testing. On the contrary, any hypothesis that differs from the null hypothesis is called an *alternate hypothesis*, which is denoted by the symbol H_1. And in hypothesis testing, it is usually to collect evidence to support H_1. In hypothesis testing, the null hypothesis is rejected only after the occurrence of the small probability event. This is preserving the null hypothesis. And H_0 always equals a certain value. There are three types of statistical tests, which are right-tailed, left-tailed and two-tailed value.

Second, select the test statistics and specify the form of the rejection region. When there is a specific sample, the rules of hypothesis testing can determine whether to accept H_0 or reject H_0. The test is equivalent to dividing the sample space into two mutually exclusive parts, W and $\overline{W}$. When the sample belongs to W, H_0 is rejected. Contrarily, H_0 is accepted. Thus, W is called the rejection region of the test, and $\overline{W}$ is called the acceptance region [2]. Table 1 indicates H_0 and H_1 for the tests of the mean.

Table 1. H_0 and H_1 for the tests of the mean

Null Hypothesis	Alternate Hypothesis and Type of Test		
H_0: $\mu = k$	H_1: $\mu < k$ Left-tailed test	H_1: $\mu > k$ Right-tailed test	H_1: $\mu \neq k$ Two-tailed test

Table 2. Probabilities associated with a statistical test.

Truth of H_0	Decision	
	Accept H_0 as true	Reject H_0 as false
If H_0if true	Correct decision with corresponding probability $1-\alpha$	Type Ⅰ error, with corresponding probability α
If H_0 is false	Type Ⅱ error with corresponding probability β	Correct decision; no error, with corresponding probability$1-\beta$

Third, since accepting or rejecting the null hypothesis is based on the sample information, it is possible that errors in judgment may occur. And there are two types of errors in hypothesis testing. If the H_0 is true, but reject H_0 in this situation, this will make an error. And this type of error called I error. On the contrary, accept H_0 when it is false, this will also lead an error. This called type II error.

Forth, when H_0 is true and the probability of rejecting it is the level of significance α. It is also the probability of type Ⅰ errors. On the contrary, β is the probability of making type Ⅱ errors [3]. Table 2 indicates how these errors occur and probabilities associated with a statistical test.

Fifth, one should calculate the value of the statistics in the hypothesis, and observe whether it falls within the rejection region. Then people should make a judgment, which accept or reject H_0. Finally, according to the judgment, the researchers should make a statistical decision [4]. For the hypothesis tests of μ, given x is normal and known standard deviation σ, the standardized test statistic is

$$u = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}, \tag{1}$$

where n is the sample size, μ is the value stated in H_0, and $\bar{x}$ is the mean of a simple random sample. When the sample size is small, and standard deviation σ is unknown. It is necessary to choose the test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}. \tag{2}$$

As a summary, Table 3 indicates tests of hypotheses about one mean.

Table 3. Tests of hypotheses about one mean.

Method	Condition	H_0	H_1	Test statistic	Rejection region
u test	σ is known	$\mu = \mu_0$ $\mu \leq \mu_0$ $\mu \geq \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$u = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	$\lvert u \rvert \geq u_{1-\alpha/2}$ $u \geq u_{1-\alpha}$ $u \leq -u_{1-\alpha}$
t test	σ is unknown	$\mu = \mu_0$ $\mu \leq \mu_0$ $\mu \geq \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$\lvert t \rvert \geq t_{1-\alpha/2}(n-1)$ $t \geq t_{1-\alpha}(n-1)$ $t \leq -t_{1-\alpha}(n-1)$

2.3. *The P-value of a statistical test*

The minimum significance level at which it can reject the null hypothesis based on the observed sample values is P-value. When conducting hypothesis tests by using statistical software, there will occur P-values. P-values serve as another basis for making decisions in hypothesis testing and it sometimes called the probability of chance.

Different P-values usually depend on H_1 and the type of test. When a test has a mean and it is a standard normal distribution, and $H_0: \mu = k$ and $H_1: \mu < k$. It is a left-tailed test, when $p \leq \alpha$, then reject H_0 at the level of significance α [5]. On the other hand, if $H_0: \mu = k$ and $H_1: \mu > k$, this is right-tailed test. When $p \geq \alpha$, then reject H_0. Figure 1(a) indicates right-tailed test of p-value. When $H_0: \mu = k, H_1: \mu \neq k,$ it is two-tailed test. When $P/2 \geq \alpha/2$ or $-p/2 \leq -\alpha/2$, then reject H_0. Figure 1(b) indicates two-tailed test of p-value.

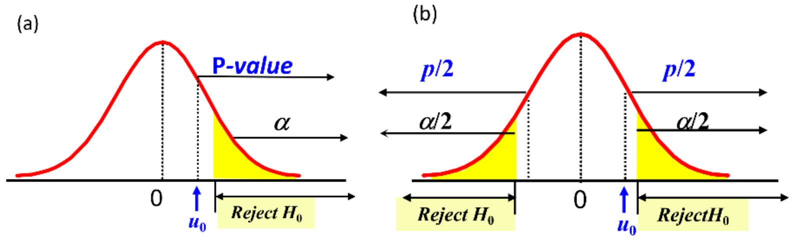

Figure 1. (a) Right-tailed (b)Test Two-tailed Test.

DOI: 10.54254/2753-8818/38/20240554

There are several advantages of using P-value. First, it is not necessary to consult tables to find critical values. It is convenient to compare P-value with α directly. Secondly, since P-value represents the actual probability of type I error, it accurately indicates the level of the significance of the test [6].

3. Examples of hypothesis testing theory plays in practical applications

3.1. Hypothesis test in HUD display accuracy test

HUD is a display that can show the external environment, and drivers can see the display at the same time. HUD display accuracy is the angular difference between the real-world position on the monitor and the HUD display symbol. HUD display accuracy test is important in the practical applications. However, there are still several problems in HUD display accuracy test [7].

There are generally nine points in the HUD display accuracy test, and if all nine points pass, the product is qualified. However, this method does not meet the requirement of 2σ (95.44%). Although there is currently a lot of research on HUD display accuracy test, the methods for evaluating HUD display accuracy are still relatively limited. For example, the incremental upper bounds for producers and users' risks have not yet been determined. Therefore, it is necessary to use hypothesis testing to improve this testing plan. The following is a comprehensive process for HUD display accuracy test. Design a simple random sample of HUD display accuracy testing

To begin with, assuming that the HUD field of view is a rectangular area. There are three regions in this field. The Figure 2 indicates the three areas of the field, and they design simple random samples for each region. The coordinates of each selected point in the field of view must satisfy a two-dimensional uniform distribution. That means the probability of every pixel can be selected is equal.

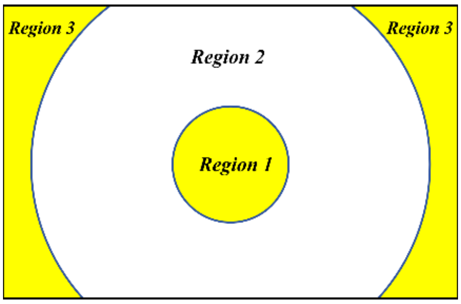

Figure 2. HUD field of view

These steps use simple random sampling theory to design a random sample model that satisfies two-dimensional uniform distribution. Among these three regions, the closer to the center of the field of view, the higher the accuracy required. On the contrary, the lower it is. The proportion of test points to the total test points in the region where they are located should exceed $2\sigma = 95.44\%$. Use hypothesis testing to obtain the process of display precision, and propose a hypothesis at beginning. The null and alternate hypothesis are expressed as H_0: Overall display precision out of tolerance rate $P = 4.56\%$, and H_1: Overall display precision out of tolerance rate $P > 4.56\%$. Then, the author chooses error probabilities α and β are 5% and 10%, respectively, and sets display precision out of tolerance as test statistics. In general testing, the number of outliers obeys hypergeometric distribution. When the sample size is less than or equal to 0.05 times the total sample size, the binomial distribution can be used instead of hypergeometric distribution. And as the number of samples increases, the error probability of β will decrease.

Size from a simple random sample to conduct the experiment, and record the display precision out of tolerance. When the sample size is less than or equal to 0.05 times the total sample size, testing precision out of tolerance is 0, and the accuracy of the display is qualified, it is easy to know the minimum initial sample size is 50. By making decisions and choosing 50 points to test, if there is no display precision out of tolerance, and the accuracy of display is qualified. However, if the display precision points appeared, it is necessary to calculate P-values. When P-value is less than 5%, and the display precision is unqualified. And there is another way to solve this problem, in which P-value is less than 5%. Increase the sample size based on the number of display precision points and the value of β. Then test these samples again. If the display precision is not going to increase, and the accuracy of display is qualified, or it is necessary to increase the sample size again until the accuracy of display is qualified.

In other words, after calculating the P-value, when p-value is less than α, then reject null hypothesis H_0. When p-value is large than α, then accept H_0. And it is necessary to calculate the value of β and increase the sample size until β is less than 10%. After obtaining process of display precision, and list the relationship between the number of out of tolerance, the minimum sample size, and the maximum allowable number out of tolerance when that sample size is less than 388. The results of experiments. Use accuracy test decision process and obtain the relationship between the number of out of tolerance, the minimum sample size, and the maximum allowable number out of tolerance. When the sample size is 50, and the number of out of tolerance is qualified. In the HUD accuracy test record, when the sample size is 84, and the result is qualified. These test results all show that that process of display precision is correct [8]. To conclude, hypothesis testing plays an important role in the HUD accuracy test decision process. And using hypothesis testing can improve the original HUD accuracy test and decrease the risk of the manufacturer and users.

3.2. Application of hypothesis testing in automotive engineering

Hypothesis testing is used in a wide range of fields. Automotive engineering is a good example of hypothesis testing being used. When companies develop vehicle performance, that most important thing is testing the quality of the engine, such as the maximum torque and maximum power. The performance of a car is greatly affected by these factors. Thus, it is necessary to use hypothesis testing to verify whether the torque and power is qualified or unqualified [9].

To start with, due to the limited time and cost, the sample size of cars sometimes is small. And in this condition, it is necessary to choose t test statistic as that of shown in Eq. (2). The value of T distribution will change with the sample size. And the probability of alternate hypothesis can be obtained when the null hypothesis is acceptable. This probability can be used as risk measure.

To proceed further, the author presents the information and calculation process. Let Max-torque be 202 N.M and Max-power be 114 KW as the parameters. These data are not true but it does not affect the results of the hypothesis testing. Then, propose a hypothesis, which are H_0: population mean of max-torque < 202 N.M, and H_1: population mean of max-torque ≥ 202 N.M. Calculate test statistic:

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1), \quad (3)$$

where $\bar{X} = 195N.m,\ u_0 = 202N.m, S = 1.93, n = 22$. Thus, $t = \frac{195-202}{1.93/\sqrt{22}} = -0.82$ [10]. Therefore, the probability $P(t > -0.82) = 1 - P(t < 0.82) = 21\%$. It can be known that the probability that the maximum torque of the engine assembly is greater than or equal to 202N.M. In other words, 100 engines are produced, of which 21 will have a maximum torque value greater than equal to 202N.m.

4. Conclusion

This article introduces the theory of hypothesis testing. The first step is determining what is null hypothesis and alternate hypothesis. Then, select the appropriate test statistic. When the sample size is small, and standard deviation σ is unknown. It is necessary to choose the t test statistic. Apply the level of significance α to determine the rejection region. P-value plays an important role in hypothesis testing. It is direct that compare the P-value with the level of significance α and it is convenient to know the rejection region. In the HUD accuracy testing, hypothesis testing improves the process of HUD accuracy testing and decrease the risk of the manufacturer and users. And in automotive engineering, the sample size of cars sometimes is small, because of the limited time and costs. In this case, it is necessary to use t test statistic, but not u test statistic. Hypothesis testing plays a very important role in many fields. In

DOI: 10.54254/2753-8818/38/20240554

the future, more scholars will learn the idea of hypothesis testing and hypothesis will be applied in more fields.

References

[1] Kukush A, Mishura Y, Ralchenko K. (2017). Hypothesis testing of the drift parameter sign for fractional Ornstein–Uhlenbeck process. Electronic Journal of Statistics, 11(1), 385-400

[2] Lou D H. (2005). Statistical thought of rank sum test. Chinese Journal of Health Statistics, 22(4), 264-265+267.

[3] Veazie, P. J. (2015). Understanding Statistical Testing. Sage Open, 5(1), 1-10.

[4] Robert V. Hogg, Elliot A. Tanis, Dale L. Zimmerman. (2021). Probability and statistical inference, 10th Edition. Pearson.

[5] Hamasaki T, Bretz F, LaVange L M, et al. (2021). roles of hypothesis testing, p-values and decision making in biopharmaceutical research. Statistics in Biopharmaceutical Research, 13(1), 1-5.

[6] Dongmei Fan. (2013). P-value in hypothesis testing. Zhengzhou: Journal of Zhengzhou economic management institute, 23(04), 70-71.

[7] Shouyuan Ynag, Hao Ding, Guijie Yu. (2023). Application of Hypothesis Test in HUD Display Accuracy Test. Luoyang: Electronics Optics& Control, 30(06), 82-88.

[8] Junhu Wang. (2021). Three Decision Criteria in Statistical Hypothesis Test, Hubei: Statistics& Decision, 37(06), 56-60.

[9] Wang J, Wu X. (2008). The results of variance analysis and rank sum test were compared in practical application. Chinese Journal of Health Statistics, 25(1): 55-58.

[10] Junping Jia, Xiaoqun He, Yongjin Jin. (2021). Statistics. Eighth Edition. Beijing: Renmin University Press.