

UNDERSTANDING HYPOTHESIS AND STATISTICAL TESTING IN RESEARCH FORMULATION

Dr. Ritu Sharma*

ABSTRACT

Under inferential statistics two topics are discussed viz. estimation of population values and testing the statistical hypotheses. Although there are two approaches to hypothesis testing, the more established is the classical or sampling theory approach. Classical statistics are found in all of the major statistics books and are widely used in research application. Hypothesis is tested on the basis of sample data collected. Researcher is entirely dependent on analysis of sample data. Second approach is known as Bayesian statistics, which are extension of the classical approach. They also use sampling data, but they go beyond to consider all other available information. This additional information consists of subjective probability estimates stated in terms of degree of belief. These subjective estimates are based on general experience rather than on specific collected data.

Keywords: *Inferential Statistics, Statistical Hypotheses, Hypothesis Testing, Sampling Theory.*

Introduction Statistical Testing

Following the sampling theory, we accept or reject a hypothesis on the basis of sampling information alone. Since any sample will almost surely vary some-what from its population, we must judge whether these differences are statistically significant or insignificant. A difference has statistical significance if there is good reason to believe the difference does not represent random sampling fluctuations only. For example, the controller of a large retail chain may be concerned about a possible slowdown in payment by the company's customers. He measures the rate of payment in terms of the average age of receivables outstanding. Generally, the company has maintained an average of about 50 days with standard deviation of 10 days. Suppose the controller has all of the customer account analyze and finds the average is now 51 days. Is this difference statistically significant from 50? Of course it is because the difference is based on a census of the account and there is no sampling involved. It is a fact that the population average has moved from 50 to 51 days.

Since it would be too expensive to analyze all of a company's receivables frequently, we normally resort to sampling. Assume a sample of 25 is randomly selected and the average number of days outstanding is calculated to be 54. Is this statistically significant? The answer is not obvious. It is significant if there is good reason to believe the average age of the total group of receivables has moved up from 50. Since the evidence consists of only a sample, consider the second possibility, that this is only a random sampling error and thus not significant. The task is to judge, whether such a result from this sample is or is not statistically significant. To answer this question, one needs to consider the logic of hypothesis testing. To describe the process of hypothesis testing the following steps may be used:

Step 1 Formulate the Practical Problem in Terms of Hypotheses

In classical tests of significance, two kinds of hypotheses viz. null hypothesis and alternate hypothesis are used. Null hypothesis is a statement that no difference exists between the parameter (a measure taken by a census of the population or a prior measurement of a sample of the population) and the statistic being compared to it(a measure from a recently drawn sample of population). Analysts

* Assistant Professor, Department of Sociology, Kamla Nehru College, University of Delhi, Delhi, India.

usually test to determine whether there has been no change in the population of interest or whether a real difference exists. The null hypothesis needs to be very simple and represents the status quo, i.e. there is no difference between the processes being tested. It is basically a standard or control with which the evidence pointing to the alternative can be compared.

This can be difficult in some cases. We should first concentrate on what is called the alternative hypothesis, H_A . This should express the range of situations that we wish the test to be able to diagnose. The alternative hypothesis may take several forms, depending on the objective of the researchers. The alternative hypothesis may be of the “not the same” or the “greater than” or “less than” form. In this sense, a positive test can indicate that we should take action of some kind. In fact, a better name for the alternative hypothesis would be the action hypothesis. These types of alternative hypothesis correspond with two-tailed test and one tailed tests. A two tailed test or non-directional test, considers two possibilities greater than or less than. A one tailed test, or directional test, place entire probability of unlikely outcome into the tail specified by the alternative hypothesis.

Step 2 Set the Level of Significance

The level of significance generally denoted by α is the probability, which is attached to a null hypothesis, which may be rejected even when it is true. The possibility of committing type 1 error is called α or level of significance. The level of significance is also known as the size of the rejection region or the size of the critical region. It is very important to note that the level of significance must be determined before drawing the samples. So that the obtained results are free from choice bias of a decision maker. The level of significance which is generally applied by researchers is 1% or 5%, depending on the severity of the consequences of making such an error. This involves specifying how great a risk we are prepared to run of coming to an incorrect conclusion. We also have to contend with the possibility of not rejecting H_0 when it is in fact false and H_1 is true. This is an error of the second type or Type II error, and the probability of this occurring is denoted by β . Thus in testing any statistical hypothesis, there are four possible situations which determine whether our decision is correct or in error. These situations are illustrated as follows:

Hypothesis Testing

		Null hypothesis	
		Accept	Reject
Null Hypothesis	True	Correct Decision	Type I error (α) = P(rejecting a good lot) Producer's risk
	False	Type II error (β) = (Accepting a bad lot) Consumer's risk	Correct Decision

Step 3 Determination of Suitable Significant Test- Identify Sample Statistic (t)

Test statistic is to be determined by applying appropriate test. Under this step researcher has to select an appropriate measure for the determination of ratio of significance under the null hypothesis and alternative hypothesis. There are two general classes of significant tests: parametric and nonparametric. Parametric tests are more powerful because their data are derived from interval and ratio measurement. Nonparametric test are used to test hypotheses with nominal and ordinal data. Parametric Tests are statistical techniques to test hypothesis based on some restrictive assumptions about the population.

- The observations should be drawn from normally distributed population
- Random selection of sample
- The measurement scale should be at least Interval or Ratio scales data

Recommended statistical techniques by measurement level and testing situation

	One Sample Case	Two Sample Case		k- sample Case	
Measurement level		Related samples	Independent samples	Related samples	Independent samples
Nominal	Binomial Chi square	McNemar	Chi square	Cochran Q	Chi square for k samples
Ordinal	Kolmogorov-smirnov Run test	Sign test Wilcoxon matched pairs	Median test Mann-Whitney U test Kolmogorov-smirnov	Friedman two-way ANOVA	Median extension Kruskal-wallis one way ANOVA
Interval and Ratio	t test Z test	t test for paired samples	t test Z test	ANOVA	One way & N-way ANOVA

Step 4: Determine the Critical Value

A critical value is a value in normal distribution that indicates the number of times of the standard error in normal distribution expected to cover certain area of the universe. The relationship of critical value and the area expected to be covered in a normal distribution is shown in the following figure:

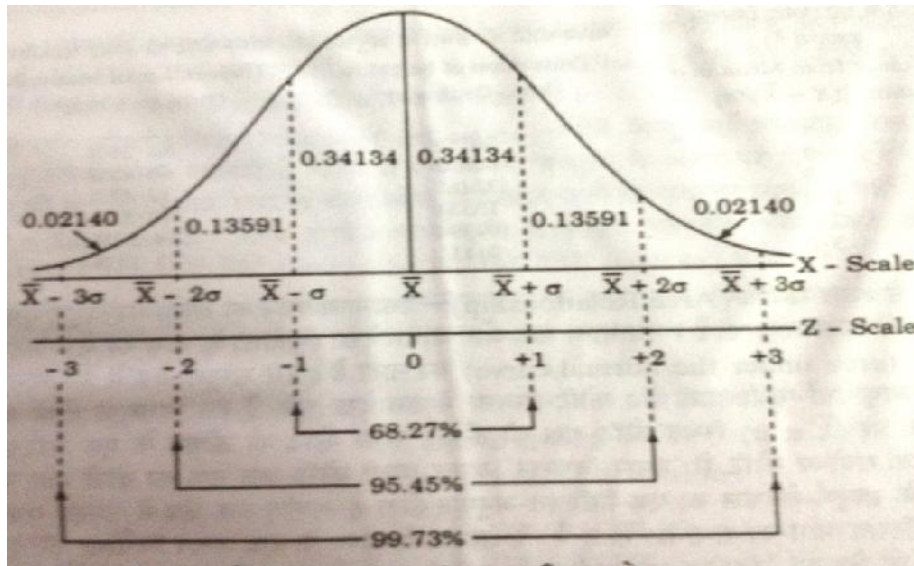


Figure: Area under normal distribution

It may be seen in the above figure that the area covered in normal distribution can be specified in terms of mean plus and minus certain time of standard deviation. This area covered and the number of times of standard deviation indicates the level of probability or level of significance and critical value respectively. This relationship is given below table

Area Covered	Critical Value	Level of Significance	Level of Significance
$\bar{X} \pm S.D.$	1	68.27%	31.27%
$\bar{X} \pm 1.96S.D.$	1.96	95%	5%
$\bar{X} \pm 2 S.D$	2	95.45%	4.55%
$\bar{X} \pm 2.58 S.D.$	2.58	99%	1%
$\bar{X} \pm 3S.D.$	3	99.73%	0.27%

Probability Values (p Values)

Hypothesis can be accepted or rejected on the basis of p values. p value method of presenting the result of a statistical test report the extent to which the test statistic disagrees with the null hypothesis. This method has become popular because analysts want to know what percentage of the sampling distribution lies beyond the sample statistic on the curve. Most of the statistical computer programs report the results of statistical tests as probability values (p values). the p value is the probability of observing a sample value as extreme as, or more extreme than, the value actually observed, given that null hypothesis is true. This area represents the probability of a type I error that must be assumed if the null hypothesis is rejected.

Determination of p Value

The p value is determined using the standard normal table. The area between the mean and a Z value is found, then such area is subtracted from level of significance say 5%, the resultant figure is p value. For example value of Z comes 2.5. The area between mean and Z value i.e. 2.5 is 0.4983 (derived from standard normal table). So p value for one tailed test would be $0.05 - 0.4983 = 0.0062$ and for two tailed test it would be $0.0062 * 2 = 0.0124$

Step 5: Decision: Interpret the Test

The computed value of test statistic (Z,t,F, etc) is compared with the tabulated value (Z,t, F) at the given level of significance. In case of parametric test if calculated value of statistic is greater than

critical value (at 5% 1.96 and at 1% 2.58) then null hypothesis is rejected otherwise accepted. In case of non parametric test critical value is derived from corresponding table at given degree of freedom and level of significance. If calculated value of statistic is less than derived table value, null hypothesis is accepted otherwise rejected. Simultaneously hypothesis can also be tested on the basis of p values. If the p value is less than the significant level, the null hypothesis is rejected, if p value is greater than or equal to the significance level, the null hypothesis is not rejected. Null hypothesis is rejected means difference between parameter and statistic is significant and there is a good reason to believe the difference does not represent random sampling fluctuations only. Null hypothesis accepted means visible difference is available only due to chance or sampling fluctuations not due to major reasons.

Step 6 Measuring the Power of Test

It is quite important to know how well a hypothesis test is working. The measure of how well the test is working is called the power of the test. Rejection of hypothesis means we run a risk of committing an error of type I but acceptance of null hypothesis means we run some risk of committing an error of type II. The probability of type one error is denoted by α (level of significance) and the probability of type II error is referred as β . Usually significance level of a test is determined in advance and once we decide it, there is nothing else we can do about alpha. Hypothesis test cannot be foolproof: sometimes the test does not reject null hypothesis when it happens to be a false and this way a type II error is made. But we would certainly like the probability of accepting null hypothesis when it is not true i.e. β to be as small as possible. Alternatively, we would like that $1 - \beta$ to be as large as possible. **If $1 - \beta$ is very much nearer to unity (i.e. nearer to 1). We can infer that the test is working quite well, and if $1 - \beta$ is very much nearer to 0 then we infer the test is poorly working.**

If test statistic lies close to the boundary of the critical region we may say that there is some evidence that H_0 should be rejected, whereas if it is at the other end of the region we would conclude there was considerable evidence. In other words, the actual significance level of T can provide useful information beyond the fact that T lies in the critical region. In general, the statistical test provides information from which we can judge the significance of the increase (or decrease) in any result. If our conclusion shows that the increase is not significant then it will be necessary to confirm that the experiment had a fair chance of establishing an increase had there been one present to establish. In order to do this we generally turn to the power function of the test, which is usually computed before the experiment is performed, so that if it is insufficiently powerful then the design can be changed. The power function is the probability of detecting a genuine increase underlying the observed increase in the result, plotted as a function of the genuine increase, and therefore the experimental design must be chosen so that the probability of detecting the increase is high. Also the choice among several possible designs should be made in favour of the experiment with the highest power.

Small Case Study

A marketing research firm conducted a survey 10 years ago and found the average household income of a particular geographic region was Rs. 10,000. Mr. XYZ, who has recently joined the firm as a vice president has expressed doubt about the accuracy of the data. For verifying the data, the firm has decided to take random sample of 200 households that yield a sample mean (for household income) of Rs. 10200. Historical data shows that population standard deviation of the household income is Rs.1200. Verify the Mr. Bhatanagar's doubt at 5% level of significance.

Solution

- Step 1 Set null hypothesis or alternate hypothesis
The null hypothesis is set as no difference or status quo that is, the average of household income has not changed. Or there is no difference between average values of household income ten years ago i.e. Rs.10000 and average value of household income derived i.e. Rs.10200 recently.
- Step 2 Set the level of significance: already given i.e. $\alpha = 5\%$
- Step 3 Determine the appropriate statistical test and apply formula
Sample size is greater than 30 and sample mean is used as statistic, so the Z test can be used for hypothesis testing. i.e. Test of significance of single mean

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

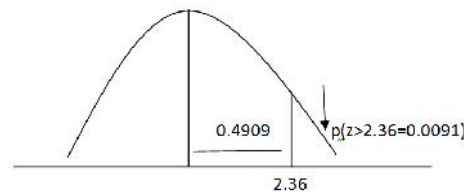
$$Z = 10200 - 10000 / 1200 / 200 = 2.36$$

Notes

- if population standard deviation remains unknown then sample standard deviation can be used.
- The z formula is based on the assumption that the sample is drawn from an infinite population.
- Step 4 Testing the hypothesis
- By using critical value approach

The calculated value of Z is 2.36 which is greater than critical value of 1.96 at 5% level of significance. Hence null hypothesis is rejected. On the basis of hypothesis testing, it can be concluded that the average household income of the particular geographic region is not Rs.10000. Mr. Bhatanagar's doubt about this average household income was right.
- By using p-Value approach

The p-Value approach for hypothesis testing is used for large samples and is sometime referred to as observed level of significance. This approach is very advantageous especially after the introduction of many statistical software programmes. Here the value of p is 0.0182 $((0.05-0.4909)*2)$ i.e. less than 0.05 hence null hypothesis is rejected.



The observed test statistics is computed as 2.36, from the normal table, the corresponding probability area for $z = 2.36$ is 0.4909. so, the probability of obtaining a Z value greater than or equal to 2.36 is $0.5000 - 0.4909 = 0.0091$ for a two tailed test, this value is multiplied by 2. Thus for a two-tailed test, this value is $(0.0091 * 2 = 0.0182)$. it has to be noted that for $\alpha = 0.05$, the null hypothesis is rejected because 0.0182 is less than 0.05.

- By setting limits

Limits = $10000 \pm 1.96 * 1200$ 200
 = $10000 \pm 1.96 * 84.86$
 = Rs. 9833 Rs10167

Sample mean is out of range hence null hypothesis is rejected

References

- ✧ Cooper R Donald R, Schindler Pamela S and Sharma J.K., Business Research Methods, McGraw Hill Education (India) private Limited, New Delhi.
- ✧ Gupta S.P., Statistical Methods, Sultan Chand & Sons, Education Publishers, New Delhi
- ✧ Phanse Sameer, Research Methodology Logic, Methods and Cases, Oxford University Press, New Delhi
- ✧ Bajpai Naval, Business Research Methods, Pearson Education in South Asia, Noida
- ✧ Sharma J.K., Business Statistics, Pearson Education in south Asia, New Delhi
- ✧ Albright Christian S, Winston L. Wayne, Business Analytics: Data Analysis and Decision Making, Cenage Learning India, Delhi

