Test of Hypothesis: Basic Principles*

In any empirical research, especially in social science, the researcher has to face the following main questions: What are the objectives of your study? What hypotheses are you going to test? What methodologies are you using for the purpose? What is your database? What are your main findings? What are your conclusions and policy implications? Hypothesis is the crucial concept in all these questions and so in this type of research. Methodologies at hand are used or new methodologies are devised to test the hypotheses proposed in accordance with the objectives in mind. They are tested with the help of data available or suitably collected to arrive at some conclusions that have policy implications.

A **hypothesis** is simply a proposition that is assumed to be true. It is a presumption. It is generally formulated by logical deductions of existing theories. It may also be the outcome of some empirical generalisations. It is tested by the observations. It is the backbone of scientific researches, especially of empirical social researches. It is assumed to be true because it has logical or general empirical base and because it is either not yet tested to be true or if tested not yet proved to be true. Repeated and varied testing of a hypothesis lead to empirical generalisation and modification of existing theories by inductive logic. This effort to enhance/reduce the boundary of theories or to strengthen/weaken the existing theories is the objective of empirical social research and this establishes the relevance of **hypothesis**.

Empirical exercises to have some empirical results or descriptive statistics or even to make some inferences about the corresponding population have little theoretical relevance unless they are governed by a set of hypotheses or theoretical predictions. A theoretical prediction or a theoretical hypothesis is more important than a statistical hypothesis. Let us consider some simple examples.

Suppose we are interested to examine the condition of Self-Help Groups (SHGs) in the district of Paschim Medinipur. For this purpose, we collect different information on a sizable sample of SHGs formed so far in the district. We may find that a significant portion of population below the poverty line is involved in the formation of SHGs, or we may find that a significant portion of members of SHGs comes from the population below the poverty line. We may also find that

---

* Lecture note prepared by Prof. Debasish Mondal, Department of Economics, Vidyasagar University for second semester students 2020 (Paper Eco 201).

the proportions of SC/ST members or female members in the SHGs are also significantly greater than the corresponding non-SC/ST or male members.

The significance tests of the above proportions are performed by testing some null statistical hypotheses against their respective suitable alternatives. For example, we can come to the conclusion that a significant portion of members of SHGs comes from the population below the poverty line by rejecting the null hypothesis that $P=0$ against the alternative that $P>0$, where P is the population proportion of members below the poverty line. But this does not ensure that the proportion of members below the poverty line is significantly greater than that above the poverty line. We can arrive at this conclusion if we can reject the null hypothesis that $P=\frac{1}{2}$ against the alternative that $P>\frac{1}{2}$. But this does not ensure that the proportion of members above the poverty line is insignificant. To arrive at that conclusion, we have to fail to reject the null hypothesis that $P=1$ against the alternative that $P<1$.

Thus, for a single aspect, we have the possibility of different types of null and corresponding alternative hypotheses and all of them have different theoretical implications. Therefore, which statistical hypothesis we shall formulate and test depends on the theoretical possibility we have in our mind. A theoretical prediction or a theoretical hypothesis is more important than a statistical hypothesis. Statistical hypotheses are formed to test directly or indirectly the theoretical hypotheses formed out of the objectives in the given theoretical environment. For example, we may have the theoretical prediction that the institutional provision for the formation of cooperatives even for day-to-day basic economic activities of subsistence consumption and employment (or SHGs) help the population below the poverty line more than that above the poverty line.

The theoretical hypothesis we are proposing may not fully reflect the objectives we have in mind; the theoretical hypothesis we are proposing may not be totally captured by a suitable statistical hypothesis; and the statistical hypothesis we are forming may not be tested directly. For example, the hypothesis that $P>0$ can not be tested directly whereas the hypothesis that $P=1$ can be tested directly.

A statistical hypothesis is not tested for its acceptance against rejection; but it is tested for its rejection against non-rejection. For example, for testing the directly testable hypothesis that $P=1$ we try to utilise the information at hand (data at hand) to reject the null hypothesis which is assumed to be true under the existing theoretical environment. If we fail to reject it then we find more confidence in saying that our presumption based on theoretical prediction is probably not wrong. Thus, theoretical prediction is very important.

Consider another simple example. The Keynesian theory of income determination suggests that consumption is a stable function of income both across sections and over time. It also suggests that the short run propensity to consume is significantly less than the long run propensity to consume. These logical suggestions of the Keynesian theory can be presented by a set of hypotheses and can be tested by relevant data.

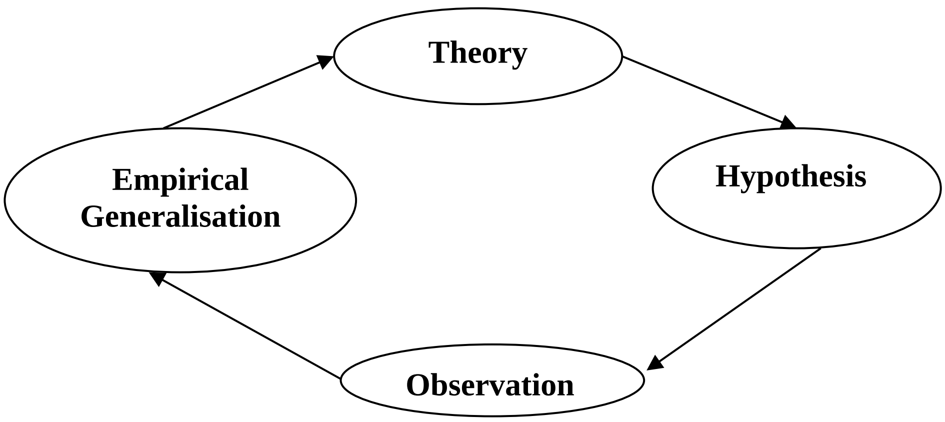
Thirdly, consider the case of variability of productivity of rice across different sub-regions in a broader region (for example across different states in India, or across different districts in West Bengal, or in different blocks in Paschim Medinipur etc.) and that over time in a specific region or in different regions. Theory suggests several factors behind these different types of variability. A hypothesis that a factor that has higher cross-section variability than inter-temporal variability will be more significant in explaining the cross-section variability than in explaining the inter-temporal variability may not be illogical.

Formulation of hypothesis is related to theoretical prediction and testing of hypothesis is meant for empirical prediction. Deductive logic is applied on the existing theories to build up propositions or predictions or hypotheses on the observable variables or attributes. These propositions are assumed to be true to examine their validity. Deductive logic is now applied under the above assumptions on statistical theories to build up deduced propositions on some statistics or measures that can be obtained from the observations. If our assumptions are true then the deduced propositions will not be found to be false and our assumptions are not found to be invalid. Thus, a hypothesis is a logical testable proposition.

Hypothesis is a key concept in scientific researches, especially in empirical social researches. The chain of events in such a research can be presented schematically as shown in figure-1 below. The urge for empirical research starts from the knowledge of existing literature in the field of interest. It consists of the <u>theories</u> developed or developing (or controversies existing) in the field and of <u>empirical generalisation</u> and <u>major empirical findings</u> in the field.

Logical deductions from the literature help us forming some propositions or predictions or <u>hypotheses</u> or presumptions that are assumed to be true for further logical deductions or for policy implications. These hypotheses further help us forming some <u>statistical hypotheses</u> that can be tested with the help of <u>observations</u>. Rejection or non-rejection of statistical hypotheses raises or reduces the confidence in the theoretical hypotheses and this ultimately raises or reduces the confidence in the existing theory or enlarges the boundary of the existing literature. The dashed line in the figure below shows the demarcation between theories and empirics.

Figure-1 Schematic presentation of the relevance of Hypothesis



Statistical hypotheses are of different types. On the one hand, we have parametric hypotheses regarding parameters only of some presumed population distribution and on the other hand, we have non-parametric hypotheses with no such presumption. Where non-parametric hypotheses have wide uses, parametric hypotheses have strong and specific inferential power. Hypotheses may be related to one, two or many univariate populations; they may also be related to bivariate or multivariate population. It may be related to classified data on variables or attributes – classified by values in case of variables and classified by qualitative features in case of attributes. There may be one-way, two-way or multi-way classification. It is not possible to discuss all of them and it is not also needed. We shall discuss mainly the principle of hypothesis testing in some simple situations with the help of some simple examples.

For example, consider the theoretical prediction that a region in India predominant with SC/ST population has mean income significantly less than that for India as a whole. Suppose that the mean income for India as a whole is truly known as μ_0. This theoretical prediction or hypothesis directly involves a simple univariate population on a quantitative variable, viz., the income of a totality (population) of households in the regions of India predominant with SC/ST population.

This theoretical hypothesis also coincides with the statistical hypothesis that $\mu < \mu_0$, where μ is the mean income of the population of the regions of India predominant with SC/ST population.

But the statistical hypothesis that $\mu < \mu_0$ cannot be directly tested. To perform this test, we take the null hypothesis $H_0 : \mu = \mu_0$ and we test it against the alternative hypothesis $H_1 : \mu < \mu_0$. Thus, the null hypothesis which is assumed to be true does not coincide with the theoretical hypothesis. The procedure of testing is about rejection or non-rejection of H_0. The procedure involves rejection of H_0 which is assumed to be true if contradictory (least likely) outcome (contradictory under the assumption) occurs. If H_0 is rejected against H_1 we feel confident in our theoretical prediction; but if H_0 is not rejected against H_1 we lose confidence in our theoretical prediction and begin to suspect it.

To test a simple hypothesis like the above with the help of a sample drawn from the population, we generally make the assumption that the variable (X) in question (in this case, the income of households in the above mentioned regions) is normally distributed with unknown mean (μ) and with known or unknown variance (σ^2). Symbolically, we assume that $X \sim N(\mu, \sigma^2)$. μ and σ^2 are called the parameters of the normal population. This is also a statistical assumption or statistical hypothesis and should be backed by some theoretical hypothesis, and if possible, should be tested before testing the main hypothesis. Secondly, to make the testing procedure simple and easy we draw the sample following simple random sampling with replacement procedure.

Under this statistical environment and under the consideration that we are interested in the theoretical hypothesis about μ we try to observe the consistency of the hypothesis in the value of a statistic in which the value of μ is reflected most. Under the above statistical environment $\bar{x}$, the sample mean, is that statistic. $\bar{x}$ is not only the statistic that reflects the value of μ most, it is also the maximum likelihood estimator of μ. This implies that the likelihood that the value of μ coincides with the value of $\bar{x}$ is greater than the likelihood that the value of μ coincides with any other statistic of the sample. The sampling distribution of $\bar{x}$ is $\bar{x} \sim N(\mu, \sigma^2 / n)$. Under the null hypothesis $H_0 : \mu = \mu_0$ we have $\bar{x} \sim N(\mu_0, \sigma^2 / n)$. Now, given the alternative hypothesis and treating 100α % (generally not more than 5%) extreme (least probable) cases insignificant or taking α as the level of significance, it does not seem inconsistent to observe the value of $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$, also known as τ_0, greater than τ_α. If it is not observed, we feel inconsistency in the null hypothesis and we reject it.

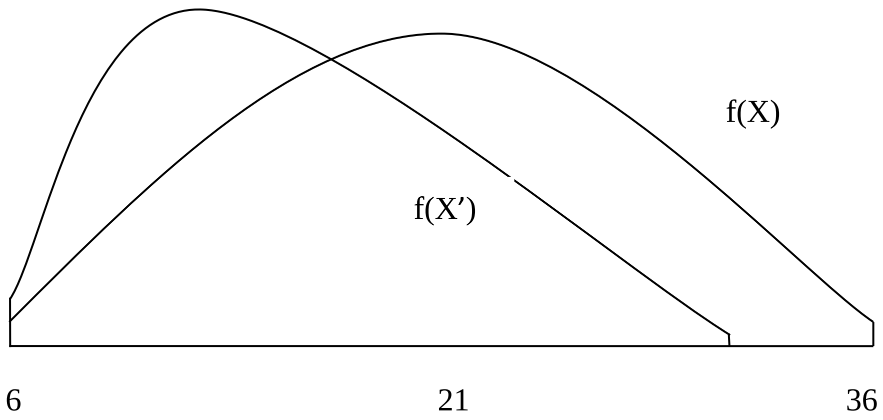
The inconsistency in our observation may also be due to the fact that the population is not normally distributed or the sample is not truly random. Thus to make our testing of the main hypothesis more reliable it is best to test the normality of the population hypothesis first and to ensure that the sample is really random in the true sense of the term.

However, we have an escape. If the population is not normally distributed but has a finite mean (μ) and a finite variance (σ^2) and if the sample is a simple random sample with replacement, and if the sample size is fairly large then also by virtue of the Central Limit Theorem the distribution of the sample mean will be as before. Moreover, if the sample is not perfectly random but fairly random then also the above testing procedure can be applied approximately. However, the testing procedure will lose its exactness property or strength.

The Central Limit Theorem mentioned above has wide applicability. By virtue of this theorem, we can justify theoretically the assumption of a normal population distribution. For this consider a simple example of a lottery play I observed in a fair in my village when I was a college student studying the theories of probability and probability distribution. In a box, there are 36 tokens – 6 with one point, 6 with two points, 6 with three points, 6 with four points, 6 with five points and 6 with six points. One has to draw six tokens at a time from the box at a charge of Rs. 2 per draw. He will obtain total points in between 6 and 36. If he can draw 6 or 36, he will receive a prize worth Rs. 1000. He will receive a prize of Rs. 500 for points 7 or 35, that of Rs. 200 for points 8 or 34, that of Rs. 100 for points 9 or 33, that of Rs. 50 for points 10 or 32, that of Rs. 20 for points 11 or 31, that of Rs. 10 for points 12 or 30, that of Rs. 5 for points 13 or 29, that of Rs. 2 for points 14 or 28 and a consolation prize of 10 paisa for points in between 15 and 27.

Though the probabilities of having a token of point one, or two, or three, or four, or five or six are all equal (1/6 each) the probability distribution for having total points in between 6 to 36 takes a unimodal symmetric shape like that shown in figure-2 by the curve f(X). For large no of values and draws it can be approximated by the normal distribution. If three tokens of six or five points are permanently removed from the box, it is no longer possible to have 36, 35 or 34 points and the probability of having lower points will increase and the probability distribution curve will now be positively skewed and will look like f(X') as shown in figure-2.

Figure 2 The mechanics of population distribution



An almost similar case is observed when we think of total marks obtained by a student in a subject or in an examination. If the probability that the student can answer a question is equal to the probability that he cannot and if it is grossly true for all questions in all subjects, the probability distribution of marks of a subject will be symmetric and unimodal like that shown by f(X) in figure-2 and the distribution will be close to the normal distribution. The probability distribution of marks in the examination as a whole will be closer to the normal distribution. However, if the probability that the student can answer a question is less than the probability that he cannot and if it grossly true for all questions in all subjects, the probability distribution of marks of a subject will be unimodal but positively skewed like that shown by f(X′); but the probability distribution of marks in the examination will be less skewed and will tend to the normal distribution.

Similarly, income of a household is the sum of incomes of different members of it and income of a member is the sum of incomes from different sources. Normally the probability of getting a high income from any source is less than that of getting low income from that source, or the probability of getting a job with low income is higher than that with high income, and the distribution of household income is positively skewed but as it maintains the law of total it will be close to the normal distribution. For income of a region comprising of a large number of households it is closer to the normal distribution. This is the theoretical backup behind the hypothesis that the variable of this sort is normally distributed. However, before testing the parametric hypothesis regarding mean or variance of the population, this primary hypothesis is to be tested with the help of observations, because if this primary hypothesis is rejected the parametric hypothesis loses it significance.

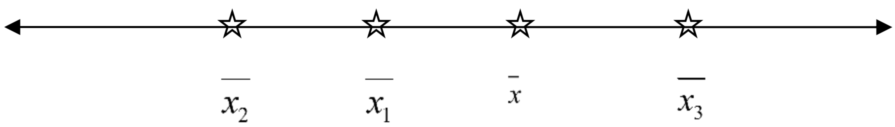
Now let us consider the case of two univariate populations. To make the estimates precise and to make the tests of significance specific we assume that these populations are independently and normally distributed. It is reminded that there should be some theoretical justification behind these assumptions and if needed should be tested first before testing the concerned socio-economic hypotheses of the research. Under the above assumptions we have two populations $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ with four parameters μ_1 comparable with μ_2 and σ_1^2 comparable with σ_2^2. Thus, null hypothesis like $H_0 : \mu_1 = \mu_2$ or that like $H_0 : \sigma_1^2 = \sigma_2^2$ can be tested with high socio-economic significance. Univariate hypotheses about the means and variances of these two populations can also be tested side by side and these tests can also be compared or tested. For example, we can test $H_0 : \mu_1 = 0$ and $H_0 : \mu_2 = 0$ side by side and if both are rejected we can try to compare the power and significance of these two tests. One usual way to do this is to compare their marginal levels of significance. If the maximum level of significance that rejects the null hypothesis about the first population is less than that about the second population, we can say that the null hypothesis about the first population is more significantly rejected than that about the second population. This maximum level of significance is called the probability value (or, P-value in short) and is reported in almost all statistical packages. One theoretical point about these comparisons of means or variances and comparisons of tests of two populations is worth mentioning in this context.

Let us think of a situation where a simple random sample with replacement of size n_1 is drawn from a normally distributed population with unknown mean, μ, and known variance, σ^2. Let us suppose that the sample mean is $\bar{x}$ and this is the minimum variance unbiased estimator of μ with sampling variance = σ^2 / n_1. If n_2 new observations are added with the existing n_1 observations and the sample mean remains unchanged at $\bar{x}$ the best unbiased estimate of μ remains unchanged but at the same time the estimate becomes better than before because the variance of the estimator is now $\sigma^2 / (n_1 + n_2)$ which is sufficiently less than σ^2 / n_1.

Under these circumstances the null hypothesis that $\mu = 0$ will be rejected at a lower level of significance than before. This is due to the fact that the unbiased estimates of the population mean are equal in two situations, but their sampling variances are different. If we want to keep the level of significance unchanged then some deviation in sample mean can be allowed. If the alternative hypothesis is $\mu \neq 0$ a deviation of sample mean on either sides is allowed (that is, the new sample mean is allowed to fall or to increase from the old mean to a fixed extent). If, on the other hand, the alternative hypothesis is $\mu > 0$, a deviation only in the downward direction is allowed.

Under this second alternative, if the sample mean falls to $\bar{x}_1$ (so that the level of significance remains unchanged); the estimate becomes equally good in terms of level of significance. However, at the same time it becomes worse than before because the null hypothesis that 'the population mean has remained unchanged' will never be rejected against the alternative that it has improved and will be rejected for some level of significance against the alternative that it has deteriorated. Thus if we are concerned with both unbiasedness and minimum variance, we have to make a compromise between the two criteria of goodness.

Figure 3 Tests of means compared



In this context, it would be convenient to set another lower limit of the sample mean $(\bar{x}_2)$ below $\bar{x}$ so that the null hypothesis that 'the population mean has remained unchanged' will be rejected at a predetermined level of significance against the alternative that it has deteriorated. Similarly an upper limit of the sample mean $(\bar{x}_3)$ can be set above $\bar{x}$ so that the null hypothesis that 'the population mean has remained unchanged' will be rejected at a predetermined level of significance against the alternative that it has improved. If the predetermined level of significance is very small, $\bar{x}_2$ will be less than $\bar{x}_1$. In this case, if the new sample mean falls below $\bar{x}_2$ the hypothesis that 'the population mean has improved' will be rejected at the predetermined level of significance. If the new sample mean lies above $\bar{x}_2$ but falls below $\bar{x}_1$ the hypothesis that 'the population mean is zero' will be rejected against the alternative that it is greater than zero at a lower level of significance. If the new sample mean lies above $\bar{x}_1$ but falls below $\bar{x}$ the hypothesis that 'the population mean is zero' will be rejected against the alternative that it is greater than zero at a higher level of significance. If the new sample mean lies above $\bar{x}$ but falls below $\bar{x}_3$ the hypothesis that 'the population mean is zero' will be rejected against the alternative that it is less than zero at a higher level of significance. Finally, if the new sample mean lies above $\bar{x}_3$ the hypothesis that 'the population mean has deteriorated' will be rejected at the predetermined level of significance.

Thus formulation and testing of hypothesis is completely based on the theoretical background we have in mind, the theoretical predictions we want to make and the ways we want use those theoretical predictions.

A short list of books and articles used

1. Goon, A. M., M. K. Gupta and B. Dasgupta (1979), *Fundamentals of Statistics, Vol. I*, Calcutta: The World Press Pvt. Ltd.
2. Johnston, J. and J. DiNardo (1997), *Econometric Methods*, 4th edition, Mc-Graw Hill Book Company.
3. Kennedy, P. (1990), *A Guide to Econometrics*, 2nd ed, Basil Blackwell, Oxford.
4. Maddala, G. S. (1989), *Introduction to Econometrics*, Macmillan Publishing Company, New York.
5. Sadhu, A. N. and A. Singh (1985), *Research Methodology in Social Sciences*, Himalaya Publishing House.
6. Singleton, R. A. and B. C. Strait (1999), *Approaches to Social Research,* Oxford University Press.