On Hypotheses Testing: Some Views

Srilakshminarayana, G.
Assistant Professor, SDMIMD, Mysore
lakshminarayana@sdmimd.ac.in

Abstract

In the last decade, the world has seen many developments in statistical methods that are being used to test hypotheses constructed under different situations. Statistical significance is being used by many scientists from different fields to test their hypotheses. Proper use of a method gives the researchers the best results and helps them to take appropriate actions. The scientific procedures developed by many scientists also have some limitations, which the researchers have to take into consideration. But some of these methods are misunderstood as robust methods and are used very frequently in the research. The data collected should satisfy the associated assumptions to every method. If the data do not satisfy the assumptions, then the conclusions drawn need not be accurate and reliable. In this context, we study the fundamental aspects of these methods and present some of the important points that a researcher has to consider when he uses these methods to draw inferences about the hypotheses constructed. These are some views that have emerged from teaching and consulting.

Keywords: *Estimation, Parametric, Non-parametric, Hypothesis, Sample Size, Power of the Test, Robust.*

Introduction

"To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of."

Prof. R.A. Fisher Presidential Address to the First Indian Statistical Congress, 1938

For many years, researchers have been using statistical methods for the analysis of their research data, to draw valid and reliable conclusions with respect to their research objectives. They are advised to meet a statistician before they use the statistical tools, in order to ensure that an appropriate statistical method can be chosen as per the research objectives and check all the underlying assumptions of the method. The above lines quoted from Prof.R.A. Fisher's presidential address to the first Indian Statistical Congress, stresses the seriousness of this fact.

The present world has seen many changes in the way statistical inference is being used to draw inferences about the hypotheses constructed for different objectives. Researchers have been using statistical tests to test the significance of their studies, depending on a sample drawn from a population. It is interesting to observe that most of the researches hailing from various disciplines rely on statistical significance to support their claims made with respect to their research. Even the journals that publish the research papers anticipates that the research hypotheses be tested using statistical tests. Introduction of statistical significance has helped researchers in finding the strength of their hypotheses.

In order to use a method under statistical significance, the researchers have to be aware of all the assumptions associated with a statistical method. For example, the randomness of the sample drawn, assumption of normality etc. The sample drawn from a population under study plays an important role in constructing estimates of unknown parameters and these estimates are in turn used to draw inferences about the hypotheses constructed on the characteristics of the population. It is important to select an appropriate sampling technique to draw a sample. Many a times the sampling technique used is non-random, which violates the assumption of randomness of the sample. Even if the sampling technique used is non-random like judgmental or convenience sampling, the researchers are advised to test for randomness of the sample using appropriate non-parametric test. One such test is run test, which is being used to the randomness of the sample. Using a sample for a parametric test without adopting one of the above mentioned ways, may lead to wrong conclusions.

The following are the lines from the paper on t-test by Gosset (1930) (written under the pseudonym name "Student") with respect to the assumption of normality

"…This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed: yet it appears probable that the deviation from normality must be very extreme to lead to serious error."

These lines indicate that a t-test is more applicable to only those populations that follow normal law. When the population distribution is non-normal, t-test or any other test which has its root distribution as normal cannot be used to test the hypotheses. For example, Chi-square and F statistics that are being used frequently by the researchers require the assumption of normality. Keeping these in mind, statisticians have cautioned the scientists from other fields about the seriousness of this assumption. The book by Henry (2002) provides several methods to check the assumption of normality.

The methods developed by statisticians have their own limitations and the one who uses these methods have to be cautious about these limitations. One has to choose a method as per the situation and as per the assumptions of the study. It is the responsibility of the researcher to check whether the sample drawn satisfy all the assumptions of a method. Violation of atleast one of the assumptions may lead to false conclusions.

The other important aspect that a researcher has to take into consideration is about the measurement errors and outliers. The popular classical parametric methods that are being used for estimation and testing are vulnerable to both measurement errors and outliers. Measurement errors may occur due to faulty measurement tools or due to response bias. Outliers can be defined as those observations which are different from normal observations. The objective of the researchers has been changed from using a method to identifying a method which is insensitive to these limitations.

The other important aspects that the researchers have to take into consideration are the errors in hypothesis testing (type-I and type-II errors), power of the test. Note that the errors are conflicting in nature (i.e. if one tries to minimize type-I error, type-II error increases and vice versa) and hence, it is recommended by the statisticians to fix type-I error and then develop the test that will minimize the type-II error. Fixing the chances of committing type-I error in terms of level of significance is the usual practice adopted by the researchers. The random sample chose helps to minimize the size type-II error. Power of the test gives the strength of the test in correctly rejecting a hypothesis that has to be rejected. Note that using classic parametric tests, when the assumptions are violated, increases type-I error and decreases the power of the test. This aspect has been noted in Wilcox (1998).

Alternative methods such as non-parametric methods (Hollander and Wolfe (1999)) have been developed by the statisticians, when the data do not satisfy the assumptions. The development of robust statistical methods (Wilcox (2001, 2005)) has created avenue for the researchers to draw inferences in a better way as compared to parametric or non-parametric methods. But these robust methods couldn't gain popularity as the researchers are not aware of these methods. David and Vikki (2008), in their paper, have stated the importance of using robust statistical methods to classical parametric or non-parametric methods. They examined why the robust statistical methods are not popular as compared with classic parametric methods. It is interesting to note that even if the assumptions of a classical test are satisfied or violated, robust statistical can be used to draw inferences. It can be observed from their paper that add-ons to softwares like SPSS, SAS have been developed to use robust statistical methods. In fact, the paper has given good number of reference papers using which the researchers can comfortably use robust statistical methods via software.

In this paper, we present some views relating to the fundamentals of statistical significance. It is an attempt to bring out some of the subtle aspects relating to the methods used. Though the views presented in this paper are not new, the emphasis is to recall those fundamental aspects that were ignored by the researchers. But, the views presented are the author's own views, which have been drawn from his experience.

Definition of Variables

Before one chooses a statistical method for analysis, the variables under study should be defined properly and clearly. This is because the method has to be chosen based on type of the variable under study. If the variable is quantitative then the parameters under study will be the population mean and standard deviation. If the variable is qualitative then the parameter under study will be the population proportion. This also plays an important role in determining the sample size required for the study. The basic idea behind this is to ensure that a proper method has been identified for the variable defined. For example, researchers use different scales to quantify the responses. These have been developed by Likert (1932) and are called

Likert scales. These scales are meant to measure the responses that are qualitative. But, they do not qualify for quantitative treatment such as finding mean, standard deviation etc. Moreover, they come under a multinomial population, where the categories are multiple in number and the responses can be categorized to one of the categories. One important point that one has to keep in their mind is that the researchers have developed justifications to the use of quantitative techniques to likert scales which are so convincing. But, the end results need not be so accurate or reliable. One has to think of the alternative techniques available to model likert scales like non-parametric statistics. We end this section by concluding that the researcher should clearly define the variable(s) before a method is selected for further analysis.

Role of Sampling and Pilot Study

In this section, we discuss the importance of sampling and conducting a pilot study before the actual study.

Sampling plays an important role in testing hypotheses statistically. In almost all the studies, the researchers depend on the sample drawn from a given population, to draw valid inferences. In order to select a true representative sample, a standard sampling technique has to be chosen. Before this, it is advised to conduct a pilot study to understand the behavior of the respondents, to decide on the budget, to test the questionnaire prepared etc. A pilot study should be constructed in such a way that it gives a complete picture of the actual survey. One of the important reasons behind conducting a pilot survey is to determine appropriate sample size for the actual survey. The determination of sample size should depend directly on the resources available like money, time, number of interviewers etc. In other words one has to design the sampling technique in such a way that the sample size should be a function of atleast some of the resources available to conduct a study. Standard sampling techniques available in the literature also provides the researchers to determine appropriate sample sizes for their study.

Very few researchers from other fields conduct a pilot study to understand the behavior of the population and to test their design of the study. By conducting a pilot study, a researcher can design his study in a better way. A pilot study not only helps the researcher to understand the limitations of his study but also reveals the true picture of the past surveys which he takes as inspiration to conduct the present study. Many a times the present study will be based on the past studies and a pilot study will help a researcher to check the validity of the procedures used in these past studies for the present context. Once he takes care of all these only then he can start the present study.

There are several sampling techniques available. It is interesting to observe that the sampling techniques are being used according to the study and discipline. For example, researchers of wild life use line-transect sampling; see Anderson et.al (1979). One can adopt either a sampling design that was constructed for a given situation or a standard sampling design, for the study. The book of Cochran

(2007) presents good description of the standard sampling techniques. We only present the fundamental idea of these methods, avoiding deep technical details, as most of them can be found in standard books on sampling theory.

Caution on Using the Word 'Random'

Most of the researchers mention in their papers or talks that the sample is drawn at random. It has become common to use the word random to those samples which are collected based on convenient or judgmental sampling techniques. If one wants to use random, start with a proper sampling frame which has the information about the sampling units. Once the construction of a sampling frame has been done, use a random sampling technique which involves probability of selecting the sampling units to collect the sample. Only then one can use the word random. It is important because the standard hypothesis testing techniques assume that the data are collected at random. At this stage, if one questions what if the sample is a convenient sampling, the answer is to use some of the non-parametric tests like run test available to test the randomness of the sample.

Standard Sampling Procedures

The first one which is a simple sampling technique that is available and most frequently used is simple random sampling. In this method we first assign equal probability of selection to all the units. Then, we use lottery method or random numbers method to take the sample. The second technique is stratified random sampling. In this, we divide the heterogeneous population into subgroups such that the units within the groups are homogeneous and between they are heterogeneous. Now we divide the estimated sample size among these groups using either proportion allocation method or optimum allocation method (Cochran (2007)). The third technique is systematic random sampling. We select a number at random and all other units are selected automatically with equal spacing until the required sample size is met.

The other random sampling methods that were in practice are cluster, probability proportional to sample size, two phase, multi-phase etc. Since our focus is on hypothesis testing, we are not giving complete details of these methods. We emphasize on selecting appropriate sampling design to select a true representative of the population. These aspects seems to be simple to most of the researchers as they focus more on finding a suitable test and draw conclusions regarding their objectives. But, one has to be more focused on choosing a best group under reach as this group forms a basis for drawing inferences.

Probability Space

Note that the space in which the parametric methods are developed is a probability space. Theoretically, a probability space is developed under a random experiment and the set of all possible outcomes are listed to construct a sample space. Out of this sample space, a set of outcomes constitute an event. Using the probability function defined on the sample space, the occurrence of the events is measured. Similarly, the space under which a random sample is drawn is a probability space. The set of all possible samples constitutes a sample space and using a sampling technique, a random sample is drawn. One can take repeated samples and calculate the estimates using the estimator for different samples. The set of all estimates will give the sampling distribution of the estimator. The probability model assumed is normal model. Hence, using the normal probability function, one can find the required probabilities with respect to selecting a sample mean as an estimator. This can be found in confidence interval estimation. Using normal model, at a chosen level of significance, a confidence interval can be constructed. Here the point to be noted is that the entire estimation and hypothesis testing depends on a random sample. This is why the major assumption in statistical inference is that the sample is a random sample.

Role of Normal Distribution in Constructing a Test Statistic

This section deals with the role of normal distribution in developing a test statistic. Recall that, the first assumption is "the population from which the sample is drawn follows a normal distribution". That is, the random variable under study follows a normal law. Normal distribution plays an important role in parametric inference. This can be observed in some of the standard testing procedures. Note that a chi-square random variable is obtained by squaring a standard normal random variable, t-random variable is the ratio of standard normal to the square root of ratio of a chi-square random variable to its degrees of freedom, and

F-random variable is the ratio of two chi-square random variables divided by their respective degrees of freedom. A student of statistics is familiar with these transformations, as they are discussed in a basic distribution theory class. These aspects of distribution theory have been used to develop a chi-square, t, or F statistics. That is why statisticians insisted on assumption of normality in testing a hypothesis statistically. It is to be noted here that in most of the cases this assumption is not verified by the resear chers who use these testing procedures. Because of this the power of the test decreases and increases the type-I error (David and Vikki (2008)). One can argue that central limit theorem can be applied to justify this assumption, when the sample size is large. But the question which has no exact answer is how large should be the sample size to use the central limit theorem. Wilcox (2001) discusses these aspects in his book and cautions on use of central limit theorem. One need to understand that, theoretical concepts can be used only when the assumptions made during their developments are satisfied by the data considered.

Important Observation Regarding the Responses

One important observation that one can make is that every response either qualitative or quantitative can be viewed as the sum of actual response and the error. We are not providing any mathematical justification for this but heuristically one can see that the responses collected involves error which can be sometimes identified and minimized or cannot be identified. The error is a random variable and assumed to follow normal distribution. Because of this assumption and the type of relationship with error, the response variable is also said to follow normal distribution. Some arguments pertaining to this can be found in Huber (1972).

Another important aspect with respect to the role of normal distribution is it plays an important role in selecting a sampling distribution for an estimator. Note that the estimator is a function of random sample. The choice of the sampling distribution depends on the value of standard deviation. It the value of standard deviation is known then, the sampling distribution is a normal distribution. If the value is unknown, then the sampling distribution is taken as t-distribution, under the assumption that the population follows a normal distribution. Hence, it is very important that either the sampling distribution or the population distribution should be normal. Another advantage we have with normal model is that most of the discrete models like Binomial, Poisson etc., can be approximated when the sample size is sufficiently large.

Checking for Outliers

This is very important with respect to the analysis of the data. Once the sample has been selected, it is very important to check for the outliers/ extremes. This is because most of the methods used to test the hypotheses tend to give wrong conclusions under the influence of the outliers/extremes. This is very common when one is using sample mean as an estimator of population mean. Barnett and Lewis (1978) discuss methods which can be used to handle the outliers. They present the methods to handle the outliers for Univariate as well as Multivariate data.

It is clear that the data plays an important role in applied research. Care should be taken in cleaning the data. Proper outlier detection methods should be used to clean the data, before using statistical methods.

Estimation

Most of the researchers are familiar with word estimation. As pointed in the previous sections, the sample collected from the population plays a significant role in estimating the parameters of the population under study. A sample which is not a true representative of the population will not provide good estimates. The second aspect that one needs to be clear is about the method of estimation and the estimator used to estimate the parameters. For example, if one wants to find an estimator for a location parameter, the obvious choice would be sample mean. But, one has to be careful in selecting sample mean as an estimator because of the limitations of the sample mean. Sample mean is affected by extreme observations and outliers present in the data. One has to either ensure that the data is free of these or decide on how to handle these observations. In case of extreme observations, one can use trimmed mean as an estimator of the location. Here we would like to recall that the sampling distribution of the sample mean is normal and this fact is being used in hypothesis testing. Similarly, if one asks the question about the sampling distribution of trimmed mean, Stigler (1973) has shown that under some conditions on trimming the distribution of trimmed mean will be asymptotically normal. He also cautioned that if the trimming is not done at proportions corresponding to uniquely defined percentiles of the population distribution, the distribution will not be normal and the using trimmed mean may lead to invalid tests or confidence intervals, even with large samples. Huber (1964) presents methods that can be used to estimate location parameter through robust estimation.

Estimation Space

Another important aspect is estimation space. Estimation space of a parameter is defined as, the space which includes all admissible values of the parameter under study. One can raise a question as how to identify this space. This space can be identified using the phenomena under study and the population from where the sample has been drawn. This step has to be taken into consideration before estimating the parameters. The estimator used to estimate the parameter should be selected with utmost care. This is because the efficiency of the estimator will have on impact on the testing procedures.

Classical Parametric and Non Parametric tests: Some concerns in Their Usage

In this section, we discuss some of the parametric and non-parametric tests available in the literature. The most frequently used testing procedures are t-tests, chi-square tests and F-test. Note that F-statistic is being used in Analysis of Variance (ANOVA). The alternate non-parametric procedures are Sign test, Kruskal-Wallis test etc. At the end of this section, we present our views on why researchers blame statistical significance, for their failures.

T-test is used when there is one (two) population(s) under study and the hypothesis is constructed on population mean(s). This is a small-sample approach i.e., it is used when we have small samples for the study. We select a sample from this population, find the estimates of the parameters and then use them in the test statistic to draw inferences. After constructing the null and alternative hypotheses, it is important to set the level of significance (size of type-I error). Let us spend some time on assumptions of this test.

1. The population from where the sample is drawn

   follows normal distribution.
2. Then population variances are unknown.
3. The responses are measured independently.
4. The sample is a random sample.
5. Equality of Variances.

The first assumption is very important with respect to the computation of the sampling distribution of the statistic. The second assumption indicates that the population variance should be estimated using the sample data. The responses should be collected independently because of the theoretical assumption.

Before using this test, one has to check whether the data satisfy these assumptions. The assumption of normality can be checked using Kolmogorov-Smirnov Test, assumption of equality of variances can be checked using F-test. There are instances where researchers claim that t-test is robust to violation of assumptions. But Sawilowsky and Blair (1992) noted that t-test is robust to violation of normality assumption when the test is two tailed, sample sizes are equal, sample sizes are 25 are more and variances are equal (in case of two sample problem).

Chi-square test also assumes that the population follows a normal distribution. This test can be used to test significance of population variance, test independence of attributes, difference between multiple proportions and goodness of fit of a probability model chosen for the situation under study. If the assumption of normality is violated, this test fails to give proper results. The major disadvantage is that it increases the size of type-I error and decreases the power of the test.

Analysis of variance (ANOVA) is used to the significance of several means. It is also used to check the effect of one or more factors on the population considered for the study. Depending on the number of factors, this test can be classified as one-way, two-way, and multi- way classification. To use this, first one has to define the factors under study and then classify the population based on the factors defined. Now samples are drawn from each group and then the test is applied to check the effect of these factors. The basic assumption here is that the errors are normally distributed. The other assumptions are errors are uncorrelated and errors have constant variance. These assumptions have to be tested before the test is used. Using this test under the failure of the assumptions may lead to wrong inferences. This test is used in design of experiments to check the effect of the factors used to design the experiments.

Sign test is an alternative to the t-test for single mean. The major difference is that this test does not need the assumption of normality and the parameter considered is Median.

Kruskal-Wallis test is an alternative to ANOVA and used when the assumptions of ANOVA are not satisfied by the data.

Note that, we are not discussing these methods in depth as the discussion about these methods can be found in any fundamental book on statistical inference. A good reference could be Levin and Rubin (2007) for the beginners and Kale (2000) for advanced users.

Caution in Using the Testing Procedures

In order to use the above mentioned test procedures, one has to understand the procedures completely and match the conditions prevailing under their studies with those conditions under which the methods have been developed or used. For example, suppose that researcher-1 uses t-test for his data and concludes his work successfully. There is a tendency that other researchers in the same field who are doing a similar study will use t-test and conclude. This continues and generations of researchers will be using only t-test. Even the guides who guide the students encourage the students to use only t-test because the last student has used t-test. The basis for using t-test is a thesis that first uses t-test for a similar study. Here they forget the fact that when researcher-1 used a t-test, he ensured that the data satisfy all the conditions of a t-test. But, other researchers use it without understanding this fact. By doing so, they misuse statistical significance and produce wrong results. This what some people call as “Cult of Statistical Significance”. Who should be blamed for this? It is the guide who suggests the student to use a procedure without knowing whether it can be applied or not. The result of this mistake is that when the results go wrong the researchers comfortably blame the statistical significance procedures. They claim that the theory failed. But the fact is that they failed to identify a suitable alternative method for their work. One has to understand that though it seems simple, it is actually creating lot of problems among the researchers. When one researcher raises concerns with respect to the method used, only then others wake up and start blaming the statisticians for developing such methods which are not suitable for their studies. They do not accept the fact that their inefficiency in selecting a suitable procedure/method.

Power of the Test

In this section we present the procedure used to compute the power of the test.

Step 1: Construct the test statistic using the sample drawn and find the critical region.

Step 2: Now find the non-critical region, which is opposite to critical region.

Step 3: Find the probability of type-II error, â.

Step 4: The power of the test=1- â.

This is the common procedure that has been adopted over years to compute the power of the test. Cohen (1988) discusses the advantage of computing the power of the test. This helps the researcher

a. To measure the strength of their procedure.

b. To choose a best alternative from the available alternatives.

There are good softwares available to compute the power of the test. For example, G*Power can be used to compute the power of the test. The test procedure should end with stating the power of the test.

To calculate the power of the test, one has to select a value for the parameter under alternative hypothesis. Since, the power is complement to the size of type-II error this value has to be chosen from critical region. We discuss this using an example. Consider t-test for single mean. Suppose that the null hypothesis is to test population mean is equal to 45 against the alternative hypothesis population mean is greater than 45, at 5% level of significance. The sample size is 25 with sample mean 47 and standard deviation 2.485. The p-value for this test is 0.00024, which is less than level of significance. Hence, we reject the null hypothesis and conclude that the sample is sufficient to provide evidence against the null hypothesis. Now the critical region for this is computed in the following way.

Step1: Reject H_0 is

$$t_{cal} > t_{\alpha}$$

$$\Rightarrow \frac{\overline{X}-\mu_0}{\frac{s}{\sqrt{n}}} > t_{\alpha}$$

$$\Rightarrow \overline{X} > \mu_0 + (t_{\alpha} * \frac{s}{\sqrt{n}})$$

This gives the upper critical region value. Now, the non-critical region is

$$\overline{X} \leq \mu_0 + (t_{\alpha} * \frac{s}{\sqrt{n}})$$

Step2: The size of type-II error â is computed using the following procedure

$$\beta = P\,(not\ rejecting\ H_0 | reject\ H_0)$$

$$\Rightarrow \beta = P\left(\overline{X} \leq \mu_0 + (t_{\alpha} * \frac{s}{\sqrt{n}}) | reject\ H_0\right)$$

Now one has to choose a value for the population mean under alternative hypothesis from the critical region. The critical region for the above problem is set of all values above 45.85. This is computed by substituting the values in the above inequality. Now one can choose a value that is greater than 45.85. Let us choose the value as 47. For this value, the power is 0.98. One can notice that if the alternative value increases then the power also increases.

One important point that we would like to mention is that. The size of type-I error has to be fixed before taking the sample and a random sample has to be drawn such that the size of type-II error decreases. This subtle aspect can also be found in the definition of null hypothesis

"Null hypothesis is a hypothesis, which is being tested for its possible rejection".

It clearly states that we do not reject the null hypothesis in the beginning but we take a random sample and test for its possible rejection. That is, if the sample is a true representative then the testing procedure developed based on the estimates of this sample will have good power reject a null hypothesis correctly, for a suitably chosen value of the parameter under the alternative hypothesis. It is interesting to note that by computing the power, one can also assess the strength of their sample, in truly identifying a wrong specification of the parameter.

At this point we would like to say that the books that discuss procedures to test the hypothesis gave emphasis to power of the test only after researchers started enquiring about methods that can be used to measure the strength of the testing procedures developed on the basis of their samples.

Cult of Statistical Significance

In this section, we present some views on misuse of statistical significance. It is interesting to note that there are group of researchers who criticize proving science based on statistical significance. We consider this an important part, as it cautions the researcher before using statistical significance to draw inferences about their research. The first paper that we would like to quote is by Ziliak and McCloskey (2009). The following are the lines taken from this paper

“We want to persuade you of one claim: that William Sealy Gosset (1876-1937)-aka “Student” of “Student’s” t-test-was right, and that his difficult friend, Ronald A. Fisher (1890-1962), though genius, was wrong. “

They mention that when Gosset introduced the concept of standard error, he instructed the user to use it consciously. But, Fisher erased the word consciously and used it. They bring out the true error in using standard error in scientific research. It is interesting to note that most of the researchers who made statistical significance popular were blamed for its misuse by the researchers in science. This is not restricted to one field.

Most of the researchers in different fields use statistical significance to prove or disprove their hypotheses based one sample. But, the question is how one can generalize the results based on one sample. If one has to generalize this then they have to take several samples in different places under different conditions and then use statistical significance. The models and hypotheses that were tested using statistical significance worked only few times and other times the researchers blamed statistical significance for their failure. This can be observed in the recent financial recession. People used a mathematical model for their predictions and finally blamed the model and the developer for its failure. But, they forgot to check the assumptions associated with the model. In such instances one cannot blame the model used or statistical significance. The user himself should be blamed for not understanding the model and using it without checking whether the data satisfy the assumptions. At this point we don't agree with Ziliak and McCloskey (2009). The reason is that when statisticians developed the models and statistical significance, they cautioned others about their misuse. But, the researchers from other fields ignored their words as they are more interested to finish their thesis work or to get an immediate answer for their hypotheses. They ignored the fact that statistical significance has its own limitations and should be used with caution. Decision making cannot depend on one p-value. It should take into consideration other aspects like availability of the resources and feasibility of applying the results etc.

Another interesting aspect is that statistical significance has become a part of almost all the research works where data and analysis of the data plays an important role. Statistical significance helps the researcher to decide whether the sample considered provide sufficient evidence either to reject the hypotheses or not. Statistical significance has become an important tool to study the significance of the results.

We finally end this section by stating that, when one wishes to use statistical significance, he has to understand that the every method has its unique limitations and it is up to the researcher/ decision maker to decide whether to choose the method or not. As mentioned in the beginning of the paper, bringing the data to a statistician is like bringing a patient's body for postmortem to a doctor. The life of a patient can be saved if he is brought to the doctor before and not after. Similarly, a researcher/ decision maker has to meet a statistician before the study and not after the study. This is because the statistician will help the researcher by mentioning the assumptions that the data has to satisfy before the study. He also can design the study as per the requirements in such a way that suitable statistical methods can be used for further analysis.

Modern Statistical Methods

Robust statistical methods are considered as the modern statistical methods, which are being used in analyzing the data sets. The main purpose of these methods is to provide a sound method which is insensitive to the measurement errors or outliers. Demand for robustness of a method is very old and has increased with the usage of statistical softwares. The first generation of robustness dealt with developing methods which can identify outliers and clean the data before any statistical treatment. The second generation of robustness is dealing with raising questions on the validity of the classical parametric methods and developing advanced methods which are insensitive to violation of assumptions and outliers. One has to understand that the classical parametric methods are developed under the assumption of a base model for example normal model. It is natural to ask what happens if an alternative base model is used instead of the one used while developing the method. This also includes the estimation procedures that are used to get optimum estimates of the population parameters. The methods developed after 1960s have provided avenues to the researchers use more advanced and robust methods, which in turn helped them draw powerful conclusions as compared with classical methods. Here power means power of the test. This is one aspect that researchers have to take into consideration when choosing a testing method. Note that robust methods can be used either the assumption of classical methods are met or not met. Morgenthaler (2007) provides an interesting discussion on robust statistics. It is a short discussion on robust statistics but it makes the reader think in new directions. This paper is tagged with comments from scientists who are working in the area of robust statistical methods.

There are several alternative robust methods available in the literature that can be used in the place of classical parametric tests, when the sample does not satisfy the assumptions. Huber's paper in the year 1972 lists some of the robust statistics. It is a review paper which deals with methods to estimate location parameter. As mentioned in the previous sections, sample mean can be used as an estimator of a population mean only if the sample observations are free from outliers/extremes. In this paper, Huber discusses several alternative methods in the place of sample mean as an estimator to location parameter. Similarly, there are methods to handle statistical significance and modeling.

Concluding Remarks

We would like to conclude the discussion by saying that the views expressed this paper may not be new but are just a reminder to the researchers who use statistical significance. This attempt was made by taking into consideration the philosophical saying that things known should be repeated again and again till they are applied in a proper way.

Acknowledgement

The author thanks the referee for giving constructive remarks that helped to present the paper in a better way.

References

Anderson, D.R., Laake, J. L., Crain, B. R., and Burnham, K.P. (1979): Guidelines for line transect sampling of biological populations. Journal of wildlife Management, 43(1), 70-78.

Barnett, V. and Lewis, T. (1978): Outliers in statistical data. John Wiley & Sons, Inc. New York.

Cochran, W.G. (2007): Sampling techniques. 3rd edition, John Wiley and Sons.

David M, E., and Vikki, W. M. (2008): Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. American Psychologist, Vol63, No.7, 591-601.

Henry, C. Thode, Jr (2002): Testing for normality. Marcel Dekker, Inc. New York.

Hollander, M., and Wolfe, D.A. (1999): Nonparametric Statistical methods. 2nd edition. John Wiley & Sons, Inc. New York.

Huber, P.J. (1972): Robust statistics: A review. The Annals of Mathematical Statistics. Vol.43, No.4, 1041-1067.

Huber, P.J., and Ronchetti, E.M. (2009): Robust statistics.2nd edition, John Wiley & Sons, Inc. New York.

Huber, P.J. (1964): Robust estimation of a location parameter. Ann.Math. Statist. Vol.35, No.1, 73-101.

Kale, B.K. (2000): A first course on parametric inference. Narosa Publishing House.

Levin, R.I. and Rubin, D.S. (2000): Statistics for Management. 7th Edition. Pearson.

Likert, Rensis (1932): A Technique for the Measurement of Attitudes. Archives of Psychology, 140: 1–55.

Morgenthaler, S. (2007): A survey of robust statistics. Statistical Methods and Applications. 15:271-293.

Sawilowsky, S. S., and Blair, R. C. (1992): A more realistic look at the robustness and type II error properties of the t test to departures from population normality. Psychological Bulletin, 111, 353-360.

Stigler S., M. (1973): The asymptotic distribution of the trimmed mean. The Annals of Statistics, Vol.1, No.3, 472-477.

Student (1908): The probable error of mean. Biometrika, Vol.6, 1-25.

Wilcox R, R. (1998): How Many Discoveries Have Been Lost by ignoring modern statistical methods? American Psychologist, Vol.53, No.3, 300-314.

Wilcox R, R. (2005): Introduction to robust estimation and hypothesis testing. 2nd edition, Elsevier Academic Press.

Wilcox R, R. (2010): Fundamentals of modern statistical methods. New York: Springer.

Ziliak, S.T. and McCloskey, D.N. (2009): The cult of statistical significance. Section on statistical education-JSM, 2302-2316.