“10 Tips to Get a Successful Statistical Analysis”

Mubashra Mehmood
0937-BH-STAT-10

Stepping into real life, you come across a lot of fields which directly or indirectly relates to the worthy “Statistics”. Doing any kind of analysis whether you are a student, a marketer or a statistician on any professional, you always wish to have an exact or true statistics or in common words-true facts and figure about the analysis which you are going to deal with. Statistics always helps you to get good analysis but not always, sometimes you do not get desired results due to some common statistical mistakes. These mistakes are easy to correct but difficult to identify. Following are their respective details:

1. Differentiate between causation and correlation
Whenever you have two variables, there is always a distinct correlation between them. Correlation is a measure of identifying dependency between two variables. For a good analysis we must have a correlation between variables and there always exists correlation in any two variables but having a good correlation doesn’t mean that these variables do have true relationship in real life. Causation lets us know whether these variables have a relationship in real life or not? There is always causation which gives guarantee to correlation. So, it must be differentiated. Relying upon causation is not enough.

2. Role of regression
Prior regression analysis always supports a researcher to overcome multiple problems while doing an analysis. One should look critically upon data, apply regression on it so that you may know what type of analysis you have? You may have a liner, multiple or irrelevant type of regression. This will help you in identifying a suitable model for your data. After identifying it, if you have multiple regressions then having a huge correlation between two variables can be labeled as a case of multicollinearity. So these possibilities should be kept in mind.

3. Residuals in regression
Keep an eye upon the residuals and their squares too. If they have different patterns and their variances change then you may encounter the statistical problem of heteroscedasticity.

4. Sample of population
The population is always unknown, that is why sample is taken from it. A sample should have all characteristics of population. For this, sample selection method should be sample per unit. A reasonable sample size is to be selected neither so large that it becomes difficult to analyze. This avoids complexity. Outlier should be avoided while sample is taken. This may cause a huge change in the averages.

5. Decision about conclusion of statistical analysis
A statistical analysis should not be concluded unless and until you run proper results of hypothesis test. Hypothesis test gives a clear direction to the conclusion about a problem. Level of certainty should be predetermined and consider necessary for the comparisons. Usually taken as 95% standard or when it is not predetermined. Unless you have significant result about the statement of problem you cannot conclude a statistical analysis.

6. Differentiate between normal or non-normal distribution
There several statistical distributions which are classified into two main categories: normal or non-normal. One should test normality first which has many other methods of identifying e.g. test of chi-square (goodness of fit) or through graphically (through histogram). Another check involves identification when you have more than one large sample group. Then according to “central limit theory” each group comprises of 30 elements, their respective means will follow a normal distribution. After differentiating you may be able to apply suitable distribution for the data.

7. Using a suitable test
While doing a hypothesis test, one should consider all variables rather than two main variables which are giving a huge contribution. This is due to the reason that in case of two variables you go for a t-test and in case of more than two, the whole analysis changes and switches to ANOVA. So, number of independent variables is to be counted. Furthermore, in case of t-test you also keep sample size into account (use for small sample size). The variances have no condition to be known.

8. Covariance analysis or correlation analysis
Correlation analysis do not depend upon the scale of measurement you opt. It always gives same results for complete data sets whether you use different scales of measurements or not. For Covariance analysis the change in the scale of measurement matters a lot. As you change the scale the whole analysis will be changed. So, difference between both of them should be considered mandatory.

9. Check for Accuracy
If you want to check the accuracy of the analysis then I suggest checking it by using single-tailed test why? Because using two-tailed test usually give results by highlighting the area of uncertainty, while on the other hand one-tailed test not only give results but also the direction of means in different direction.

10. Analysis for unknown population with small sample size
When we have small sample size and our population is not known then we usually consider that population is normal and use t-test. In contrary, sometimes population is known and sample size is small but the known population is not normal then we use non-parametric test or distribution free tests. These tests are valid rather than t-test.

The author is an associate editor of The Scientific Ravi.

Quality for customer; part and parcel

Syed Yasar Mahmood
Lecturer

Undoubtedly, there is always a hard decision of choice among the competing things or products and it becomes even more than a headache when products are almost of
same standard. This fact is rife; either the buyer is an individual, an industrial organization or any other institution. Normally one remains in an ambiguous state regarding which product is preferred over the other. But the answer of the above question can only be delivered by a person called “Customer or Consumer” because a consumer has his own priorities for the product to be used in daily life but one thing which is never compromised is “Quality”. So the only line of discrimination between the competing things is quality which throws a clear picture between a bad quality product and a good one. But the decision of bad and good product is again on the disposal of consumer. We observed this whole war in between customer’s own choice. Every potential customer has his own definition of quality and measures to scan a product. If a product is a quality product for one customer, it may be a bad one for the other as satisfaction of customer varies from one another. A quality product may be defined as a product which meets certain specifications. Due to various parameters associated with it, it is very difficult to define the term quality. Secondly, the tolerance of customers for products has rapidly decreased. Almost every customer looks for the following aspects in a product.

Some prefer how well a product will perform specified functions for which it is designed. Reliability of the product is one of the major characteristics which appeals a customer. Different appliances require repair after some time. A product is considered to be of good quality which requires less number of repairs. What is the effective service life of a product is another question. Of course, one will prefer a product with long service life. If a repair is required, how easily it can be repaired is another question. Customer will feel at ease if it is possible for a product to be repaired. A major dimension of quality is product present look. Aesthetics play a vital role in sale of a product. A high percentage of customers feel satisfied and have purchasing mood even when the product is not of good quality but they still prefer the product just based on aesthetics. Style, color, packing etc. are salient features of aesthetics. Added features in a product are of significant importance. It’s obvious; a customer will purchase a product with added features. Like, if two mobile companies are of equal standard and manufacturing mobiles of almost equal price. Customer choice will be for that product which has more features. Quality also termed as perceived quality. Every manufacturer or brand has its own repute in the market. A customer will show his/her inclination towards that product which is well known by the other customers due to certain features. Certainly, it takes time for every product to get people’s attention and a commanding place in the market where no other product is its rival. These are all the parameters which customers usually considered when purchasing the products.

Statistical quality control is one of the subjects of statistics which is purely based on the quality improvement and quality control. Different techniques are being used for quality improvement and control. Different industries have their own quality control departments and their officials continuously monitor the quality of a product. Statisticians work in different quality control sections and analyze the collected data to have a continuous check on the product. Whether the product is doing the intended job? Main focus of the manufacturers is customer. Manufacturers include those features in products which customers demand. Whatever the aesthetics, features and how well the product performs; does not matter but the customer satisfaction does. If a customer is satisfied then everything is good otherwise a good quality product may not appeal the customer and it may be declared a product of inferior quality.

The pith and marrow of the whole discussion is that quality and customers are both very important for the manufacturers. Competing companies never want their products to be rejected or returned for replacement and rework. To meet these challenges, companies arrange surveys in turn to know the popularity of their products and their perceived quality. To sum up, quality is part and parcel for customers and hence quality can be referred to as “Customer Satisfaction”.

### Suitable Statistical Models with respect to the nature of Studied Variables

Zeeshan Ali Shammas
901-BH-STAT-09

The purpose of this article is to help the reader in selection of suitable statistical models with respect to the nature of studied variables. By a statistical model, we mean the inexact relationship between the studied variables.

There are usually two types of variables which are studied in statistical models: dependent variables and independent variables. As the name suggests, a variable that depends on other variable(s) is known as dependent variable; the alternative names for dependent variable are response variable, controlled variable, needy variable, regressant variable, predicant variable, non-orthogonal variable, outcome, explained variable and endogenous variable. Statistical models are generally used to estimate the effect of other variable(s) on this response variable. These “other variables” are known as independent variables, control variables, regressor variables, predictor variables, orthogonal variables, explanatory variables, stimulus, factor, covariates and exogenous variables. (In this article, the terms response variable for dependent variable and explanatory variable for independent variable have been used).

Response variable and explanatory variable(s) may either be quantitative or qualitative. A quantitative variable is one which describes a (numerical) quantity; it can either be continuous (measurable quantitative variable e.g. time, speed etc.) or discrete (countable quantitative variable e.g. number of rooms, number of students etc.). If a variable is qualitative in nature, it means that it describes a quality. Qualitative variables are also known as categorical variables. Categorical variable can be dichotomous
/binary (having just two categories e.g. pass and fail, good items and defective items etc.), nominal (having just names e.g. gender (male, female), eye colors, religion etc.) or ordinal (having name and order both e.g. stages of cancer (1st stage, 2nd stage), top three positions in a car race etc.). In statistics, qualitative variables do not have any role directly. Firstly, these variables are converted into some measurable or calculable variables by giving the categories (or qualities) some codes; then these are studied in statistics.

The choice of a suitable statistical model depends upon the nature of response and explanatory variables. If a model is selected without caring about the nature of studied variables (response variable and explanatory variable(s)), then the analysis would not exhibit the proper inference about population. Therefore, selecting a suitable model is very important. Some guidelines for this purpose are given below.

1. If both the response and explanatory variables are quantitative, “Simple Linear Regression Model” should be used. Response in this model is always continuous.
2. If response variable is quantitative (continuous) and all explanatory variables are qualitative (binary or dummies), then “Analysis of Variance Model” (ANOVA Model) should be used.
3. If response variable is quantitative (continuous) and explanatory variables are a mixture of quantitative and qualitative (binary or dummy) variables, then “Analysis of Covariance Model” (ANCOVA Model) should be used.
4. If response variable is binary and all explanatory variables are discrete or continuous or mixture of both discrete and continuous variables, then “Binary Logistic Regression Model” should be used. Besides that if response variable is dichotomous (binary), then “The Linear Probability Model” (LPM) or the “Probit Model” can also be used.
5. If response variable is nominal and all explanatory variables are discrete or continuous or a mixture of both discrete and continuous variables, then “Multinomial Logit Regression Model” or “Conditional Logit Model” or “Probit Regression Model” or “Nested Logit Model” (depending on the situation) can be used.
6. If response variable is ordinal and all explanatory variables are discrete or continuous or a mixture of both discrete and continuous variables, then either “Generalized Ordinal Logit Regression Model” or “Generalized Ordinal Probit Regression Model” can be used.
7. For count data, “Log Linear Model” is used.

Vital Statistics In Our Daily Life

Ahtisham Ali
928-BH-STAT-10

In our daily life many events take place. Some events are related to the changes in the composition and size of human population. These events are termed as vital events. For example births, deaths, migrations, divorces and adoption etc. When births take place it means that the population is increased and when a death occurs then population decreases. All these events bring changes in the size of human population.

But some events bring changes in the composition of human population. For example divorce cases, marriages, adoptions, migration and legitimating etc.

Births, Deaths and Migration are termed as components of population growth. All these events are vital events and the study of these events is called Vital statistics. All these events are recorded continuously. We often use the term demography will throw some light on what the demography actually means.

Demography is a Greek term which simply means the description of the people. Demography is defined by the United Nations as follows:

“The scientific study of the human population primarily with respect to its size, structure and development. In demography, we first collect the data on the vital events and then the analysis of the data is begun using different rates and ratios. Finally, we give the interpretations of the findings with reference to the social, biological economic and historic background.

Registration of vital events in Pakistan:-

In Pakistan, the provincial government registers the vital events through the local bodies. It is the responsibility of provincial government to report these registered vital events to the corresponding registration offices. The registration of vital events in the rural and urban areas of Pakistan is carried out separately.

- **In rural areas:-**
  The vital events are registered in the rural areas of Pakistan by local councils. The head of the household is required to register the vital events within 4 days of its occurrence. The household may report to the chowkidar or may get it registered at union council office.

- **In urban areas:-**
  The municipal committees, municipal corporations and cantonment boards are required to register the vital events. In case of birth, the midwife is also required to send report to the registration office.

In vital statistics, different rates and ratios are used to see the characteristics of a population but I will deal with some important rates and ratios.

**Sex Ratio:-**
It is the ratio between males and females in a human population. It tells us that how many males are there against 100 females. If ratio is greater than 100, it will indicate that the number of males in a population is higher than the number of females in a population.

**Birth-Death ratio:-**
It is the ratio between total number of births and total number of deaths in a population in a particular year. If the ratio is greater than 100 then it means that the population is increasing and if it is less than 100 then it indicates that the population is decreasing.

**Age specific death rate:-**
In age specific death rate, the death rates are computed for each age group separately. If we calculate age specific
death rates for males and females separately then it will be called as age sex specific death rate.

**Infant Mortality rate:-**

This rate is very important because it is used to determine the health level of a society. A low infant mortality rate indicates that the medical facilities are well enough, and hygienic conditions are also good.

**Importance of vital statistics:-**

- To estimate the population in a particular area. So in this way we will be able to determine the needs of these people. For example, how much wheat should be cultivated to fulfill the demand of wheat in a particular area?
- The death rates or mortality rates are the health indicators. With the help of these rates, we can come to know that whether the health and medical care conditions are sufficient or not.
- Vital statistics is also used by the Government for planning of facilities. For example, education facilities, medical facilities etc.
- Vital statistics tell us about the current trend in the population of a community i.e. whether it is increasing or decreasing.

---

**Is statistical analysis any worth of giving importance?**

Syeda Maryam Tahira Naqvi  
1174-BH-ECON-12

Statistics is a subject which is concerned with the mathematical study of data. In massive amount of data sets it becomes impossible and impractical to quickly analyze each tiny piece of data. So, then a sample of the data is studied thoroughly for the remainder of results of the data which are taken out from the sample data. This gives mathematicians faster results and an access to other mathematical exercises. Various other types of data interpretation as well as trend analysis and frequency and distribution analysis can make a good use of this faster and much more efficient technique of studying only a part of large data.

Statistical analysis primarily refers to a collection or a variety of methods used to process vast amount of data and to report overall trends. It is particularly useful when dealing with noisy data and in addition to that it provides ways to objectively report on how unusual and infrequent an event can be based on a historical data.

Many scientists, sociologists, pollsters along with various other professionals use statistical analysis in order to extract the information they need. It also provides meaning to what otherwise would have had been called a collection of data numbers and values. Not only in academics but also in business affairs, statistical analysis plays an important role in helping the administration in accumulating immediate feedback from the customers or clients.

Statistical Analysis has many uses, but its two main purposes include the generation of hypothesis regarding the population from which the data was taken out for sampling and to be useful to the researchers while verifying the hypothesis and theories by keeping in mind a specific purpose for the collection of data. In any case, the type of data chosen for analysis plays an important role as an element which affects the entire analysis procedure.

Statisticians often use chi-square statistics in their analysis of nominal data. Similarly, the use of ordinal data allows the researchers to hold the ordering of classification during the procedure of analysis.

Researchers often use numerical values as a method of formulating one face to another in the case of dealing with the interval data which takes the form of all possible numbers, apart from the number zero which has no meaning for an interval data.

Ratio data analysis is used in statistical analysis and is quite similar to interval data as it is also in the numerical form apart from one thing which distinguishes it well from the interval data and that is the number zero which plays a significant role in it. These objective definitions enable the researchers to interpret the ratio data well. Types of statistical analysis which examine the ratio data are many with common linear form of models and logistic type of regression.

Statistics is considered to be a strong tool and it can provide the people with many helpful answers. However, there are a few drawbacks of the statistical analysis. As every coin has two sides, similarly Statistical Analysis has both of its benefits and drawbacks together with it. One of the main disadvantages is that while performing a statistical test we normally use a sample which is taken out from a population, then using the entire sample data we reach conclusions about that population. Many of the other disadvantages of Statistics arise from the use of Samples.

All in all, we can safely conclude that Statistical Analysis is worth of giving importance as its advantages outweigh its disadvantages and its use in regards to the purpose it serves in analyzing the vast amount of data cannot be ignored. Hence, the use of Statistical Analysis provides an easier way and less time consuming technique to the people when they want to investigate large amount of given information.
Backscatter

Sequential Profile Adjustment by Regression (SPAR)

A new statistical analysis practice that identifies and removes systematic predisposition, blare and apparatus-based artifacts from observational data could escort to more clear-cut and consistent measurement of nanomaterials and nanostructures probable to have future industrial applications. Beyond nanomaterials and nanostructures, the method could also perk up consistency and meticulousness in nanoelectronics measurements and in study of certain larger scale system. Exact perceptive of these properties is significant to the development of future elevated industrial applications for nanomaterials and nanostructures because producers will necessitate reliability in their yield. Known as sequential profile adjustment by regression (SPAR), the procedure could also shrink the amount of observational data required to make conclusions, and help differentiate true nanoscale phenomena from experimental miscalculation. Researchers plans to use the SPAR technique on future work, and to explore past research for prospective new findings.