CONTINUOUS MEDICAL EDUCATION (CME)
pISSN 0976 3325│eISSN 2229 6816
Open Access Article
www.njcmindia.org

# Medical Statistics Series: Inferential Statistics (Part-I)

**Swati Patel1**

1Statastician cum Assistant Professor, Department of Community Medicine, SMIMER, Surat

## ABSTRACT

Hypothesis testing (or statistical inference) is one of the most important applications of biostatistics. Most of medical research begins with a research question that can be framed as a hypothesis. There are two type of hypothesis in inferential statistic, Null hypothesis reflects that no difference in comparison to baseline or between groups, whereas an investigator/researcher has some reason to accept difference in comparison to baseline or between groups is known as alternative hypothesis. Since $H_0$ must be either true or false, there are only two possible correct outcomes in an inferential test; correct rejection of $H_0$ when it is false, and retaining $H_0$ when it is true. Therefore, there are two possible errors that can be made which have been termed Type I and Type II errors. A type I error occurs when $H_0$ is incorrectly rejected. This is commonly termed a false positive. A type II error occurs when $H_0$ is retained when it is in fact false. This error is commonly termed a false negative. This article explained the statistical hypothesis, type of errors, confidence interval, P- Value and concept of normality.

**Key words:** Statistical Inference, Hypothesis, Type of Errors, P-Value, Concept of Normality

## INTRODUCTION

Instead, inferential statistics helps to suggest explanations for a situation or phenomenon whereas, descriptive statistics is used to describe numerical and graphical presentation and in summarization of observations. It allows drawing conclusions based on extrapolations, and is in that way fundamentally different from descriptive statistics which merely summarize the measured data.[1]

Biological phenomena are essentially variable and in this age of “evidence-based medicine” an understanding of such variation through statistical approaches is essential not only for the medical researcher who intends to draw inferences from his sample, but also for the practicing clinician, and the medical teacher whose responsibility is to critically appraise the presented inferences before accepting them into practice or the curriculum. Development of new drugs, devices, and techniques is heavily dependent, nowadays, upon statistical analyses to prove their effectiveness. It is also frequently alleged that statistical analyses can be misused and misrepresented. This should be a further impetus for understanding inferential statistics.

Inferential statistics mostly includes the testing of hypothesis, various statistical methods (parametric and non - parametric), prediction analysis etc. This article discusses about the hypothesis testing, When to apply parametric and non-parametric test.

## HYPOTHESIS TESTING

The main objective of the testing of hypothesis in any research is to know whether the change observed or effectiveness is by chance or actually it occurs.

Hypotheses are the research questions to be answered from experiments or studies. They are made in reference to populations. In, statistics there are two type of hypothesis (i) Null Hypothesis (ii) Alternative Hypothesis.

The null hypothesis is the one that states there are no effect/ no difference /no change between two or more groups, whereas the alternate hypothesis is defined about effect/change or difference.

**How to cite this article:** Patel S. Medical Statistics Series: Inferential Statistics (Part-I). Natl J Community Med 2021;12(7):204-208.

**Financial Support:** None declared **Conflict of Interest:** None declared

**Date of Submission**: 04-07-2021; **Date of Acceptance**: 26-07-2021; **Date of Publication**: 31-07-2021

**Correspondence:** Dr. Swati Patel (E-mail: Swati84patel@gmail.com)

Statistical hypothesis testing is the use of data in deciding between two (or more) different possibilities in order to resolve an issue in an ambiguous situation. Hypothesis testing produces a definite decision about which of the possibilities is correct, based on data. The procedure is to collect data that will help decide among the possibilities and to use careful statistical analysis for extra power when the answer is not obvious from just glancing at the data.

The first step in testing hypotheses is the transformation of the research question into a null hypothesis and alternative hypothesis, which is denoted as H0 and HA respectively.

The null and alternative hypotheses are two mutually exclusive statements about a population. A hypothesis test uses sample data to determine whether to reject the null hypothesis.

**Null hypothesis (H0)**

The null hypothesis states that a population parameter (such as the mean, the standard deviation, proportion and so on) is equal to a hypothesized value. The null hypothesis is often an initial claim that is based on previous analyses or specialized knowledge.

H0: No association between smoking and lung cancer

**Alternative Hypothesis (H1)**

The alternative hypothesis states that a population parameter is smaller, greater, or different than the hypothesized value in the null hypothesis. The alternative hypothesis is what you might believe to be true or hope to prove true.

H1: Smoking is associated with lung cancer.

There are two type of Alternative hypothesis – one tail and two tail. When comparison occurs between two groups or researcher claim that Drug A is better than Drug B to increase the weight (i.e $(\bar{x}_A = \bar{x}_B)$ it is known as one tail hypothesis whereas in two tail hypothesis we only check whether the drug A and B are individually effective or not to increases the weight.

The sample data must provide sufficient evidence to reject the null hypothesis and conclude that the effect exists in the population. Ideally, a hypothesis test fails to reject the null hypothesis when the effect is not present in the population, and it rejects the null hypothesis when the effect exists.

Statisticians define two types of errors in hypothesis testing. Creatively, they call these errors Type I and Type II errors. Both types of error relate to incorrect conclusions about the null hypothesis.

After a study is completed, the investigator uses statistical tests to try to reject the null hypothesis in support of its alternative. Depending on whether the null hypothesis is true or false in the target population, and assuming that the study is free of bias, four situations are possible, as shown in below. In two of these, the findings in the sample and reality in the population are concordant, and the investigator's inference will be correct. In the other two situations, either a type I (α) or a type II (β) error has been made, and the inference will be incorrect.[1]

The researcher establishes the maximum chance of making type I and type II errors in advance of the study. The probability of committing a type I error (rejecting the null hypothesis when it is actually true) is called α (alpha) the other name for this is the level of statistical significance.

| | | Reality | |
|---|---|---|---|
| | | Positive | Negative |
| Study Finding | Positive | **True Positive** (Power) (1-β) | False Positive **Type I Error** (α) |
| | Negative | False Negative **Type II Error** (β) | **True Negative** |

**Figure 1: Type 1 and type 2 error**

If a study of lung cancer and smoking is designed taking Type I error (α = 5%), i.e A researcher has already set there are 5% chance of incorrectly reject the null hypothesis.

The probability of failing to reject the null hypothesis when it is actually false is called Type –II Error (β).

Power (1-β) is chance of not making type-II error (β).

If β is set at 0.10 0r 10%, then the researcher has decided that he/she is willing to accept a 10% chance of missing an association of a given effect size between smoking and lung cancer. This represents a power of 0.90, i.e., a 90% chance of finding an association of smoking in lung cancer.

Many studies set alpha at 0.05 and beta at 0.20 (a power of 0.80). These are somewhat arbitrary values, and others are sometimes used; the conventional range for alpha is between 0.01 and 0.10; and for beta, between 0.05 and 0.20. In general, the investigator should choose a low value of alpha when the research question makes it particularly important to avoid a type I (false-positive) error, and he should choose a low value of beta when it is especially important to avoid a type II error.[1]

Analysis plan includes decision rules for rejecting the null hypothesis. In practice, statisticians describe these decision rules in two ways - with reference to a P-value or with reference to a region of acceptance.

### P- Value

The concept of the P value is used in all over the world in hypothesis testing. Technically, P denotes the probability of obtaining a result equal to or more extreme than what is actually observed, assuming that the null hypothesis is true. The boundary for more extreme is dependent on the way the hypothesis is tested. Before the test is performed, a threshold value is chosen called the significance level of the test (also denoted by α) and this is conventionally taken as 5% or 0.05. If the P value obtained from the hypothesis test is less than the chosen threshold significance level, it is taken that the observed result is inconsistent with the null hypothesis, and so the null hypothesis must be rejected. This ensures that the Type I error rate is at the most α. typically the interpretation is:

A small P value (<0.05) - It is rejected the null hypothesis, which indicates strong evidence against the null hypothesis, so it is rejected. The alternative hypothesis may be accepted although it is not 100% certain that it is true. The result is said to be statistically "significant".

An even smaller P value (<0.01) indicates even stronger evidence against the null hypothesis. The result may be considered statistically "highly significant".

A large P value (>0.05) indicates weak evidence against the null hypothesis. Therefore, it cannot be rejected, and the alternate hypothesis cannot be accepted.

A P value close to the cut off (≈0.05 after rounding off) is considered to be marginal. It is better to err on the side of caution in such a case and not reject the null hypothesis.

## CONFIDENCE INTERVAL

A p-value is calculated to assess whether trial results are likely to have occurred simply through chance (assuming that there is no real difference between new treatment and old, and assuming, of course, that the study was well conducted), whereas Confidence intervals are preferable to p-values, as they tell us the range of possible effect sizes compatible with the data.

Confidence intervals are frequently calculated around the estimates from statistical hypothesis tests. They may be calculated for the t test, chi-square test, analysis of variance, regression, and most other tests of inference. A 95% CI is a range of values within which 95% of the results of repeated samples from the overall population would lie; this is the most frequently reported CI level. The confidence limits are related to the P value. If one calculates the 95% CI of the difference in means between two samples, and zero is within the range of the 95% CI, then the P value will not be significant at the level less than 0.05.

In epidemiology, a common way to indicate a measurement's precision is by providing a confidence interval. A narrow confidence interval indicates high precision; a wide confidence interval indicates low precision.

Confidence intervals are calculated for some but not all epidemiologic measures. It can also be calculated for some of the epidemiologic measures covered, such as a proportion, risk ratio, and odds ratio.

## NORMALITY OF DATA

For the continuous data, test of the normality is an important step for data analysis. When our data follow normal distribution, parametric tests otherwise nonparametric methods are used to compare the groups.

There are various methods to check the normality of the continuous data, out of them; most popular methods are Shapiro–Wilk test, Kolmogorov–Smirnov test, skewness, kurtosis, histogram, box plot, P–P Plot and Q–Q Plot. The two well-known tests of normality, namely, the Kolmogorov–Smirnov test and the Shapiro–Wilk test are most widely used methods to test the normality of the data. Normality of data can easily check using SPSS software and any other statistical software like Minitab, R, SAS, SYSTAT, and STATA.

Several statistical methods used for data analysis make assumptions about normality, including correlation, regression, t-tests, and analysis of variance. Central limit theorem states that when sample size has 100 or more observations, violation of the normality is not a major issue.[2,4,5,6]

If a continuous data follows normal distribution, then we present this data in mean value. Further, this mean value is used to compare between/among the groups to calculate the significance level (P value). If our data are not normally distributed, resultant mean is not a representative value of our data. A wrong selection of the representative value of a data set and further calculated significance level using this representative value might give wrong interpretation.[3,4,7] That is why, first we test the normality of the data, then we decide whether mean is applicable as representative value of the data or not. If applicable, then means are compared using parametric test otherwise medians are used to compare the groups, using nonparametric methods.

### Methods used for check the normality of data

Histogram, Q-Q plot, P-P Plot and box plot is graphical methods to check the normality of data.

Histogram: When a histogram's shape approximates a bell-curve it suggests that the data may have come for a normal population. (It can be generated using SPSS software https://www.spss-tutorials.com/ creating-histograms-in-spss/)

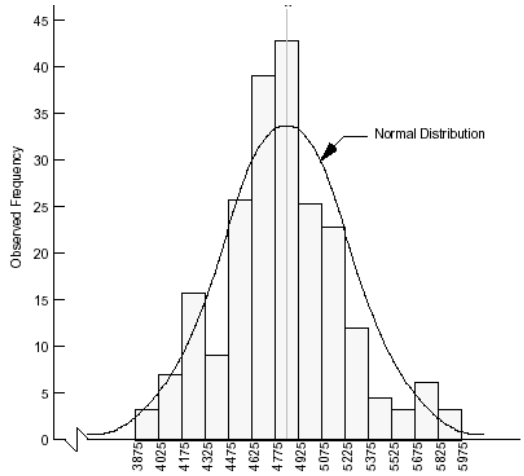


**Figure 2: Histogram with frequency polygon curve**

The Q-Q plot, or quantile-quantile plot, is a graphical tool to help us assess if a set of data plausibly came from some theoretical distribution such as a Normal or exponential. If we run a statistical analysis that assumes our dependent variable is normally distributed, we can use a Normal Q-Q plot to check that assumption. https://ezspss.com/test-for-normality-in-spss/

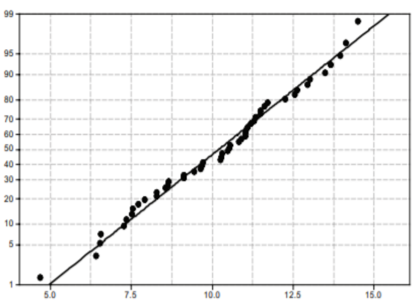

**Figure 3: P-P plot graph**

In descriptive statistics, a box plot or also known as box and whisker plot is a type of chart often used in descriptive data analysis. Box plots visually show the distribution of numerical data and skewness through displaying the data quartiles (or percentiles) and averages. Box plots show the five-number summary of a set of data: including the minimum score, first (lower) quartile, median, third (upper) quartile, and maximum score.

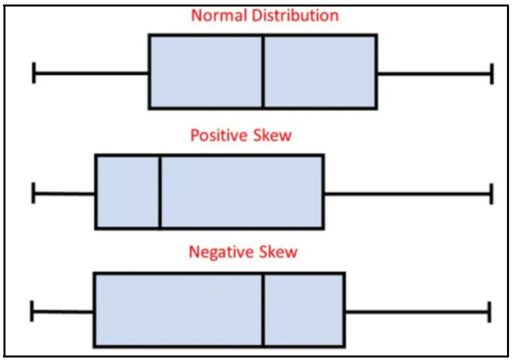


**Figure 4: Box plot graph**

## Statistical test for checking normality of data

The main tests for the assessment of normality are Kolmogorov-Smirnov (K-S) test, Lilliefors corrected K-S test, Shapiro-Wilk test, Anderson-Darling test, Cramer-von Mises test, D'Agostino skewness test, Anscombe-Glynn kurtosis test, D'Agostino-Pearson omnibus test, and the Jarque-Bera test. Among these, K-S is a much-used test and the K-S and Shapiro-Wilk tests can be conducted in the SPSS software and in others software also.[8,9,10,11]

Hypothesis for test of normality
H0: The sample data are not significantly different than a normal population.
H1: The sample data are significantly different than a normal population.
When testing for normality interpretation of P-Value:
Probabilities > 0.05 indicate that the data are normal.
Probabilities < 0.05 indicate that the data are not normal.

**Normally Distributed Data**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| Asthma Cases | .069 | 72 | .200* | .988 | 72 | .721 |

In SPSS output above the probabilities (P- values) are greater than 0.05 (the typical alpha level), so we accept H0… these data are not different from normal, which indicates data follow the normality.

**Non-Normally Distributed Data**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| Average PM10 | .142 | 72 | .001 | .841 | 72 | .000 |

In the SPSS output above the probabilities are less than 0.05 (the typical alpha level), so we reject H0… these data are significantly different from normal, which indicates data don't follow the normality.

**Figure 5: Kolmogrov-Smimov and Sahpiro-Wilk test output in SPSS and its interpretation**

## REFERENCES


1. Hypothesis testing, type I and type II errors, Amitav Banerjee, U. B. Chitnis, S. L. Jadhav, J. S. Bhawalkar, and S. Chaudhury1 ;Ind Psychiatry J. 2009 Jul-Dec; 18(2): 127–131.doi: 10.4103/ 0972-6748.62274; https://www.ncbi.nlm.nih.gov/ pmc/arti-cles/PMC2996198/

2. Principles of Epidemiology in Public Health Practice, Third Edition: An Introduction to Applied Epidemiology and Biostatistics Section 7: Measures of Spread. Available on https:// www.cdc.gov/csels/dsepd/ss1978/lesson2/section7.html. Accessed on June 18, 2021

3. Testing for Normality using SPSS Statistics. Available on https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php. Accessed on June 18, 2021

4. Prabhaker Mishra, Chandra M Pandey, Uttam Singh, Anshul Gupta,1 Chinmoy Sahu,2 and Amit Keshri3 Descriptive Statistics and Normality Tests for Statistical Data. Ann Card Anaesth, Jan-Mar2019; 22(1): 67–72. (https://www.ncbi.nlm. nih.gov/ pmc/articles/PMC6350423/)

5. Altman DG, Bland JM. Statistics notes: The normal distribution. BMJ. 1995;310:298. [PMC free article] [PubMed] [Google Scholar]

6. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. Int J Endocrinol Metab. 2012;10:486–9. [PMC free article] [PubMed] [Google Scholar]

7. Indrayan A, Satyanarayana L. Essentials of biostatistics. Indian Pediatr. 1999;36:1127–34. [PubMed] [Google Scholar]

8. Oztuna D, Elhan AH, Tuccar E. Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. Turkish Journal of Medical Sciences. 2006;36(3):171–6. [Google Scholar]

9. Elliott AC, Woodward WA. Statistical analysis quick reference guidebook with SPSS examples. 1st ed. London: Sage Publications; 2007. [Google Scholar]

10. Altman DG, Bland JM. Detecting skewness from summary information. Bmj. 1996;313(7066):1200. [PMC free article] [PubMed] [Google Scholar]

11. Peat J, Barton B. Medical Statistics: A guide to data analysis and critical appraisal. Blackwell Publishing; 2005. [Google Scholar]