Medical Statistics Series: Inferential Statistics (Part-II)

Swati Patel[1]

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

ABSTRACT

The purpose of this article is to provide detailed information of application of different inferential statistical methods and interpretation of its outcome values. The following are the general steps for statistical analysis: formulate a hypothesis, select an appropriate statistical test according to the objective of study, prepare data for analysis, summarize with appropriate measure of descriptive statistics, check of tests, run the analysis, tabulate the results and their interpretation.

Key words: Parametric test, one sample t-test, Independent t-test, Paired t-test, one way ANOVA, Z-test of proportion

INTRODUCTION

Inferential statistics takes data from a sample and makes inferences about the larger population from which the sample was drawn. The goal of inferential statistics is to draw conclusions from a sample and extrapolate them to a population. There are various methods in inferential statistics which could be applicable according to the objective of an experiment and type of data. To know whether a treatment /drug/procedure is actually effective or its effectiveness is by chance, the researcher needs to go for hypothesis testing which is the important part of inferential statistics.

Statistical test can be applied in study when researcher has defined hypothesis of study. In some study there is no hypothesis; the researcher just want to know or determine various facts. For example, in a prevalence study there is no hypothesis to test, and the size of the study is determined by how accurately the investigator wants to determine the prevalence. If there is no hypothesis, then there is no statistical test. It is important to decide *a priori* which hypotheses are confirmatory (that is, are testing some presupposed relationship), and which are exploratory (are suggested by the data).

Selection of a particular statistical test in a study depends on the objective of the study. Depending upon the objective of study a hypothesis can be decided .If a study includes measure of summarization variables such as mean, the hypothesis will have comparison of means of two or more than two groups. Likewise other measures of summarisation such as proportion, and median etc can be compared.

The main purpose of this article is to understand the concept and application of different test statistics according to the objective of study.

Parametric test

To analyse the continuous outcome variables, parametric test will be applied when normality assumptions are satisfied otherwise equivalent non parametric test will be applied. Normality, Homogeneity of variance, linearity and independence are the assumptions of the parametric test that should be satisfied before applying the test.

One sample t-test

In a One Sample *t* Test, the test variable's mean is compared against a "test value", which is a known or hypothesized value of the mean in the population. Test values may come from a literature review, a trusted research organization, legal requirements, or industry standards.

According to CDC, the mean height of U.S. adults ages 20 and above is about 66.5 inches .In our sample data; we have a sample of 435 college students from a single college. Let's test if the mean height of students at this college is significantly different than 66.5 inches.

The null and alternative hypotheses of this test will be:

H_0:μ_{Height} =66.5("the mean height is equal to 66.5")
H_1: μ_{Height} ≠ 66.5 ("the mean height is not equal to 66.5")

Before applying the test

In the given sample data, the variable is Height, which a continuous variable representing each respondent's height in inches. The heights exhibit a range of values from 55.00 to 88.41. Normality is checked using Shapiro –Wilk`s test (P>0.05), which indicates data (height) following the normality.

	t_{cal}	Mean difference	P- value
Height	5.810	1.53	P<0.001

Interpretation:
One sample t-test can be applied on data manually as well as using the statistical software's. If researchers go for manually calculation he/she first go for calculation of test using the formula given in the table. 1 ($\bar{x}$ = study sample mean, μ_0= pre decided hypothetical value, SD= standard deviation of sample, n = sample size) .After calculating it, need to compare the tabulate value according to the value of degree of freedom at specific level of significance (95 % or 99%).Tabulate value will be get for statistical bles.($t_{cal} > t_{tab}$ at 95 % or 99% ,reject the null hypothesis).

When one sample –test is done by any software (SPSS, SYSTAT, SAS, Graph pad, Open Epi, etc) it gives P-value (sig.2 tailed) to decide null or alternative hypothesis will be reject or accept. In the table .1 one sample t-test outcome values are given, which indicates P- value <0.001 i.e. we reject the null hypothesis. There is a significant difference in the mean height of the students at this college and the overall adult population in the U.S. (p <0 .001).The average height of students at this college is about 1.5 inches taller than the U.S. adult population average (95% CI [1.013, 2.050]).

Independent t-test or unpaired t-test

When two quantitative variables from same population (continuous data- interval or ratio) are given, which follow the normality and summarised as mean .Also follow the homogeneity of variance (i.e. variance should be approximately equal across the groups).When this two assumptions following the given both group data independent t – test applied to know the significance difference between two groups mean.

When two quantitative variables (continuous data-interval or ratio) are given, which follow the normality summarised as mean? But don`t follow the homogeneity of variance (i.e. variance unequal across the groups).When the assumption of homogeneity of variance violated and sample size of each group differ ,Welch t-test applied for to know the difference between the mean of two independent test. No outliers should be present in the data. Shapiro-Wilk`s and Levence test use to check the normality and homogeneity of variance respectively.

Hypothesis for independent t-test and Welch t-test

H_0:μ_1 =μ_2 ("the two sample means are equal")
H_1: μ_1≠ μ_2 ("the two sample means are not equal")

Example:

A researcher decided to investigate whether an exercise or weight loss intervention is more effective in lowering cholesterol levels. The researcher recruited a random sample of inactive males that were classified as overweight. This sample was then randomly split into two groups: Group 1 underwent a calorie-controlled diet and Group 2 undertook the exercise-training programme. In order to determine which treatment programme was more effective, the mean cholesterol concentrations were compared between the two groups at the end of the treatment programmes

Group	Mean	SD	t_{cal} (P-value)
Group1(n=20)	6.145	0.5196	2.428 (0.02)
Group 2 (n=20)	5.795	0.3818	

Interpretation:

The value of cholesterol is normality distributed in group1 and group 2, which follow the normality (P >0.05). Leven`s test value showing that the variances of both of this groups are approximately equal (P-value = 0.314), so mean of cholesterol for this two group can be compared by Independent t-test. (Shapiro- wilk`s Tets for normality & Leven`s test for Homogeneity can be calculated using SPSS) The mean cholesterol level of Group 1 and Group 2 are not same, which is statistically significant (P-Value is 0.02). In other words we can say that the mean cholesterol level is higher in Group 1 as compared to Group.

Paired t-test

The paired *t*-test (for matched samples) are probably the most widely used methods in statistics for the comparison of differences between two dependent samples and or matched samples. To apply the paired *t*-test to test for differences between paired measurements, the following assumptions need to hold:

• Subjects must be independent. Measurements for one subject do not affect measurements for any other subject.

- Each of the paired measurements must be obtained from the same subject. For example, the before-and-after weight for a smoker in the example above must be from the same person.
- The measured differences are normally distributed.
- The data are continuous (not discrete).
- We also have an idea, or hypothesis, that the difference between pairs is zero. Here are three examples:
- A group of people with dry skin use a medicated lotion on one arm and a non-medicated lotion on their other arm. After a week, a doctor measures the redness on each arm. We want to know if the medicated lotion is better than the non-medicated lotion. We do this by finding out if the arm with medicated lotion has less redness than the other arm. Since we have pairs of measurements for each person, we find the differences. Then we test if the mean difference is zero or not.
- We measure weights of people in a program to quit smoking. For each person, we have the weight at the start and end of the program. We want to know if the mean weight change for people in the program is zero or not.
- An instructor gives students an exam and the next day gives students a different exam on the same material. The instructor wants to know if the two exams are equally difficult. We calculate the difference in exam scores for each student. We test if the mean difference is zero or not.

Example (Before and after comparison)

The study was conducted to assess the effectiveness of iron supplementation in anaemic patients. Researcher included total 14 patients according the inclusion criteria of study, h/s recorded baseline (before) and after treatment value of Hb to calculate difference.

Measure	Hb gm%
Mean gain in Hb	1.86
SD of difference	1.1
n	14
df	13-1
P- value	P<0.001

The above table reveals that the iron supplementation has significantly improved the Hb level (P<0.001)

Example (Matched Comparison)

In addition to the time related samples, paired *t*-test is also introduced in the data analysis of matched sampling. Such sampling is a method of data collection and organization which helps to reduce bias and increase precision in observational studies.[2] For example, consider a clinical investigation to assess the repetitive behaviours of children affected with au-

tism. A total of 10 children with autism enrolled in the study. Then, 10 controls are selected from healthy children with matched age and gender which may be the confounding factors in the study. Each child is observed by the study psychologist for a period of 3 hours. Repetitive behaviour is scored on a scale of 0 to 100 and scores represent the percent of the observation time in which the child is engaged in repetitive behaviour (see table).[1]

Children with autism	68.8±8.95
Children without autism	51.5±10.89
Difference	17.3±6.07
P- value	P<0.0001

The above table revels that in this example, there are 20 subjects. However, each subject in the experimental group is matched with a subject in the control group. We also need to use the matched pair *t*-test to compare the mean values of the two groups. The paired *t*-test statistic equals 9.00. From the *t*-distribution with *df*= 9, we obtain the p-value <0.0001, which shows strong evidence to reject the null hypothesis.

One sample Z-test of proportion:-

The One-Sample Z- test of Proportion is used to assess whether a population proportion (P) is significantly different from a hypothesized value (p)/ sample proportion. Examples are, comparing the proportion of subjects with various forms of infections or malnutrition or health disorder with those of the population.

A researcher planed study to compare the prevalence of whooping cough in rural school age children sample proportion (prevalence)12%, he/she took the prevalence rate(population proportion) of whooping cough was 11% from data of health survey by government of Gujarat to compare the sample proportion. Number of subjects studied was 100.

Null Hypothesis: The sample proportion and population proportion of whooping cough is same (p=P)

Alternative Hypothesis: The sample proportion and population proportion of whooping cough is not same (p≠P)

| Study Variable	Prevalence	Z Cal	P- value
Sample Prevalence of Whooping cough(p)	12%	0.3	0.7583
Population Prevalence of whooping cough(P)	11%		

The above table depicts that no significance difference observed in between the sample and population prevalence of whooping cough (P-value is 0.7583). Thus, we are 95% confident that the true proportion of whooping cough in rural school age children is between 32.9% and 36.1%.

Two sample Z-test of Proportion

The two sample proportion test is a hypothesis test to compare the proportions of one certain event oc-

curring in two populations following the binomial distribution.

- Null Hypothesis(H_0): $P_1 = P_2$
- Alternative Hypothesis(H_a): $P_1 \neq P_2$
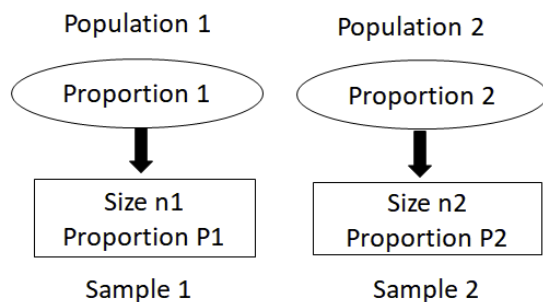
Two Sample Proportion Test Assumptions

- The sample data drawn from the populations of interest are unbiased and representative.
- There are only two possible outcomes in each trial for both populations: success/failure, yes/no, and defective/non-defective etc.
- The underlying distributions of both populations are binomial distribution.
- When $np \geq 5$ and $np (1 - p) \geq 5$, the binomial distribution can be approximated by the normal distribution.

The purpose of two sample Z test is to compare the random samples of two populations. Use two sample z test of proportion for large sample.

A researcher plan population survey to study nutritional deficiencies and want to compare two villages. Is there any significance of difference in the prevalence/proportion of nutrition deficiency between the village A and village B?

Details	Village A	Village B
Sample size	150	160
Subject with clinical deficiency	129	110
Prevalence/proportion with clinical deficiency	86%	68.75%
Z cal	3.61	
P-Value	P<0.05	

The above table illustrated that the prevalence/ proportion of nutrition deficiency between village A and village B is not same, which is statistically significant (P- value <0.05). For Z test of one sample and two sample proportion calculated value and P- value can be calculated using Epi tools software. (https://epitools.ausvet.com.au/ztestone & https://epitools.ausvet.com.au/ztesttwo)

The Differences and Similarities between Two-Sample T-Test and Paired T-Test

One way Analysis of Variance (ANOVA)

The differences in the means of more than two groups that are mutually independent and satisfy both the normality and equal variance assumptions can be obtained by comparing them using Analysis of Variance (ANOVA). However, the present article aims to examine the necessity of using a one-way ANOVA instead of simply repeating the comparisons using Student's t-test. The present article aims to use a conceptual illustration to explain how the difference in means can be explained by comparing the variances rather by the means themselves.

Null Hypothesis: $\mu_1 = \mu_2 = \cdots = \mu_n$
Alternative hypothesis: $\mu_1 \neq \mu_2 \neq \cdots \neq \mu_n$

Example

A researcher planed randomised control trial carried out to compare the effect of three treatments A, B and C for the reduction of serum cholesterol levels in obese patients. The main objective of this study is to find out the difference in reduction (Baseline serum cholesterol value-After intervention value) of serum cholesterol level after intervention by A, B and c. Outcome of analysis is given in the below table.

Null Hypothesis: Mean reduction of serum cholesterol level after intervention by treatment A, B and C is same.

Alternative Hypothesis: Mean reduction of serum cholesterol level after intervention by treatment A, B and C is different (not same).

Before applying the ANOVA the assumptions of it should be checked, which are normality and homogeneity of variables. The Normality and Homogeneity is checked using shapiro-Wilk`s test and Leven`s test respectively, which full field the assumptions to applied the ANOVA.

Treatment	n	Mean ±SD	P-Value
Treatment A	20	21.62±3.21	P<0.05
Treatment B	20	32.76±2.87	
Treatment C	20	23.75±1.08	

The above table revels that mean reduction of serum cholesterol level after intervention by treatment A, B and C is different, which is statistically significant (P<0.05). Moreover it shows that treatment B is more effective as compared to treatment A and treatment B. Statistical value of ANOVA can be calculate using the statistical software (SPSS, SYSTAT, SAS, Graph pad, Open epi etc).

REFERENCES

1. Xu M, Fralick D, Zheng JZ, Wang B, Tu XM, Feng C. The Differences and Similarities between Two-Sample T-Test and Paired T-Test. Shanghai Arch Psychiatry. 2017 Jun 25;29(3):184-188. doi: 10.11919/j.issn.1002-0829.217070. PMID: 28904516; PMCID: PMC5579465.

Annexure: Summary of applications of statistical test

Details	Name of test	Formula	Software
For comparison of One sample mean with theoretical value/	One sample t-test	$t_{cal} = \dfrac{\lvert \bar{x} - \mu_0 \rvert}{\frac{sd}{\sqrt{n}}}$	Epi tools SPSS STATA
For comparison of two dependent samples like Pre treatment out is compare with post treatment.	Paired t test	$t_{cal} = \dfrac{\lvert \bar{d} \rvert}{\frac{sd(\bar{d})}{\sqrt{n}}}$	SPSS EXCEL Epi tools
For comparison of two Independent samples like mean Hb of male and female of Malaria cases.	Independent t-test or Unpaired t test	$t_{cal} = \dfrac{\lvert \bar{x_1} - \bar{x_2} \rvert}{\sqrt{\dfrac{SD_1^2}{n_1} + \dfrac{SD_2^2}{n_2}}}$	Epi tools Open Epi Excel SPSS
One sample (qualitative data) For comparison population proportion (study population) with hypothesized value like proportion of anaemia in pregnant women in Surat city is compared with the prevalence /proportion of anaemia in pregnant women reported by NFHS in Gujarat.	One sample Z-test of proportion	$Z_{cal} = \dfrac{\lvert P - p \rvert}{\sqrt{\dfrac{PQ}{n}}}$	Epi tools
For comparison of two Independent samples like proportion of obesity among Indian and American adult.	Two sample Z-test of proportion	$Z_{cal} = \dfrac{\lvert\ P_1 - P_2\ \rvert}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}}$	Epi tools
Comparison of more than two independent group mean	ANOVA	---	Open Epi