



Formation, Testing of Hypothesis and Confidence Interval in Medical Research

Senthilvel Vasudevan¹ 

Assistant Professor of Statistics (Biostatistics), Department of Pharmacy Practice, College of Pharmacy, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia.

Abstract

Background: Statistics which help us in arriving at the criterion for such decisions in any research. Hypothesis means assumptions. It is an important activity of pharmacy or medical fields and its related research.


Materials and Methods: Statistical inferences play an important role in biological statistical tests and arriving at some conclusion. Some suitable examples were also worked out in this section.

Results: Confidence intervals provide a method of stating the precision or closeness of the sample statistics. It contains lower and upper limits.

Conclusion: We concluded that the hypothesis is very much useful and essential tool in medical, nursing, pharmacy and other science and biomedical sciences as well as in its research fields. Some numerical illustrations with suitable examples also there.

Key Words: hypothesis testing, type I error, type II error, confidence interval, medical research

Article Summary: Submitted: 20-July-2022 Revised: 10-August-2022 Accepted: 05-September-2022 Published: 30-September-2022

Quick Response Code:	Web Site http://ijmsnr.com/	This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non-Commercial-ShareAlike 4.0 International License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.
	DOI 10.55349/ijmsnr.2022232227	
		Corresponding Author: Dr. Senthilvel Vasudevan, Assistant Professor of Statistics, Department of Pharmacy Practice, College of Pharmacy, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, KSA. Email ID: senthilvel99@gmail.com

Introduction

The theory of testing of hypothesis was initiated by J. Neyman and E.S. Pearson and employs statistical techniques to arrive a decision in certain situations where there is an element of uncertainty based on a sample size is fixed in advance. Hypothesis means assumption. Its testing is an important activity of pharmacy/medical fields and its related research. Good formation of assumption is fifty percentage answer for the study theme/question [1]. knowledge of the subject and working knowledge of the concepts are very important. Confidence intervals [2] provide different information from that arising from hypothesis tests. Hypothesis testing produces a decision about any observed difference: either that the difference is 'statistically significant' or 'statistically non-significant'. The present paper discusses the methods of hypothesis formation, statistical concepts of hypothesis testing, confidence interval in Pharmacy.

Testing of Hypothesis:

The test of hypothesis discloses the fact whether the difference between the computed statistic and hypothetical parameter is significant or otherwise. Hence, the test of hypothesis is also known as the test of significance. It is concerned with the formation of a hypothesis based on estimation from sample data and then testing whether the hypothesis laid down is true or not [3, 4].

How to cite this article: Vasudevan S. Formation, Testing of Hypothesis and Confidence Interval in Medical Research. Int J Sci and Med Res 2022;2(3):22-27

The main structure of hypothesis testing:

All quantitative research has some issues or problems that are trying to investigate and the focus in hypothesis testing is to find ways to structure these in such a way that we can test them effectively. We follow the following steps:

- a) **Define the research hypothesis:** A statistical hypothesis or simply a hypothesis of a hypothesis is a tentative conclusion logically drawn concerning any parameter of the population. Example:
 1. A given medicine cures 97% of the patients taking it.
 2. A hormone thyroxine (T₄) increases the respiratory metabolism in 98% of the cases.
 3. In childbirth, there is an equal chance of male and female birth.
 4. The average consumption of food of the two populations of rabbits is equal.
- b) **Formation of the null and alternative hypothesis:** In these two alternatives as follows:
 1. The hypothesis is correct and accepted because the observed value of an attribute of a sample does not show much deviation from the expected value of that attribute of the population.
 2. The hypothesis is not accepted or is rejected because the observed value of the sample distinctly varies from the expected value.

Based on this, two types of hypotheses are there:

- a. Null Hypothesis (H₀) b. Alternative Hypothesis (H₁)
- a. **Null Hypothesis:** A statistical hypothesis which is to be tested for the purpose of possible acceptance is known as null hypothesis. It is denoted by H₀. According to Prof. R. A. Fisher, null hypothesis is the hypothesis which is test for possible rejection under the assumption that it is true.

Example: Suppose the average life of man is 70 years. Then, the H₀ is set as, $\mu = 70$.

While we are setting up a null hypothesis (H₀), we should taken into a consideration of the following:

(i). If we want to test the significance of the difference between a statistic and the parameter or between a statistic and the parameter or between two sample statistics, then we set up a null hypothesis that difference is not significant. This means that the difference is just on account of fluctuations of sampling.

$$H_0: \mu = \bar{x}$$

(ii). If we want to test any statement about the population, we set up the null hypothesis that it is true. For instance, if we want to know if the population mean has specified value μ_0 , then we set up H₀ as follows:

$$H_0: \mu = \mu_0$$

- b. **Alternative Hypothesis:** Any hypothesis which is complementary to the null hypothesis is called an alternative

For example: If we want to test the null hypothesis (H₀) that the average birth weight of a newborn baby in a OBG ward of 2500 kg.

- c. **Alternative Hypothesis:** Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis. This is denoted by H₁ or H_A.

For example: If we want to test the null hypothesis (H₀) that the average birth weight of a new born baby in a OBG ward of 2500 kg.

$$H_0: \mu = 2500 \text{ kg.} = \mu_0$$

Then the alternative hypothesis will be as follows:

$$H_1: \mu \neq 2500 \text{ kg.}$$

$$H_1: \mu < 2500 \text{ kg.}$$

$$H_1: \mu > 2500 \text{ kg.}$$

Two types of errors in testing of hypothesis [1]: At the time of writing a conclusion for any study, to take a conclusion of accepting or rejecting null hypothesis (H₀). Normally, some error will be happened in any kind of study or research. There were two possible types of errors in the method of hypothesis testing. They are as follows: (a). Type I error and it is denoted by the symbol ‘ α ’ and (b). Type II error and it is denoted by ‘ β ’.

Type I error: It is true when the null hypothesis is rejecting and denoted by the symbol

Type II error: It is false when the null hypothesis is accepting and denoted by the symbol

The following table shows the four possible types of situations

Actual	Decision	
	Accept H ₀	Reject H ₁
H ₀ is true	Correct decision (No error) Probability = 1 - α	Wrong (Type I error) Probability = α
H ₀ is false	Wrong (Type II error) Probability = β	Correct decision (No error) Probability = 1 - β

While accepting or rejecting a null hypothesis, our main aim is to reduce the probability of making a type I error. The probability of making type I error is denoted by Greek word α (alpha). Therefore, the probability of making a correct decision is (1 - α).

- c) Explain how you are going to operationalize (that is, measure or operationally define) what you are studying and set out the variables to be studied.

- d) Set the level of significance (α). The statistical tests fix the probability of committing type I error (α) at a certain level called the level of significance (LOS).

Always, levels of significance (α) have to be fixed as 5% (i.e., $5/100 = 0.05$) and 1% (i.e., $1/100 = 0.01$). If research chooses $\alpha = 5\%$, it means that the experiment/event is 5 times error/false out of 100 experiments or events then a researcher has to reject a correct H_0 . In other way, 95% confidence that decision to reject H_0 is correct. α desired is to be fixed in prior to apply the statistical test to any kind of study.

- e) **Rejection region and make a one or two-tailed test's prediction:**

Rejection region: Whole area of a standard normal curve (SNC) is 1 and it is representing probability distribution. Testing of hypothesis, LOS is set up in order to know the probability of making a 'α' of rejecting the hypothesis which is true. Really, region of the SNC corresponding to a pre-determined LOS should be known, because when the test statistic computed to test the hypothesis falls in the region, it is advisable to reject the hypothesis as it is believed to be corresponding to a pre-determined 'α' that is fixed for knowing the chance of making the type I error of rejecting the hypothesis which is true, is known as the "rejection region (RR)" of "critical region (CR)". Region of SNC is not covered by the RR, is called "accepted region (AR)". The testing of hypothesis falls in the acceptance region, then it is to be hypothesis accepted as it is and taken as 'true value'.

In one tailed test: CR may be shown by a researcher of the area under the normal curve in following ways: 'two-tails' under the curve which is either the right tail or the left tail. Both the tails under the normal curve is called two-tailed test/two-sided test. If CR is represented by only one tail then the test is one tailed or one-sided test.

In two tailed test: Two tailed test is used in cases where it is considered either a +ve or -ve difference between sample and population mean is towards rejection of null hypothesis. Otherwise, when the sample and population mean is significantly different from it is not considered to be due to chance, the two-tail test is to be used and viz versa. In former case, the right tail test and in the latter case the left tail test is to be applied.

- f) Determine whether the distribution that you are studying is normal. Then only you must decide the suitable type of statistical tests for your collected data.
- g) Select an appropriate statistical test based on the variables you have defined and whether the distribution is normal or not.
- h) Run the statistical tests on your data and to give an interpretation about the output for the study.
- i). The probability of obtaining a sample mean, given that the value stated in the null hypothesis is true, is stated by the p value. The p value is a probability: It varies between 0 and 1 and can never be negative. The criterion or probability of

obtaining a sample mean at which point we will decide to reject the value stated in the null hypothesis, which is typically set at 5% in behavioral research. To decide, we compare the p value to the criterion.

A p value is the probability of obtaining a sample outcome, given that the values stated in the null hypothesis is true. The p value for obtaining a sample outcome is compared to the level of significance.

Significance, or statistical significance, describes a decision made concerning a value stated in the null hypothesis. When the null hypothesis is rejected, we reach significance. When the null hypothesis is retained, we fail to reach significance.

When the p value is less than 5% ($p < 0.05$), we reject the null hypothesis. We will refer to $p < 0.05$ as the criterion for deciding to reject the null hypothesis, although note that when $p = 0.05$, the decision is also to reject the null hypothesis. When the p value is greater than 5% ($p > 0.05$), we retain the null hypothesis. The decision to reject or retain the null hypothesis is called **significance**. When the p value is less than 0.05, we reach significance; the decision is to reject the null hypothesis. When the p – value is greater than 0.05, we fail to reach significance; the decision is to retain the null hypothesis.

- j). **Accept or reject the null hypothesis:** At last, a decision is taken as to whether the null hypothesis is to be accepted or rejected. If the calculated value of the test statistic is less than the table value, the computed value of the test statistic falls in the acceptance region and the null hypothesis is accepted. If, on the contrary the computed value of the test statistic falls in the rejection region and null hypothesis is rejected. Normally, 5% level of significance ($\alpha = 0.05$) is used in testing a hypothesis and taking a decision unless otherwise any other level of significance is specifically stated [3,4].

Example:

For the comparison of sample mean with population mean: In a School Health Survey of children in a school, the mean haemoglobin level of 55 boys was found to be 10.2 g per 100ml with a standard deviation of 2.1. Can it be considered that this group of boys is identified from a population with a mean of 11.0 g/100ml?

Given data,

Sample size (n) = 55, Sample mean (\bar{x}) = 10.2 g / 100 ml.
 Population mean (μ) = 11.0 g / 100 ml.
 Standard deviation of sample (s or σ) = 2.1 g / 100 ml

Null hypothesis (H_0): The population from which a sample of 55 boys is taken has mean haemoglobin level of 11.0 g/100 ml blood.

Alternative hypothesis (H_1): The population from which a sample of 55 boys is not taken has mean haemoglobin level of 11.0 g/100 ml blood.

Alternative hypothesis (H_1): Standard error of mean: The standard error of mean is calculated from the following formula:

$$\begin{aligned} \text{Standard error of mean} &= \sigma / \sqrt{n} \quad (\text{or}) \quad S / \sqrt{n} \\ &= 2.1 / \sqrt{55} \\ &= 0.283 \end{aligned}$$

Critical ratio:

$$\begin{aligned} \text{Critical ratio} &= \frac{\text{Difference in means}}{\text{Standard error of mean}} \\ &= \frac{\bar{x} - \mu}{S / \sqrt{n}} \\ &= \frac{\text{Sample mean} - \text{Population mean}}{\text{Standard error of mean}} \\ &= \frac{10.2 - 11.0}{0.283} = \frac{0.8}{0.283} = 2.83 \end{aligned}$$

Comparison with table value:

Theoretical value of critical ration from probability of normal distribution (at $\alpha = 0.01$ or at 1%) = 2.58. Observed value of critical ratio = 2.83. It means observed value of critical ratio 2.83 is greater than 2.58. This indicates that the probability of getting the value 2.83 or greater than 2.83 by chance is greater than 0.01.

$$\begin{aligned} \text{Probability (P)} &= \frac{\text{Theoretical value of critical ratio}}{\text{Observed value of critical ratio} \times \text{sample size}} \\ &= \frac{Z_{\alpha}}{Z_c \times n} \\ &= \frac{2.58}{2.83 \times 55} = 0.017 \end{aligned}$$

When probability is less than 0.01 the null hypothesis is rejected.

Inference / Interpretation: The sample is not taken from the population with mean 11.0 g / 100 ml blood.

Confidence Interval [5]

Confidence Intervals are interval estimates of a range of values with a specified high probability of containing the population parameter. In any Gaussian distribution, 95% of the values will fall within 1.96 (~2) standard deviation of the population mean. If a sample mean \bar{x} , lies on the horizontal axis under the shaded area, as show below, then the true population mean μ , is likely to be included in the interval $\mu \pm 1.96\sigma$. If, however our observed value lies in the tail of the curve, ie., outside the shaded area, the interval estimate is not likely to include the true population mean. This forms the basis for the estimation of Confidence Interval.

The general formula used for estimation of confidence intervals around the population mean μ , is

$$\text{C. I} = \text{Population mean} \pm \text{Confidence Coefficient} \times \text{Population Standard Deviation}$$

$$\text{C. I} = \mu \pm \text{C.C} \times \sigma$$

But usually, the population mean as well as population standard deviation are not known to us. Therefore, we have to fall back upon the sample mean (\bar{x}) and the standard error of the sample mean as estimates of the respective population parameters. The working formula for confidence interval estimation therefore becomes,

$$\text{C. I} = \bar{x} \pm \text{Confidence Coefficient} \times \text{SE of sample mean}$$

Standard Error of sample mean can also be denoted as s/\sqrt{n}

$$\text{Therefore, C. I} = \frac{\bar{x} \pm \text{C.C} \times s}{\sqrt{n}}$$

The confidence coefficient can be kept 1.96 or 2.576 or 3.29 depending on the level of confidence, α being 0.05 or 0.01 or 0.001 respectively. This implies that, we have 95%, 99% and 99.9% confidence respectively for $\alpha_{0.05}$, $\alpha_{0.01}$, $\alpha_{0.001}$ and the confidence coefficient used would be 1.96 (~2), 2.5 or 3.3 respectively.

(i). Quantitative Data:

To estimate the range of individual values.

When population standard deviation (σ) is known.

Two assumptions made are sample size is large and follows Gaussian distribution.

Example: What are the confidence limits for observed mean height 170.6 cm in a sample of 725 children, where a standard deviation of 6.25 cm was seen.

For 95% confidence interval, we will use the formula,

$$\text{C. I} = \bar{x} \pm \text{Confidence Coefficient} \times \text{SE of sample mean}$$

$$\text{S. E} = \text{S.D}/\sqrt{n}$$

$$= \frac{6.75}{\sqrt{725}} = \frac{6.75}{26.93} = 0.25$$

$$\begin{aligned} \text{The Confidence limits would be} &= 170.6 \pm 1.96 \times 0.25 \\ &= 170.6 \pm 0.49 \\ &= 169.60 \text{ to } 171.09 \text{ cms.} \end{aligned}$$

When population SD (σ) is unknown and sample size is small.

In some cases, the population (σ) is not known, and the sample standard deviation (s) is then used as an estimate of σ . Here there is additional source of variability, and we obtain the confidence coefficient by using student's 't' distribution. The t distribution is also a symmetrical bell-shaped curve, but flatter than the Gaussian distribution and with thicker tails. Here the thickness being determined by the number of degrees of freedom (df). When the df is very large, the t distribution is almost like the Gaussian distribution.

The formula used to calculate the CI, $CI = \bar{x} \pm t(s/\sqrt{n})$

Example:

15 patients randomly selected from a Male Medical Ward and assessing their serum urea levels, we find an average of 36.8 mg% with a standard deviation of 2.7 mg%. We want to find the 99% confidence interval estimate of the true mean serum urea levels for all patients in the ward.

Given, $\bar{x} = 36.8, s = 2.7, n = 15$

$$S.E = s / \sqrt{n}$$

$$\text{Standard Error} = \frac{2.7}{\sqrt{15-1}} = \frac{2.7}{\sqrt{14}} = \frac{2.7}{3.74} = 0.72$$

Degrees of freedom (df) for t – distribution is computed as $n - 1 = 10 - 1 = 9$.

The 99% Confidence Coefficient means a CI encompassing as area $(1 - \alpha) = 0.99$. Thus, the area outside the interval, $\alpha = 0.01$. For a one tailed procedure, the 99% confidence coefficient in terms of the t – distribution would be,

$t_{(1-0.01)} = t_{0.99} = 2.821$ (t – distribution table at $df = 9$). By substitution we get,

$$\begin{aligned} C.I_{\mu} &= \bar{x} \pm t(s/\sqrt{n}) \\ &= 36.8 \pm (2.821) \times 0.72 \\ &= 36.8 \pm 2.03 \\ &= 34.77 \text{ to } 38.83 \text{ mg\%} \end{aligned}$$

The interpretation is, we can be 99% certain that the population mean serum urea for the patients of medical ward lies in between 34.77 and 38.83 mg%.

(ii). Qualitative Data:

To estimate a Single Population Proportion:

We often have data representing the proportion of subjects possessing a particular characteristic.

Example: Study to access the risky sexual behavior among truck drivers in a town. If 15 of the randomly selected 50 truck drivers admitted of having recent unprotected sex with a commercial sex works. What is the 95% confidence interval estimate of the population of truck drivers in the town who have had unprotected risky sexual behavior in the recent past?

Let, P be the proportion of truck drivers in the whole town, who indulge in risky behavior.

Here, we have p (Sample Proportion) = $15/50 = 0.3$, which is an estimate of P.

We will use a modified formula, $C.I_p = p \pm C.C \times S.E_p$

Assuming Gaussian distribution in our observations, the 95% confidence coefficient,

$$Z_{(1-\alpha/2)} = Z_{(1-0.05/2)} = Z_{(1-0.025)} = Z_{0.975} = 1.96$$

Substituting this value in the above equation, we get,

$$\begin{aligned} C.I_p &= p \pm 1.96 \sqrt{\frac{p(1-p)}{n}} \\ &= 0.3 \pm 1.96 \sqrt{\frac{0.3(1-0.3)}{50}} \\ &= 0.17 \text{ to } 0.43 \end{aligned}$$

Inference: 95% confident that the unknown proportion of truckers in this town who have had unprotected/risky sexual behavior recently, lies between 17% to 43%.

Confidence Interval in Odds ratio:

The 95% confidence interval (CI) is used to estimate the precision of the OR. A large CI indicates a low level of precision of the OR, whereas a small CI indicates a higher precision of the OR. It is important to note however, that unlike the p value, the 95% CI does not report a measure’s statistical significance. In practice, the 95% CI is often used as a proxy for the presence of statistical significance if it does not overlap the null value (e.g. OR=1). Nevertheless, it would be inappropriate to interpret an OR with 95% CI that spans the null value as indicating evidence for lack of association between the exposure and outcome.

In the study, 186 of the 263 adolescents previously judged as having experienced a suicidal behavior requiring immediate psychiatric consultation did not exhibit suicidal behavior (non-suicidal, NS) at six months follow-up. Of this group, 86 young people had been assessed as having depression at baseline. Of the 77 young people with persistent suicidal behavior at follow-up (suicidal behavior, SB), 45 had been assessed as having depression at baseline. What is the OR of suicidal behavior at six months follow-up given presence of depression at baseline?

- a = Number of exposed cases (++) = 45
- b = Number of exposed non-cases (+-) = 86
- c = Number of unexposed cases (-+) = 32
- d = Number of unexposed non-cases (--)= 100

$$\begin{aligned} OR &= \frac{a/c}{b/d} = \frac{ad}{bc} \\ &= \frac{45/32}{86/100} \\ &= 1.63 \end{aligned}$$

B) Calculating 95% confidence intervals

What are the confidence intervals for the above calculated OR?

Confidence intervals are calculated using the formula shown below

$$\text{Upper Limit 95\% CI} = e^{\ln(\text{OR}) + 1.96 \sqrt{1/a + 1/b + 1/c + 1/d}}$$

$$\text{Lower Limit 95\% CI} = e^{\ln(\text{OR}) - 1.96 \sqrt{1/a + 1/b + 1/c + 1/d}}$$

Plugging in the numbers from the table above, we get:

$$\text{Upper 95\% CI} = e^{\ln(\text{OR}) + 1.96 (1/45 + 1/86 + 1/32 + 1/100)} = 2.80$$

$$\text{Lower 95\% CI} = e^{\ln(\text{OR}) - 1.96 (1/45 + 1/86 + 1/32 + 1/100)} = 0.96$$

Since, the 95% CI of 0.96 to 2.80 spans 1.0, the increased odds (OR 1.63) of persistent suicidal behavior among adolescents with depression at baseline does not reach statistical significance. In fact, this is indicated in Table 1 of the reference article, which shows a p value of 0.07. Interestingly, the odds of persistent suicidal behavior in this group given presence of borderline personality disorder at baseline was twice that of depression (OR = 3.8, 95% CI: 1.6 - 8.7), and was statistically significant (p = 0.002).

Author Contributions: SV – Conceived and designed the analysis; SV – Calculations, wrote the full paper; Wrote and checked the article.

Here, SV – Senthilvel Vasudevan

Conflict of Interest: The author not having any kind of conflict of interest in this study.

Source of funding: I didn't receive any kind of financial support from the parent institution or any other financial institutions or organizations.

References

1. Amitav Banerjee, U.B. Chitnis et al. Hypothesis testing, type I and type II errors. *Ind Psychiatry J.* 2009;18(2):127–131.
2. Introduction of hypothesis testing. **Available on:** http://www.sagepub.com/upm-data/40007_Chapter8.pdf [Access on 18/05/2021].
3. Visweswara Rao K. *Biostatistics – Test of Significance*. Second Edition (Jaypee Brothers Medical Publishers). New Delhi, India: 2007:126 – 130.
4. Veazie, P. J. (2015). *Understanding Statistical Testing*. *SAGE Open*, 5(1). <https://doi.org/10.1177/2158244014567685>

5. Sim J, Reid N, Statistical Inference by Confidence Intervals: Issues of Interpretation and Utilization, *Physical Therapy* 1999;79(2):186–195.

DOI: <https://doi.org/10.1093/ptj/79.2.186>