INTERNATIONAL JOURNAL OF RESEARCHES IN BIOSCIENCES, AGRICULTURE & TECHNOLOGY
© VISHWASHANTI MULTIPURPOSE SOCIETY (Global Peace Multipurpose Society) R. No. MH-659/13(N)
www.vmsindia.org

USE OF STATISTICAL METHODS IN LIFE SCIENCES

Gayatri Behere
Assistant Prof. in Statistics, Institute of Science, Nagpur
(gayatribehere70@gmail.com)

Abstract:. Statistics for the life sciences is almost synonymous with biostatistics. It incorporates quantitative modeling and methods of data analysis for clinical and epidemiological research (e.g. survival analysis), which in the past twenty years have become indispensable in medical research. In this paper, we study the use of large sample tests in life sciences which is illustrated with the help of examples.
Keywords: Parameter, statistic, hypothesis, critical value, level of significance, large sample test.

Introduction: Statistics is the art of drawing conclusions about phenomena in which chance plays a role. Statistical techniques have proved to be extremely useful in the study of natural sciences like botany, medicine, meteorology, etc. The purpose of statistical surveys is to obtain information about populations.

Definitions and terminology:

Population: By population we understand a group of units defined according to the aims of a survey. Thus population is an aggregate of objects under study. This population may be finite or infinite.

Sample: Since it is usually impracticable to test every member of a population, a sample from the population is the best approach available. A finite subset of statistical individuals on a population is called a sample. And the number of individuals in a sample is called the sample size.

Statistic: A statistic is defined as a numerical quantity (such as mean) calculated in a sample. Such statistics are used to estimate parameters.

Parameter: A parameter is numerical quantity measuring some aspect of a population of scores.

Sampling error.: The error involved in approximation is called as sampling error.

S.E. : The standard deviation of the sampling distribution of a statistic is known as its standard error, abbreviated as S.E.

Critical region: A region in the sample space S which amounts to rejection of H_0 is termed as critical region or region of rejection.

Level of significance: The probability α that a random value of the statistic t belongs to the critical region is known as the level of significance.

Hypothesis testing: It is a method of inferential statistics. An experimenter starts with a hypothesis about a population parameter called the null hypothesis. Data are then collected and the viability of the null hypothesis is determined in the light of the data. If the data are very different from what would be expected under the assumption that the null hypothesis is true then the null hypothesis is rejected. If the data are not greatly at variance with what would be expected under the assumption that the null hypothesis is true, then the null hypothesis is not rejected.

Procedure for testing of hypothesis in large samples :

1) Set up Null Hypothesis H_0:
2) Set up Alternative Hypothesis H_a :This will enable us to decide whether we have to use a right tailed test or left right tailed test or a two right tailed test.
3) Fix an error level you are comfortable with (something like 10%, 5%, or 1% is most common). Denote that " error level" by the letter "α" If no prescribed comfort level α is given, use 0.05 as a default value
4) Test Statistics: Compute the test statistic: $Z = (t - E(t)) / (S.E.(t))$ under the null hypothesis.
5) Conclusion: We compare the value of z calculated in step 4 with the tabulated value Z at the given level of significance.

If $|Z| < Z_\alpha$ then we accept H_0 and If $|Z| > Z_\alpha$ then we reject H_0

Level of Significance	Critical value of test statistic Z
1%	2.58
5%	1.96
10%	1.64

Distribution of the standardized test statistic and rejection region

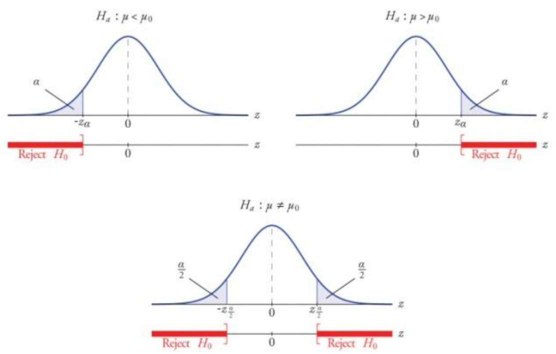

Illustration 1): A nutritionist is interested in whether two proposed diets " X" and "Y" work equally well in providing weight-loss for customers. In order to assess a difference between the two diets, she puts 50 customers on "X" diet and 60 other customers on the ``Y " diet for two weeks. Those on the former had weight losses with an average of 11 pounds and a standard deviation of 3 pounds, while those on the latter lost an average of 8 pounds with a standard deviation of 2 pounds. Do the diets differ in terms of their weight loss? (use α = 0.01)

Solution: Let μ_x and μ_y be the true average weight loss for X and Y diets, respectively. Then, we are asked to test the hypotheses:

$$H_0: \mu_x - \mu_y = 0 \quad VS \quad H_1: \mu_x - \mu_y \neq 0$$

For such a test, the appropriate test statistic is

$$\therefore\ Z = \frac{\overline{X_x} - \overline{X_y}}{\sqrt{\frac{{\sigma_x}^2}{n_x} + \frac{{\sigma_y}^2}{n_y}}} = 6.11$$

and the two-sided rejection region is

$$Z < -z_{0.01} = -2.575 \quad \text{and} \quad Z > z_{0.01} = 2.575.$$

Conclusion: So, we can clearly reject the null hypothesis and say that the two diets differ at the 1% significance level.

Illustration 2): Twenty people were attacked by a disease and only 18 survived. Will you reject the hypothesis, that the survival rate, if attacked by this disease is 85%, in favour of the hypothesis, that it is more, at 5% level.

Solution: Given: n=20,

X = No. of persons who survived after attacked by a disease =18

P = Proportion of the persons survived in the sample = 18/20=0.9

Ho: P=0.85 i.e., The proportion of the persons survived after attack by a disease is 85%

H_1: P > 0.85 (Right Tailed Alternative)

Under Ho, the test statistic is

$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$ (since, sample is large)

$$Z = \frac{0.90-0.85}{\sqrt{\frac{0.85 \times 0.15}{20}}} = 0.633$$

Conclusion: The significant value of Z at 5% level of significance for right tailed test is 1.645. Since, computed value of Z=0.633 is less than 1.654, it is not significant and we may accept the null hypothesis at 5% level of significance.

Illustration 3) A group of 120 boys obtained a mean intelligence quotient (I.Q.) of 84 while a group of 80 girls obtained mean intelligence quotient (I.Q.) of 86. If the standard deviation of the intelligence quotients is given to be 10, can we conclude that there is a significant difference between their performances.

Solution: $H_0: \mu_1 = \mu_2$ and given that $\overline{x_1} = 84, \overline{x_2} = 86, n_1 = 120, n_2 = 80$

$$\therefore\ Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} = 2.7716$$

Conclusion: Calculated $|Z| > 1.96$ at 5% level of significance. i.e., 2.7716 >1.96

∴ We reject H_0 i.e., there is a significant difference between their performances.

Remarks: Quantitative methodologies have always been central in life sciences and some of these belong to the general area of Statistics. Other methods of data analysis have their part to play and it is of ongoing importance for the field of Statistics to adapt to these developments. Statistical methods generate new knowledge. Statistical data from official sources and elsewhere play important roles in life sciences.

References:

1) Fundamentals of Mathematical Statistics by S.C. Gupta & V.K. Kapoor : S. Chand & Sons

2) ***Statistical Methods for Research Workers by*** R.A. Fisher **:** Edinburgh: Oliver & Boyd, 1925 (1st edition); London: Macmillan, 1970 (15th edition)

3) ***Statistical Methods by*** George W. Snedecor **:** 1937, Collegiate Press

4) ***Principles and Procedures of Statistics with Special Reference to the Biological Sciences.by*** Steel, R.G.D, and Torrie, J. H.: McGraw Hill (1960)

5) ***Biometry: The Principles and Practices of Statistics in Biological Research*** by Robert R. Sokal; F. J. Rohlf : 3rd ed. Freeman & Co. (1994)

6) www.google.com

❖❖❖❖