UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Professor Irfan Ali Khan

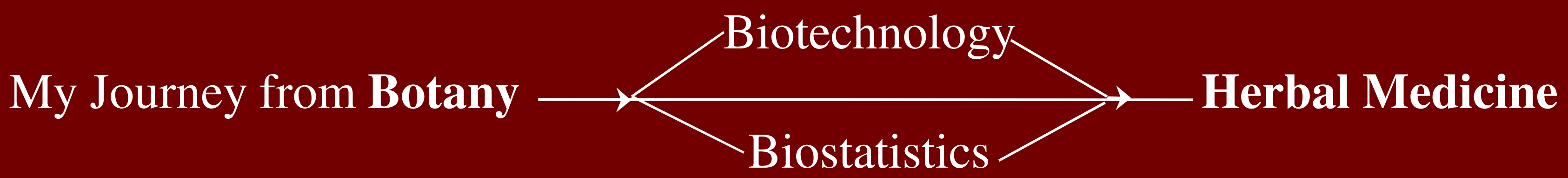
Former Director : Nawab Shah Alam Khan Centre For Post Graduate Studies and Research, Anwarul Uloom College Campus, New Mallepally, Hyd-500001

Editor-in-Chief : Annals of Phytomedicine : An International Journal

Managing Director: Ukaaz Publications, Hyderabad, A.P.

Email: ukaaz@yahoo.com, Website: www.ukaazpublications.com

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS



My Journey from **Botany**

Biotechnology

Biostatistics

Herbal Medicine

“You must do the things you think you can not do” Eleanor Roosvelt

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Thematic map

1. **Introduction**
2. **Measures of Central Tendency and Dispersion**
   - i. Introduction
   - ii. Arithmetic Mean
   - iii. Variance
   - iv. Standard Deviation
   - v. Standard Error of Mean
   - vi. Coefficient of Variation
3. **The Test of Significance: The ‘t’ test**
4. **Least Significance Difference : The “LSD” test**

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Formulae used

Mean : $\bar{x} = \dfrac{\Sigma x}{n}$

Variance : $s^2 = \dfrac{\Sigma(x - \bar{x})^2}{n - 1}$

Standard Deviation : $s = \sqrt{\dfrac{\Sigma(x - \bar{x})^2}{n - 1}}$

Standard of Error of mean : $SE_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

Coefficient of Variation : $CV\ (\%) = \dfrac{s}{\bar{x}} \times 100$

Test of Significance : $t = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Least Significance Difference : $LSD_\alpha = (t_\alpha)\ (SE_{\bar{d}})$

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

1. INTRODUCTION

i. **Biostatistics :** Biostatistics has been defined as the application of statistical methods to biological sciences. Therefore, a good understanding of biostatistics is essential as the methods of biostatistics are indispensable tools for the design and analysis of data and in the interpretation of experimental results for dependable conclusions.

ii. **Data:** The first step of biostatistical study is the collection of data, such as age, weight, height, body temperature, blood pressure, marital status *etc*. These characteristics are referred to as variables and the values of the observations recorded for them are referred to as data. There are two types of biostatistical data.

Primary data and Secondary data

iii. **Parameter :** In respect of any variable, the numerical quantities which characterise a population are called parameters of the population. For example, if the variable is length and the measurements of length are taken for a large population of animal or plant, the mean length can be regarded as parameter.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

iv. **Variable :** The specific characteristics like age, sex, height, weight *etc*. which may be assessed for a certain population are referred to as variables. Variables may be categorized further as qualitative or quantitative variables.

v. **Population :** The population in a statistical investigation refers to any well defined groups of individuals or of observations of a particular type. For example, all plants of a variety of wheat present in one acre of land, referred to as population. Similarly all fishes of one species present in a particular pond and also all patients of a hospital of Delhi may be considered as population. Example: Finite and Infinite population.

vi. **Sample :** In our day-to-day life, we adopt the sampling technique almost every moment of our existence. In medical sciences, a few drops of blood are taken and tested microscopically or chemically to know whether the blood contains some abnormalities or not. Whatever is observed in the few drops is true for the whole blood of the body. From the above discussion, it is apparent that a **few items are selected from the population in such a way that they are representative of the population.** Such a section of the population is called a ***sample*** and the process of selection is called ***sampling***.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

"Sample may be defined as fraction of a population, drawn by using a suitable method so that it can be regarded as representative of the entire population".

vii. **Size of the Sample:** In agricultural, biological and also medical sciences, the size of the sample plays an important role. The size of the sample is the number of sampling units, which are selected from a population for investigation. If the size of the sample is small, it may not represent the population and it will not be possible to ascertain the accuracy of results. On the other hand, if the size of the sample is large, greater will be the representation of the items of the population in it and it will be quite difficult to manage. Therefore, the sample size should be neither too big nor too small. It should be "optimum". However, **"An optimum sample is one which fulfils the requirements** of efficiency, representativeness, reliability and flexibility.

When the size of the sample is small, the difference will be relatively high. However, in the samples of large size, the variance will show little difference. Therefore, while computing variance when the sample size is less than 30, one should divide it by n-1 to obtain an estimate of variance.

viii. Clinical trials : A clinical trial is an experiment that seeks to determine the effectiveness of a new drug or treatment under controlled conditions. A clinical trial involves a comparison of two or more comparable groups of patients. The treatment group, which receives a potentially therapeutic agents, is compared with a similar control group, which receives a placebo. In such type of study, the investigator is interested not only in establishing comparable groups, but also in limiting the amount of bias entering a trial. One way to do this is to design the experiment as a single-blind study. In this type of experiment, the patient does not know whether he or she is in the treatment or the control group. The other better way is to design it as a double-blind study, in which neither the patient nor the experimenter knows to which group the patient is assigned. A neutral person keeps the code as to whose who and discloses it only at the end of data collection.

Phases of Clinical Trials

Phase 0 trials : They are the first-in-human trials. Single subtherapeutic doses of the study drug are given to a small number of subjects (10 to 15) to gather preliminary data on the agent's pharmacodynamics (what the drug does to the body) and pharmacokinetics (what the body does to the drugs).

Phase I trials: Researchers test an experimental drug or treatment in a small group of people (20–80) for the first time. The purpose is to evaluate its safety and identify side effects.

Phase II trials: The experimental drug or treatment is administered to a larger group of people (100–300) to determine its effectiveness and to further evaluate its safety.

Phase III trials: The experimental drug or treatment is administered to large groups of people (1,000–3,000) to confirm its effectiveness, monitor side effects, compare it with standard or equivalent treatments, and collect information that will allow the experimental drug or treatment to be used safely.

Phase IV trials: After a drug is approved by the FDA and made available to the public, researchers track its safety, seeking more information about a drug or treatment's risks, benefits, and optimal use.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

2. MEASURES OF CENTRAL TENDENCY AND DISPERSION

i. Introduction

Although frequency distributions serve useful purposes, there are many situations that require other types of data summarization. Measures of central tendency are sometimes needed to make meaningful interpretation of the data. Generally, it is found that in any distribution, values of the variable tend to congregate around a central value of the distribution. This tendency of the distribution is known as its central tendency and the measures devised to consider this tendency are known as ***measures of central tendency*.**

Average

However, one of the most important objectives of this statistical analysis is to get a single value that describes the characteristic of the entire mass of data. Such a value is called an "***average***".

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Type of Averages

Mathematical averages : Arithmetic mean, Geometric mean, Harmonic mean

Positional averages : Median, Mode

Dispersion

The different measures of the central tendency, gives an idea of the central value, whereas measures of dispersion are the measures of spread about an average. The central value may be the same in two or more distributions, but may differ in respect to dispersion. Dispersion may be defined as the measure of variation of the variable. Dispersion is the degree of the variation of the variable about a central value.

The Measures of dispersion are:

Range, Mean deviation, Variance, Standard deviation

ii. Arithmetic Mean

Introduction

The most familiar and widely used measure of central tendency is the arithmetic mean. It represents the entire data by one value which is obtained by adding together all the values and dividing this total by the number of observations. In other words, the arithmetic mean is the sum of all observations divided by the total number of observations. However, each independent value plays an equal part in the determination of the mean.

Calculation

The arithmetic mean is computed by summing up the observations and dividing the sum by the total number of observations. Symbolically;

$$\overline{x} = \frac{x_1 + x_2 + x_3 ------ + x_n}{n} \quad \text{or} \quad \overline{x} = \frac{\Sigma x}{n}$$

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

where $\bar{x}$ = arithmetic mean

$\bar{x}$ = sum of all values of the variable, x *i.e.* x_1, x_2, x_3 ----x_n

n = number of observations

Example

The following data represent the number of patients admitted in different hospitals in Hyderabad. Calculate the mean number of patients per hospital.

Number of patients = 10, 11, 10, 11, 9, 7, 9, 11, 12, 10

Solution

Σx **= 100** n = 10 $\bar{x}$ **=** $\frac{\Sigma x}{n}$

$$\bar{x} = \frac{100}{10} = 10$$

$\bar{x}$ = 10

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Application : OGTT of *Costus pictus* leaves methanolic extract and its fractions

Blood glucose level (mg/dl)				
		After glucose administration		
Groups	**Fasting**	**60 min**	**90 min**	**120 min**
Normal	76.0±9.63	124.6±15.44	115.6±12.10	83.4±8.65
Metformin	88.8±5.96	52.4±7.10***	40.2±6.19***	46.2±7.05*
CPME-200	55.4±6.38	81.8±5.36	71.4±6.38	46.0±6.07*
CPME-500	56.8±5.89	73.2±7.02*	64.2±7.02*	44.6±8.25*
CPHF-200	62.6±7.75	101.6±8.89	92.8±7.64	74.2±7.21
CPHF-500	62.6±7.75	93.8±5.44	80.6±10.25	72.2±7.16
CPEAF-200	41.4±3.39	75.2±7.00*	70.6±4.93*	54.6±4.91
CPEAF-500	51.0±10.83	61.6±13.76**	56.6±7.14**	45.2±7.42*

“Values are expressed as Mean ± SE$_{\bar{X}}$, n = 5” *p<0.05, **p<0.01, ***p<0.001 when compared to normal group CPME (CP Methanol extract), CPHF (CP Hexane fractional) and CPEAF (CP Ethyle acetate fraction)

Annals of Phytomedicine : 2(1) : 89-94 (2013)

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Conclusion: The main aim of this presentation is to evaluate the antidiabetic potential of different fractions of methanolic extract of *Costus pictus* D. Don in Swiss Albino mice. The antidiabetic effect of different fractions will be evaluated, based on the decreasing trend in mean values.

During OGTT, the oral administration of CPME-500 and CPEAF-500, blood glucose level was decreased significantly at each time interval (60, 90, and 120 min). However, CPEAF-200 exhibited significant (p<0.05) antihyperglycemic effect after 60 min and 90 min whereas CPME-200 showed significant (p<0.05) changes after 120 min only. While CPHF failed to pose any change in oral glucose tolerance at the tested doses. These results clearly indicated that CPEAF is active fraction of CPME. Therefore, further studies should be conducted in this direction for valuable findings.

iii. Variance

Introduction

This term was first introduced by R.A. Fisher in 1913. The term "Variance" is used to describe the square of the standard deviation. Therefore, the analysis of variance helps us in isolating the effects of various factors. To calculate the variance, the deviations of the variables from the mean are squared and then added. The sum of the squares of deviations is divided by the number of observations to get the variance of the sample. This measure has an advantage over the mean deviation because the sum of squares of deviations is always positive. The variance is defined as the mean of squares of deviations.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Calculation

Sample variance : $s^2 = \frac{\Sigma(x - \bar{x})^2}{n-1}$ or $\frac{\Sigma(d\bar{x})^2}{n-1}$

Population variance : $\sigma^2 = \frac{\Sigma(x - \bar{x})^2}{n-1}$ or $\frac{\Sigma(d\bar{x})^2}{n-1}$

where $\bar{x}$ = arithmetic mean

n = number of observations

Check : The results of the above mentioned formula can be checked up with the help of the following formula :

Variance :

$$s^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}$$

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Example

Calculate the variance from the data, recorded on a medicinal important plant.

Length of leaves (cms.) = 6.5, 6.6, 6.7, 7.0, 7.5, 7.6, 8.0, 9.0, 9.5, 10.0

Solution

Length of leaves

x	**x^2**
6.5	42.25
6.6	43.56
6.7	44.89
7.0	49.00
7.5	56.25
7.6	57.76
8.0	64.00
9.0	81.00
9.5	90.25
10.0	100.96
$\Sigma x = 78.4$	$\Sigma x^2 = 628.96$

$$s^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}$$

$$= \frac{628.96 - \frac{(78.4)^2}{10}}{10-1} = \frac{628.96 - \frac{6146.56}{10}}{9}$$

$$= \frac{628.96 - 614.66}{9}$$

$$= \frac{14.30}{9} = 1.59$$

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Application : Mean ± $SE_{\bar{x}}$, Variability and shift in mean for plant yield in M_2 generation.

Treatments	Mean± $SE_{\bar{x}}$	s^2	Shift in mean	Significance
c	6.28 ± 0.16	14.20	-	-
20 kR	7.66 ± 0.36	16.11	1.38	s
0.2% EMS	9.66 ± 0.31	12.12	3.38	s
0.2% NMU	7.00 ± 0.38	14.30	0.72	ns

Selection studies in M_3 generation for 0.2% EMS

Sections	Mean± SE_x	Shift in mean	CV(g) %	h^2%	Gs
c	06.94 ± 0.21	-	-	-	-
Random Selection	10.50 ± 0.28	3.56**	07.36	18.16	1.96
Positive Selection	11.00 ± 0.34	4.06**	11.08	22.00	2.34
Negative Selection	07.00 ± 0.37	0.06^{ns}	05.46	10.78	1.00

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Conclusion

Radiation (20 kR) as well as chemical mutagen (0.2% EMS and 0.2% NMU) have induced more variability in M_2 generation. Increase in mean values in M_2 and M_3 generations of plant yield might be due to elimination of plants and also due to genetic nature of changes induced after mutagenic treatments. The increase in mean value may also be due to recovery effect in the successive generations.

The selection were planned based on the significance of mean values as well as the variability in the treatments. In this case, 0.2% of EMS was found to be most effective treatment to make the selection studies in M_3 generation. In the present study, it was observed that the random and the positive selections gave a positive direction for further studies.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

iv. Standard Deviation

Introduction

Standard deviation is the most commonly used absolute measure of dispersion. The concept of standard deviation was first introduced by Karl Pearson in 1893. The term "standard" is assigned to this measure of variation is probably because it is the most commonly used and also most flexible in terms of variety of applications of all the measures of dispersion. It is clear that standard deviation is a measure of the spread in a set of observations. Standard deviation is defined as the **square root of the variance**.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Calculation

(i) Standard deviation, $s = \sqrt{\frac{\Sigma(x - \overline{x})^2}{n - 1}}$ or $s = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n \text{ or } n-1}}$

As it measures the dispersion or variability of a distribution, the larger the standard deviation mean, the greater is the value of variability. On the other hand, there will be homogeneity in a series when the standard deviation is small.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Example : Calculate the standard deviation from the following data :
Variable (x) = 10, 13, 17, 22, 27, 30, 31, 32

Solution

Variable (x)		$(d\overline{x})^2 = (x - \overline{x})^2$
10	-12.75	162.56
13	-9.75	95.06
17	-5.75	33.06
22	-0.75	0.56
27	4.25	18.06
30	7.25	52.56
31	8.25	68.06
32	9.25	85.56
Σx = 182		$\Sigma\,(x - \overline{x})^2 =$ 515.48

$$\overline{x} = \frac{\Sigma x}{n} = \frac{182}{8} = 22.75$$

$$s = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n-1}} = \sqrt{\frac{515.48}{8-1}} = \sqrt{\frac{515.48}{7}}$$

$$= \; = \sqrt{73.64} = 8.58$$

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Applications : Phenolic compounds in methanolic extract and various fractions of methanolic extract of coriander seeds

Sample extract	Total phenolic (mg/100g) [Gallic acid equivalents (GAE)]	Standard deviation	Total flavonoids (mg/100g) [Rutin equivalents (RE)]	Standard deviation	Total flavonols (mg/100g) [Rutin equivalents (RE)]	Standard Deviation	Tannins (mg/100g) [Catechin equivalents (CE)]	Standard Deviation
Methanol	2.77±0.04	8.11	0.51±0.09	9.00	1.40±0.02	9.10	0.76±0.04	10.10
Hexane	5.63±0.02	10.10	2.21±0.05	11.10	1.35±0.04	10.20	0.02±0.04	10.20
Benzene	2.66±0.06	11.92	1.63±0.03	12.82	0.64±0.02	11.30	0.45±0.10	11.30
Ethyl acetate	7.77±0.13	8.00	3.40±0.10	8.49	1.80±0.06	7.12	0.39±0.06	12.10
n-butanol	5.76±0.17	11.55	3.30±0.09	11.00	1.60±0.07	12.30	0.80±0.08	9.00
Aqueous	2.54±0.02	12.00	0.50±0.04	11.99	0.26±0.03	10.00	0.24±0.03	10.99

Annals of Phytomedicine : 2(2) : 2013

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Conclusion

The ethyl acetate fraction of coriander seeds was found to possess maximum amount of total phenolics, flavonoids and flavonols. The above mentioned findings indicate that many of the phytochemicals got extracted in the ethyl acetate fraction and therefore ethyl acetate fraction can be further subjected to sub-fractionation either using column chromatography or any other suitable analytical technique in order to identify and isolate its active components which can be further tested for biological activity. If proved effective, then these compounds can be isolated in large quantities and can be used as a source of antioxidants suitable for application in nutritional and pharmaceutical fields.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

v. Standard Error of Mean

Introduction

Standard error is a statistical term that measures the accuracy with which a sample represents a population. In statistics, a sample mean deviates from the actual mean of a population; this deviation is the standard error.

The smaller the standard error, the more representative the sample will be of the overall population. The standard error is also inversely proportional to the sample size; the larger the sample size, the smaller is the standard error.

Calculation

$$SE_{\overline{x}} = \frac{s}{\sqrt{n}}$$

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Example

Maturity data recorded on an early maturing mutant variety of Castor (Aruna). Calculate the standard error of mean, from the following data:

Days to maturity = 140, 140, 141, 141, 142, 145, 146, 150, 150, 155

Solution

x	$(x-\bar{x})$	$(x-\bar{x})^2$
140	-5	25
140	-5	25
141	-4	16
141	-4	16
142	-3	9
145	0	0
146	1	1
150	5	25
150	5	25
155	10	100
Σx = 1450		$\Sigma(x-\bar{x})^2$ =242

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1450}{10} = 145$$

Variance, s^2 $= \frac{\Sigma\,(x-\bar{x})^2}{n-1} = \frac{242}{10-1}\ \frac{242}{9} = 26.89$

s $= \sqrt{\frac{\Sigma\,(x-\bar{x})^2}{n-1}} = \sqrt{\frac{242}{10-1}} = \sqrt{\frac{242}{9}}$ $= \sqrt{26.89} = 5.19$

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{5.19}{\sqrt{10}} = \frac{5.19}{3.16} = 1.64$$

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Application: Effect of *Semicarpus anacardium* Linn. seed extract on ethanol induced gastric ulcer in rats.

Group	Treatment	Ulcer Index (Mean ± SE$_{\bar{x}}$)	% Protection from Ulcer
I	Control	4.000 ± 0.632	…..
II	Omiprazol (20 mg/kg)	1.167 ± 0.307*	70.82
III	SAAE (250 mg/kg)	4.33 ± 0.881 ns	– 8.33
IV	SAEE (250 mg/kg)	0.500 ± 0.223**	87.50
V	SAPE (250 mg/kg)	1.16 ± 0.477*	70.82

Values are expressed as Mean ±SE$_{\bar{x}}$ at *p<0.05, at **p<0.01, ns-indicates non-significant. SAPE= *Semicarpus anacardium* Linn. petroleum ether extract, SAEE= *Semicarpus anacardium* Linn. ethanolic extract and SAAE= *Semicarpus anacardium* Linn. aqueous extract.

Annals of Phytomedicine: 1(1) 105-109 (2012)

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Conclusion

It is concluded that the effect of *Semicarpus anacardium* Linn. on ethanol induced gastric ulcer in rats. Administration of ethanol in control group produced maximum ulcer, the ulcer index being 4.000 ± 0.632, the group treated with standard drug omiprazole, indicated the ulcer index being reduced to 1.167 ± 0.307, showing the reduction of 70.82 % at $p < 0.05$. No protection from ulcer was observed in animals which were treated with SAAE. However, significant reduction of 70.82 % and 87.50% were observed in animals treated with SAPE and SAEE, respectively.

The present study can be, further concluded that the *Semicarpus anacardium* Linn. seed extracts not only provides an excellent preventive effect in gastric ulcer models, but also possesses significant hepatoprotective effect. This may be due to the antioxidant nature of flavonoids present in them. Further studies on antioxidant parameters may be performed on the same lines.

vi. Coefficient of Variation

Introduction

The standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original data are collected. The standard deviation of the height of plants cannot be compared with the standard deviation of the weight of plants because they are expressed in different units, *i.e.* heights (cms.) and weights (gms.). Therefore, the standard deviation must be converted into a relative measure of dispersion for the purpose of comparison. The relative measure of dispersion is known as the ***coefficient of variation***.

Coefficient of variation is of great practical significance and is the best measure of comparing the variability or consistency of two or more samples. The smaller the coefficient of variation, the greater is its consistency. On the other hand, if the coefficient of variation is greater, the sample is said to be more variable or less homogeneous.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Calculation

$$CV\% = \frac{S}{\overline{X}} \times 100$$

where s = standard deviation

$\overline{X}$ = mean

Example : From the following data, compute the values of coefficient of variation. Compare the two series for the data consistency.

Series A : $\overline{X} = 34.60;$ $s = 14.00$

Series B : $\overline{X} = 27.10;$ $s = 8.00$

Solution : **Series A** **Series B**

$C.V.\% = \frac{s}{\overline{X}} \times 100 = \frac{14.0}{34.60} \times 100$ $C.V.\% = \frac{s}{\overline{X}} \times 100 = \frac{8}{27.10} \times 100$

$C.V.\% = 40.46\%$ $C.V.\% = 29.52\%$

This shows that the data of series 'A' are more variable and less consistent as compared to the data of series 'B' which are more consistent.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

ExamplJe : Two types of drugs were used on 10 patients for reducing their weight. Drug A was imported and drug B was indigenous. The decrease in weight after the use of drug for six months is as follows:

Drug A	12	10	10	13	12	11	10	12	14	13
Drug B	10	11	12	11	10	9	10	11	9	10

DRUG A			**DRUG B**		
x		$(x-\bar{x})^2$	**x**		$(x-\bar{x})^2$
12	**0.3**	**0.09**	**10**	**-0.3**	**0.09**
10	**-1.7**	**2.89**	**11**	**0.7**	**0.49**
10	**-1.7**	**2.89**	**12**	**1.7**	**2.89**
13	**1.3**	**1.69**	**11**	**0.7**	**0.49**
12	**0.3**	**0.09**	**10**	**-0.3**	**0.09**
11	**-0.7**	**0.49**	**9**	**-1.3**	**1.69**
10	**-1.7**	**2.89**	**10**	**-0.3**	**0.09**
12	**0.3**	**0.09**	**11**	**0.7**	**0.49**
14	**2.3**	**5.29**	**9**	**-1.3**	**1.69**
13	**1.3**	**1.69**	**10**	**-0.3**	**0.09**
$\Sigma x = 117$	$\Sigma(x-\bar{x})^2 = 18.10$		$\Sigma x = 103$	$\Sigma(x-\bar{x})^2 = 8.10$	

Source :
Professor Mohd. Ibrahim
Nizam Institute of Pharmacy,
Hyderabad

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

DRUG A

$$\bar{x} = \frac{\Sigma x}{n} = \frac{117}{10} = 11.7$$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{18.10}{10-1} = \frac{18.10}{9} = 2.01$$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{18.10}{10-1}} = \sqrt{\frac{18.10}{9}} = \sqrt{2.01}$$

$$= 1.42$$

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{1.42}{\sqrt{10}} = \frac{1.42}{3.16} = 0.45$$

$$C.V. = \frac{s}{\bar{x}} \text{ X } 100 = \frac{1.42}{11.7} \times 100 = \frac{142}{11.7} = 12.14$$

DRUG B

$$\bar{x} = \frac{\Sigma x}{n} = \frac{103}{10} = 10.3$$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{8.10}{10-1} = \frac{8.10}{9} = 0.90$$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{8.10}{10-1}} = \sqrt{\frac{8.10}{9}} = \sqrt{0.9}$$

$$= 0.95$$

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.95}{\sqrt{10}} = \frac{0.95}{3.16} = 0.30$$

$$C.V. = \frac{s}{\bar{x}} \text{ X } 100 = \frac{0.95}{10.3} \times 100 = \frac{95}{10.3} = 9.22$$

Conclusion: The mean weight reduced by drug A is higher than the drug B. However, the coefficient of variation is also higher for the drug A than drug B. It clearly indicates that the drug B (Indigenous) is more consistent or homogenous than the drug A (Imported) because the coefficient of variation is less than drug A.

3. THE TEST OF SIGNIFICANCE : The “t” test

Introduction

So far we have discussed problems relating to large samples. When the size of the sample is small (less than 30), the above tests are not applicable because the assumptions on which they are based, generally do not hold good in case of small samples. While dealing with small samples, the results will be true only if the parent population is normal.

Considerable attention has been paid in developing suitable tests for dealing the problems of small samples. The contribution to the theory of small samples is that of R.A. Fisher and William Scaly Gosset. Fisher gave a test, popularly known as ‘z’ test in 1926 and Gosset gave another test known as ‘t’ test respectively. Gosset published his discovery in 1908 under the pen name or pseudonym “student”. The ‘t’ distribution is based on the degrees of freedom of a distribution, *i.e.*, n – 1.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Example: (Source : Dr. Parth Patel : Medicity Hospital, Hyderabad**)**

A sample of 10 adults yielded a mean blood sugar level of 96 mg% with a standard deviation of 9 mg%. Estimate the mean of the blood sugar level with 95% confidence.

$\bar{x}$ = 96 mg%

s = 9 mg%

n = 10

SE $= \frac{s}{\sqrt{n}} = \frac{9}{\sqrt{10}} = \frac{9}{3.16} = 2.85$

Degrees of freedom = n – 1 = 10 – 1 = 9

Table value of "t" at 0.05 with 9 df = 2.26

96 ± 2.26 (2.85) = 96 ± 6.44 = 89.56 to 102.44 mg%

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Paired 't' Test

i. If we record the blood pressure of a person initially, and then after injecting a drug at appropriate intervals, the two recording of blood pressure constitute paired observations.

ii. If we take blood sample for estimation of blood glucose, divided it into two parts and have each part estimated in a separate laboratory, the results of the two tests again constitute paired observations.

iii. If we record the weights of the rights and left kidneys of an individual, the two observations can again be paired.

Unpaired 't' test

i. Applied to two independent groups *e.g.*, diabetic patients versus non-diabetics (or the control group).

ii. Similarly, to compare whether systolic blood pressure differs between a control and treated group, between men and women or any other two groups.

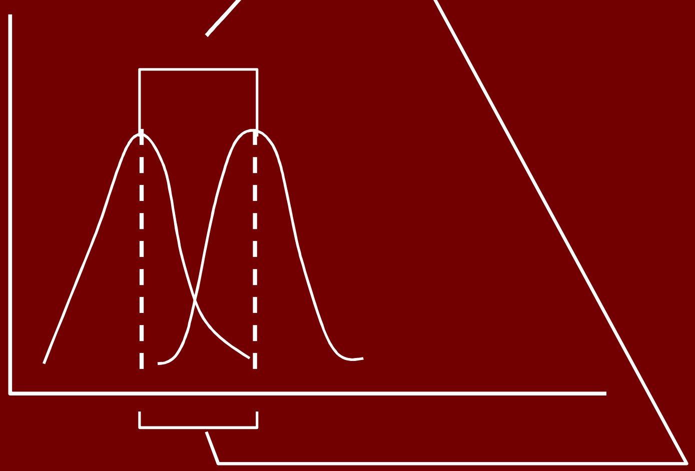
UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Statistical Analysis of the t-test

The formula for the t-test is a ratio. The upper part of the ratio is just the difference between the two means and the lower part is a measure of the variability. In the given formula of t-test, it shows how the numerator and denominator are related to the distributions.

$$t = \frac{\text{difference between groups means}}{\text{variability of groups}} = \frac{\bar{x}_t - \bar{x}_c}{SE(\bar{x}_t - \bar{x}_c)}$$

$= \text{t-value}$



Formula for the t-test

The upper part of the formula is to find the difference between the means. The lower part is called the **standard error of the difference**. To compute it, we take the variance for each group and divide it by the number of observations in that group. We add these two values and then take their square root. The specific formula is given below.

$$SE_{\bar{x}}(\bar{x}_t - \bar{x}_c) = \sqrt{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}}$$

Remember, that the variance is simply the square of the standard deviation.

The final formula for the t-test is as follows:

$$t = \frac{|\bar{x}_t - \bar{x}_c|}{\sqrt{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}}}$$

Calculation: Example : Paired 't' Test

Let us suppose that you are conducting an investigation into the weight of students in a school. Half the subjects are boys and half are girls. You think that the boys might have a significantly bigger mean weight than the girls. You record the weight of all the individuals and come up with the following informations:

Boys : Mean weight = 54 kg, Variance = 16 (n = 25)

Girls : Mean weight = 47 kg, Variance = 19 (n = 25)

Calculate the value of 't' for these two sets of data:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad t = \frac{|54-47|}{\sqrt{\frac{16}{25} + \frac{19}{25}}}$$

$$t = \frac{7}{\sqrt{\frac{16}{25} + \frac{19}{25}}} = \frac{7}{\sqrt{0.64+0.76}} = \frac{7}{\sqrt{1.4}} = \frac{7}{1.18} = 5.93$$

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Conclusion

Calculated “t” value = 5.93

Tabulated value at 5% = 2.021

Degree of freedom $n_1 + (n_2 – 2) = 25 + (25 – 2) = 48$

The calculated value of “t” is 5.93 which is bigger than the tabulated value both at 5% (2.021) and at 1% (2.704) levels of probability. There is indeed a significant difference between the means of the two sets of data. This also rejects the null hypothesis that there is no difference of means.

4. Least significance Difference : The LSD test

Introduction

In agricultural, biological and medical research, pair comparison is more common. Comparison is always made among the treatments and also the doses of the drug to find out the most effective treatment.

The most commonly used test procedure for such type of pair comparison is the Least Significance Difference (LSD) test which is suited for planned comparisons. This procedure provides for a single LSD value; at a prescribed level of significance which serves as the boundary between significant and non-significant differences between any pair of comparisons. It can be concluded that the treatment or dose of the medicine is significant if the difference exceeds the computed LSD value; otherwise it is not significant. This test is applied only in situations where number of comparisons are small.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Explanation

The LSD test can be applied in pair comparisons which is as follows :

Compute the mean difference between the two means.

$$\bar{d}_{ij} = \bar{x}_i - \bar{x}_j$$

where $\bar{x}_i$ and $\bar{x}_j$ are the means of two treatments.

Compute the LSD value at a level of significance.

$$LSD_\alpha = (t_\alpha)\ (SE\ )$$

Where $SE_{\bar{d}}$ is the standard error of the mean difference and t_α is the tabulated 't' value at 0.05 or 0.01 level of probability.

Compare the mean difference and LSD value, the difference of the specific pair of means is significant at the "α" level of significance if the value of is greater than the LSD value otherwise it is not significantly different.

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Example

In an experiment, fresh weight of the plant was recorded for control and 5 doses of a chemical mutagen (EMS). Compare the mean fresh weight of control and five doses (n = 120).

Doses	Fresh weight (gm)	$SE_{\bar{x}}$
Control	12.41	0.6941
0.1% EMS	13.40	0.7241
0.2% EMS	15.69	0.8211
0.3% EMS	13.70	0.6224
0.4% EMS	10.10	0.6112
0.5% EMS	13.00	0.7001

UNDERSTANDING BASIC BIOSTATISTICS FOR PROPER APPLICATION OF STATISTICAL METHODS

Calculation

Doses	Fresh Weight (gms)	$SE_{\bar{x}}$	Difference from Control	LSD value at	
				0.05	0.01
Control	12.41	0.6941	-	-	-
0.1% EMS	13.40	0.7241	0.99^{ns}	1.42	1.87
0.2% EMS	15.69	0.8211	3.28**	1.61	2.12
0.3% EMS	13.70	0.6224	1.29*	1.22	1.61
0.4% EMS	10.10	0.6112	2.31**	1.20	1.58
0.5% EMS	13.00	0.7001	0.59^{ns}	1.37	1.81

Hint : Calculation : Example – 10 kR

at 5% = $LSD_{5\%} = (t_{\alpha})\ (SE_{\bar{d}})$

(1.960) (.7241)

1.960 × .7241

= 1.42

at 1% = $LSD_{1\%} = (t_{\alpha})\ (SE_{\bar{d}})$

(t 2.576) (.7241)

2.576 × .7241

= 1.87

ns = non-significant

* = significant at 5 % level of probability.

** = significant at 1 % level of probability

Conclusion

The difference between control and different doses indicates that the 0.2% and 0.4% EMS doses are highly significant. Whereas 0.1% and 0.5% EMS doses are not significant.

“TRY NOT TO BECOME A MAN OF SUCCESS BUT A MAN OF VALUE”- Albert Einstein

THANK YOU FOR HEARING ME OUT