18-12-2021

EXPERIMENT NO: 09

AIM OF THE EXPERIMENT: Statistical analysis using Microsoft Excel, and perform t-test.

THEORY:

Microsoft Excel is one of the most powerful tools in the Microsoft office package. It has inbuilt simple as well as complex mathematical analytical programs, and any calculation may be done using the Excel. In addition, the software may be used to develop graphical presentations, like pie charts, bar diagrams, line diagrams etc. The key of using MS Excel is the first input of data.

MATERIALS REQUIRED:

Sample data, Computer, Ms Excel Software, Notebook, Pen.

The blood glucose level is two groups of individuals, normal and diabetics in mg/dL is given below.

Table 1: The blood glucose level data in mg/dL.

Groups	
Normal	Diabetic
98.00	154.00
114.00	132.00
102.00	141.00
123.00	133.00
102.00	141.00
110.00	210.00

PROCEDURE:

(A) There are two groups of data on glucose levels, one in normal people and one in Diabetics.

(B) We are to perform some basic statistics and then see if there is any difference (statistically significant) between the means of the two groups.

(C) The data for both the groups was typed/entered in Microsoft Excel.

(D) Calculating Count (n): Count refers to the total number of observations per groups. Count is done by typing = count(data) in a clear cell of the Excel Spreadsheet. Here, and all throughout, data is given in array of cells, e.g. D1-D6.

$$S_x = \sqrt{\frac{\sum (X_i - \bar{X})^2}{(n-1)}}$$

where, X_i is the individual value

$n-1$ is the degrees of freedom, which denotes how many other possible values are there for a particular observation. Here, one observation is itself, and thus other possible ob observations = (total observations) -1, which is called degrees of freedom.

n is the count or number of observations.

① Standard Error (SE): Here, we calculate the Standard Error of the Mean (SEM) of the mean of the data. SEM detects the quality of data. Lower the SEM, better is the data and its reproducibility. It is calculated in Excel by the following formula, by typing = stdev(data)/sqrt(n) in excel.

$$\text{Standard Error} = \frac{S}{\sqrt{n}}$$

(J) Student-test:

T-test is a statistical analysis that determines whether two data sets and their relevant means, which are different, are due to some specific reason or simply due to by chance. Here, we assume that the mean of the Diabetic individuals and those of normal are different or not. A statistical validation is done by t-test, developed by Grossett.

For a t-test, certain basic details are needed. First we determine whether the data are from same individuals or different groups. In the former case, it is called paired t-test and in latter case it is called un-paired t-test. In MS Excel, paired tests are denoted by number 1 and unpaired by 2 or 3.

Further, if we assume that the data of the Diabetics is higher than that of the normal individuals, we go for one-tailed t-test, and if we are unaware whether the data will be variable in either directions, then we go for two-tailed test. For tests that are one-tailed, it is denoted by 1 in MS Excel, and for two-tailed tests, we refer by 2.

Before performing a t-test, we propose two hypotheses:

Null Hypotheses: which states that the two means are same, or the difference between the two means is just by chance.

Alternative hypothesis: It states that the two means are different, and the difference between the two means is not just by chance, rather there is some other factor affecting the difference. In the present case, it is diabetes which is affecting the blood sugar level.

Finally, following the analysis (t-test), we either accept or reject the Null hypothesis.

The command ttest in MS Excel is used for finding the p value. This is a probability value, say x, which denoted that the probability that the two means are different is x. In statistics, we consider a probability of over 95% to be significant. So, if ttest returns a p value less 0.05, we say that the difference between the means is significant, and not due to chance.

So, the t-test was performed in MS Excel by typing = ttest (data 1, data 2, 1, 2) in a clear cell.

Here, data 1 is array 1 (normal), data 2 is array 2 (diabetic), 1 refers to tails (one-tailed in our data) and 2 refers to type (unpaired t-test).

RESULTS:

Table 2: Results of the statistical analysis.

Test parameters	Groups		Formula used in excel
	Normal	Diabetic	
	98.00	154.00	
	114.00	132.00	
	102.00	141.00	
	123.00	133.00	
	102.00	141.00	
	110.00	210.00	
Count (n)	6.00	6.00	= count (data)
Mean	108.17	151.83	= average (data)
Median	106.00	141.00	= median (data)
Mode	102.00	141.00	= mode (data)
Standard Deviation	9.35	29.57	= stdev (data)
Standard Error	3.82	12.07	= stdev(data)/sqrt(n)
Variance	87.4	874.2	= stdev^2
ttest (pvalue)	0.007		= ttest(data1, data2, 1, 3)

The t-test value thus returned by the command is the p value, which is 0.007.

INTERPRETATION :

A p value of 0.007 may be interpreted by several ways. First, anything smaller than 0.05 is regarded statistically significant, that is the means are different, and there is actually something else which is affecting the differences in the means. In our case, the means glucose levels of normal and diabetic individuals are different, which is not due to any chance, rather due to some other factor, i.e. diabetes. So, we REJECT the Null hypothesis and conclude that Diabetes causes increase in blood sugar level.

Again, the p value denotes that out of 100 such sampling or tests, the probability that we get different means is equal to 100 - p value = 100 - 0.007 = 99.993, or 99.993%. In other words, there is only 0.007% chance that we may gt same means.

CONCLUSION :

Based on the statistical analysis, we conclude as follows:

① The two data sets have different means.

② The median of the two groups, normal and diabetic, are 106 mg/dL and 141 mg/dL respectively.

③ The mode of the two groups, normal and diabetic, are 102 mg/dL and 141 mg/dL respectively.

④ The Standard deviations of normal individuals is 9·35, while that of Diabetics is quite higher. 29·57. For the normal group, the spread of the data is quite lesser.

⑤ Standard error of mean (SEM) of normal individuals is 3·82, which is quite promising indicating good reproducibility of the data, or high quality of the data. However, in the diabetic group, the SEM is larger. 12·07.

⑥ The calculated p values 0·007. This indicates that we reject the Null hypothesis.

⑦ Thus, the means of the two groups are different not just by chance. So, we infer that diabetes causes increase in blood sugar level.

⑧ Further, the chance that we get different means is 99·993%, and that we get same mean is ~~more~~ nearly 0·007%.

DATA PRESENTATION:

The means of the samples (data sets) was presented graphically using a bar diagram. The numbers inside the bars represent the respective means. The error bars represent standard errors.

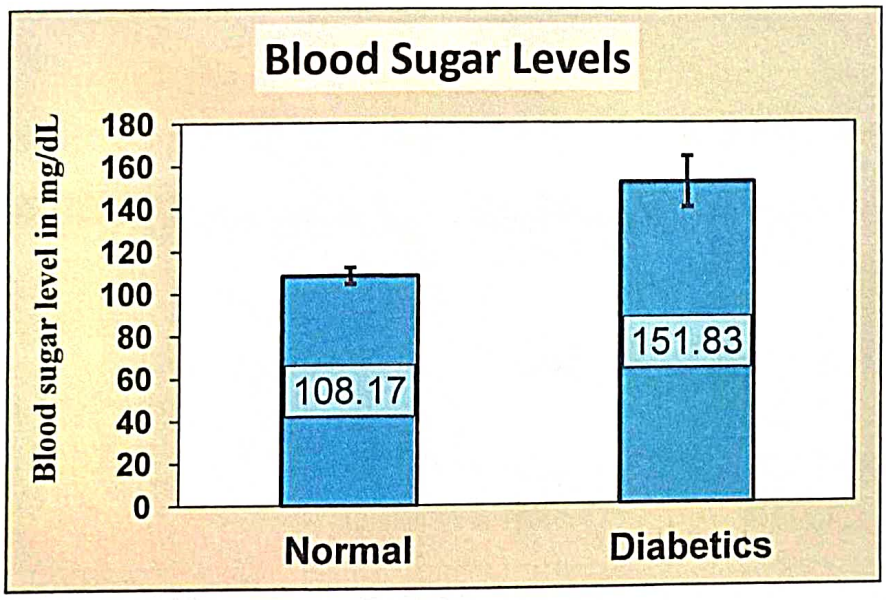

Fig: Blood sugar levels in normal and diabetic person.

18/12/21