

## **A Comparative Study on Statistical Solutions by Manual Calculation with Megastat-(An Excel Addin) Output**

**Dr.M.Thamizhsudar<sup>1\*</sup>, Dr.A.K.Bhuvaneswari<sup>2</sup>**

<sup>1,2</sup>Assistant Professors, Department of Mathematics,  
Aarupadai Veedu Institute of Technology,  
Vinayaka Mission's Research Foundation, TamilNadu, India

### **ABSTRACT**

The aim of this paper is to emphasize the importance of study of statistics which can be used as a tool for any kind of research studies, particularly to study the statistical calculations through MEGASTAT, which can be installed as add-in of Microsoft Excel 2007 which is specially designed for statistical calculations. Applications were illustrated with simple examples and the corresponding outputs and graphs were shown for easy understating.

### **Key words**

Statistical methods, examples, Megastat, output, Graphs

### **Introduction**

Statistics becomes an important component especially for analyzing and interpreting the data. Though the word statistics is defined by many authors, one of the best definitions given by "Croxtton and Cowden" and they define statistics is the science which deals with the Collection, Analysis and interpretation of numerical data.[5]. The study on statistics has wide applications in almost all sciences-Social as well as physical such as Biology, Psychology, education, Economics, Business management etc., For example, in diagnosing the correct disease the doctor has to rely heavily on actual data like temperature of the body, pulse rate, blood pressure. In fact, it is very difficult to find any branch of science without using statistical methods.

The quick changes in technological development suggest a need for new research on students' use of technology to solve real problem scenarios that use probability models [3]. The importance of proper study on Statistics in the field of Medical research given in [1], in which the five common statistical errors in medical manuscripts are identified. Moreover Statistics is an important tool to handle the vast data of present era as statistics can interpret all the information in such a beauty that so many conclusions can be extracted from it [8].

Statistics has two main basic types: descriptive statistics (set of methods to describe the collected data) and inferential statistics (set of methods to generalize, predict and decide by using information from a sample and inferring something about population). Collecting, organizing, and summarizing the data (discrete or continuous) comprise descriptive statistics, and making valid conclusions based on sample data is inferential statistics. Correlation, regression of various types, Chi-square test, t-test, and analysis of variance (ANOVA) are some simple statistical tests applied for inferences [7].

In this paper some of the fundamental statistical methods were discussed and few examples which were illustrated both manual calculation and using Megastat. For more detail on "Megastat" one can refer [6].

### **Graphical Representation**

Statistical results can be presented through diagrams and graphs. Though diagrams are more attractive than graphs, only the graphs are more helpful than diagrams for analyzing the results. Graphical presentation of data is an important part of research. By this way we can present the result in proper and effective way. Graphs and diagrams provide a visual method of examine qualitative and quantitative data [10]

Out of several methods, histogram is most widely used in practice. In particular, frequency polygon and ogives (cumulative frequency curves) are all more useful for comparing two or more frequencies. All these presentations

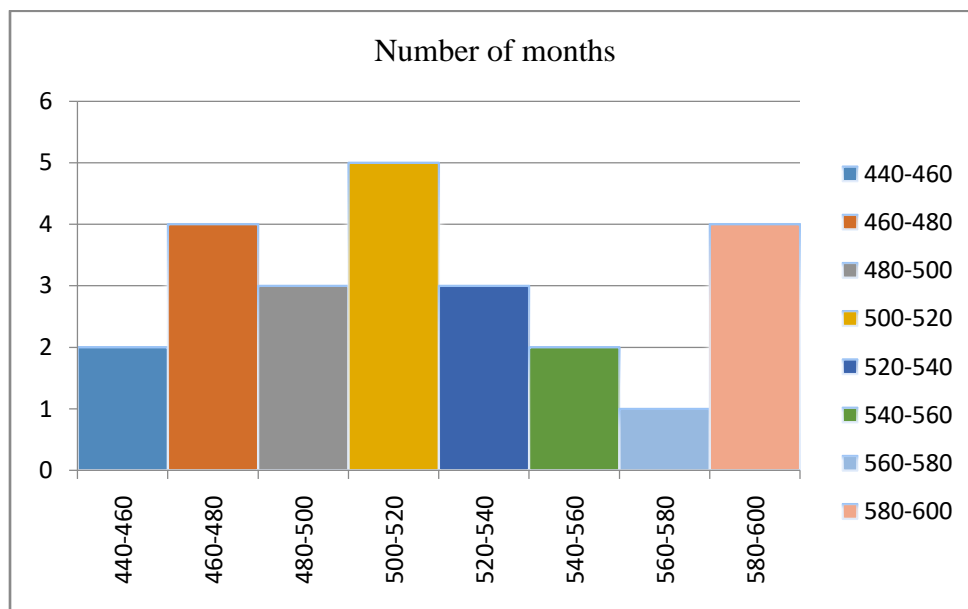
can be done easily with the help of Microsoft excel. For more study on Statistics with Microsoft excel, one can refer [4].

**Example: 1** Draw a histogram for the following data which shows the monthly cost of living index of a city in a period of 2 years .Problem taken from [2]

Cost of living index	Number of months
440-460	2
460-480	4
480-500	3
500-520	5
520-540	3
540-560	2
560-580	1
580-600	4
Total	24

One can easily obtain the required histogram from Excel.

Output from Excel



**Figure 1: Histogram**

## Statistical Inference

It is the branch of statistics commonly used for decision-making. Actually it refers to the process of selecting and using an appropriate sample statistic to draw an inference about a population parameter based on the sample which was drawn from the population. statistical inference deals two different problems Hypothesis testing and Estimation, the former one is to test some hypothesis about population from which the sample is drawn and the later one is to use the 'Statistics' obtained from the sample as estimate of the unknown parameter of the population from

which the sample is drawn. In both these cases the problem at hand is structured in such a way that inferences about relevant population values can be made from sample data.

### Procedure: Hypothesis Testing

#### Step 1:

The process starts with the assumption called a Hypothesis that too about the population parameter. There can be several types of hypothesis. Prof.MorrisHamburg described hypothesis as a quantitative statement about a population.

Setup a hypothesis: - Here one may have to set up two hypothesis namely Null and alternative Hypothesis. Null hypothesis is assumed in such a manner that there is no real difference in the sample and population about the particular thing which we considered for test. Example:

Suppose we want to test the effect of particular drug in curing malaria, we may take the null hypothesis as “the drug is not effective in curing Malaria” The rejection of null hypothesis indicates that the differences have statistical significance whereas the acceptance indicates the differences are due to chance. Since many practical problems are aimed to establish a significance of differences, the rejection of null hypothesis leads to a success in statistical project. An alternative hypothesis is the one against of null hypothesis. Null and alternative hypothesis are respectively denoted as  $H_0$  and  $H_a$  or  $H_1$ .

#### Step 2:

Rejecting or accepting null hypothesis is depend upon the level of significance adopted.

The significance level is expressed as a percentage (5% or 1%) is the probability of rejecting null hypothesis if it is true that is the risk of rejecting a true hypothesis in 5 out 100 occasions.

#### Step 3: Test Criterion:

Means applying most appropriate probability distributions. . ‘t’, ‘F’ and  $\chi^2$  are the standard test procedures for small samples and for large samples, normal distributions are used.

#### Step 4: Computation of test statistic:

Depending on the test criterion which is chosen for testing, statistic and standard error are calculated by using the appropriate formula. An example is illustrated for better understating.

**Step 5: Making decision**  
Inference has been drawn depending on whether the calculated value instep4 falls in rejection region or acceptance region.

#### Example: 2

Two types of drugs were used on 5 and 7 patients for reducing weight. Drug A was imported and drug B indigenous. The decrease in the weight after using these for six monyhs was as follows:

Drug-A	10	12	13	11	14		
Drug-B	8	9	12	14	15	10	9

Is there any significance difference in the efficiency of the two drugs? If not, which drug should you buy? (for  $v = 10, t_{0.05} = 2.223$ )

Solution:

Let us take the hypothesis that there is no significant difference in the efficiency of the two drugs. Applying t-test:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$\bar{X}_1 = \frac{\sum X_1}{n_1} = \frac{60}{5} = 12$$

$$\bar{X}_2 = \frac{\sum X_2}{n_2} = \frac{77}{7} = 11$$

$$S = \sqrt{\frac{\sum (X_1 - \bar{X}_1)^2 + \sum (X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{10 + 44}{5 + 7 - 2}} = 2.324$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{12 - 11}{2.324} \sqrt{\frac{5 \times 7}{5 + 7}} = 0.735$$

Calculated t value  $0.735 < \text{table value } 2.223$ , Hence the hypothesis is accepted. That is there is no significance in the efficacy of two drugs. Since drug B is indigenous and there is no significance in the efficacy of imported and indigenous drug, we should buy indigenous drug B.

The above problem can be easily illustrated with MEGASTAT- additional tool in EXCEL (can easily be downloaded) and the output is shown below.

**Table 1: Output – t value**

Group 1	Group 2			
12.00	11.00	mean		
1.58	2.71	std. dev.		
5	7	n		
	10	df		
	1.000	difference (Group 1 - Group 2)		
	5.400	pooled variance		
	2.324	pooled std. dev.		
	1.361	standard error of difference		
	0	hypothesized difference		
	0.73	t		
	.4793	p-value (two-tailed)		

In this output, t value =0.73 which is same as our calculated value 0.735.

## Correlation and Regression

The study of relationship between the variables can be studied by Correlation. By correlation, one can measure the degree of relationship between the existing variables. By Regression, we can estimate the value of one variable with the value of another where the variables should possess some kind of relationship between them. Actually Regression is a useful tool for engineers. It could help the engineer identify the relationship between variables based on test data. [11]

Using Megastat, one can easily find out the correlation coefficient along with p-value

### Example 3

The following table gives the distribution of items of production and also the relatively defective items along them, according to size groups. Find the correlation coefficient between sizes.

Size-Group	15-16	16-17	17-18	18-19	19-20	20-21
No. of Items	200	270	340	360	400	300
No. of Defective Items	150	162	170	180	180	114

**Solution**

The output of the problem using Megastat is given here.

**Table 2:** output- Regression and p value

Regression Analysis						
	$r^2$	0.267	n	6		
	r	0.517	k	1		
	Std. Error	23.902	Dep. Var.	Y		
ANOVA table						
Source	SS	df	MS	F	p-value	
Regression	832.0703	1	832.0703	1.46	.2940	
Residual	2,285.2630	4	571.3158			
Total	3,117.3333	5				
Regression output						
variables	coefficients	std. error	t (df=4)	p-value	95% lower	95% upper
Intercept	102.7937	47.8556	2.148	.0982	-30.0748	235.6621
X1	0.1814	0.1503	1.207	.2940	-0.2359	0.5988

## Probability Distribution

In statistics, probability distribution is the mathematical function which gives probabilities of different possible occurrence. [9]

### Example 4:

Consider the problem “If 10% of the screws produced by an automatic machine are defective, find the probability that out of 20 screws selected at random, there are (i) exactly 2 defective (ii) atleast 2 defective.

### Solution (Manual calculation)

This problem can be solved by applying Binomial distribution;

Given  $n=20$  and  $p=10\%=0.1$

Hence  $q = 1 - p = 0.9$

By Binomial distribution we have

$$p(X = x) = (n_{c_x} p^x q^{n-x}), \quad n = 0, 1, \dots, 20$$

$$(i) \quad P(X = 2) = (20_{c_2} (0.1)^2 q (0.9)^{20-2}) = 0.2$$

$$(ii) \quad P(X \geq 2) = 1 - P(X < 2) \\ = 1 - \{(20_{c_0} (0.1)^0 q (0.9)^{20}) + (20_{c_1} (0.1)^1 q (0.9)^{20-1})\}$$

= 0.6

It has been calculated that the probability of exactly two screws are defective is 20% and the probability of atleast two screws are defective is 60%.

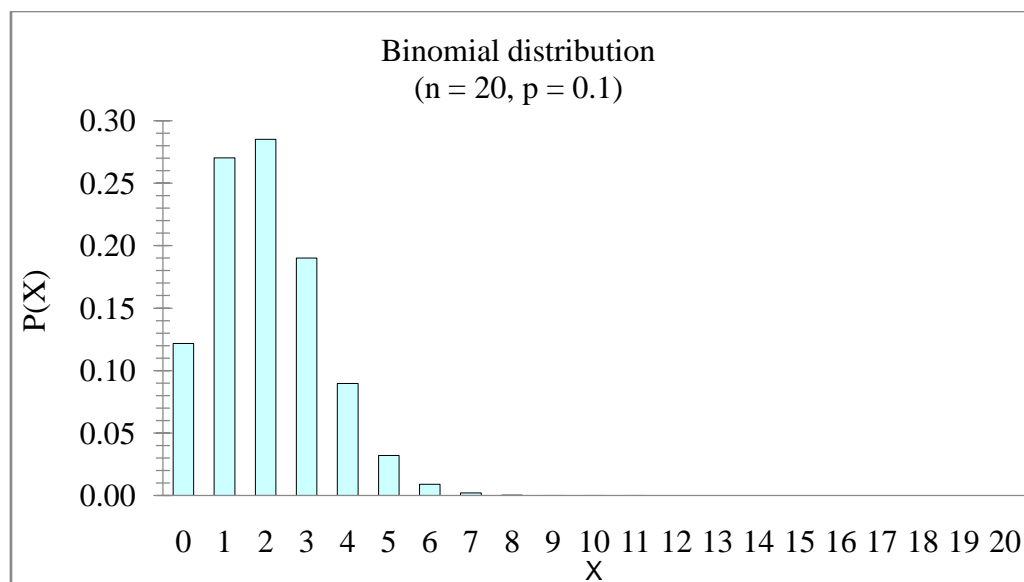
### MEGASTAT Calculation

The above problem can easily be solved for all values of x along with mean and variance through Megastat by giving inputs n and p. The output is shown below, and it is to be noted that the graph will be generated automatically.

#### Binomial distribution

20 n		
0.1 p		
X	P(X)	cumulative probability
0	0.12158	0.12158
1	0.27017	0.39175
2	0.28518	0.67693
3	0.19012	0.86705
4	0.08978	0.95683
5	0.03192	0.98875
6	0.00887	0.99761
7	0.00197	0.99958
8	0.00036	0.99994
9	0.00005	0.99999
10	0.00001	1.00000
11	0.00000	1.00000
12	0.00000	1.00000
13	0.00000	1.00000
14	0.00000	1.00000
15	0.00000	1.00000
16	0.00000	1.00000
17	0.00000	1.00000
18	0.00000	1.00000
19	0.00000	1.00000
20	0.00000	1.00000
1.00000		
2.000 expected value		
1.800 variance		
1.342 standard deviation		

**Figure 2:** Output-Binomial Distribution



**Figure 3:** Graphical output of Binomial Distribution

### Conclusion

As this paper provides some of the statistical methods with illustrated examples, a student can apply these in their field and particularly they can make use of Megastat which is an exclusive tool for statistical methods.

The future of the world Wide Web(WWW)will depend on the development of many new statistical ideas and algorithms apart from the need for Secure Cryptographic Schemes. Hence it is concluded that the statistical computations are becoming essential in almost every fields and those can be done effectively through a Microsoft Excel add in-Megastat .

### References

- [1] Alexander M. Strasak, QamruzZaman, Gerhard Marinell, Karl P. Pfeiffer, and Hanno Ulmer. (2007) *The American Statistician*, , Vol. 61, No. 14, 47-55
- [2] Arora,P.N., Arora,S. (2013) Statistics for Management. *S.Chand Publications, NewDelhi*.
- [3] Batanero C., Chernoff E.J., Engel J., Lee H.S., Sánchez E. (2016) Research on Teaching and Learning Probability. *Research on Teaching and Learning Probability*. ICME-13 Topical Surveys. Springer, Cham. [https://doi.org/10.1007/978-3-319-31625-3\\_1](https://doi.org/10.1007/978-3-319-31625-3_1)
- [4] DeCoursey W.J. (2003) Statistics and Probability for Engineering Applications with MicrosoftExcel*Elsevier Science (USA)*
- [5] Gupta,S.P. (2014) Statistical Methods. *Sultan Chand & Sons Educational Publishers, NewDelhi*,
- [6] [http://WWW.connectmath.com/Training/megastat\\_user\\_guide\\_pdf](http://WWW.connectmath.com/Training/megastat_user_guide_pdf)
- [7] Huma Zahir, Aisha Javaid, Rehana Rehman, Zahir Hussain. (2014). Statistical Concepts In Biology And Health Sciences. *J Ayub Med Coll Abbottabad*,26(1),95-97.
- [8] Iram Saleem, Muhammad Aslam and Muhammad Azam. (2013) The use of Statistical Methods in Mechanical Engineering. *Research Journal of Applied Sciences, Engineering and Technology*,5(7): 2327-2331
- [9] [Probability distribution - Wikipedia](#)
- [10] Swati Patel, Prakash Patel, RK Bansal, (2012) Some Basic Application of Bio-Statistics In Biomedical research. *National journal of community Medicine*, Vol.3, Issue 4, 748-749.
- [11] Wei Zhan, Rainer Fink, Alex Fan Texas A&M University, USA. (2010) Applications of Statistics in Engineering Technology Programs *American Journal of Engineering Education*, Volume 1 Number 1, 65-78.