Review Article [Sharma *et al.*, 5(7): July, 2014:3648-3650]
CODEN (USA): IJPLCP **ISSN: 0976-7126**

INTERNATIONAL JOURNAL OF PHARMACY & LIFE SCIENCES
(Int. J. of Pharm. Life Sci.)

Need of excel in Biostatics

Sonia Sharma[1*], Rachit Kumar[2], Hemlatha Vasishtha[3] and Deepankar Sharma[4]
1, 2, Sai Nath University, Ranchi - India
3, IIMT Institute of Engineering & Technology, Meerut IIMT, Meerut - India
4, D.J. College of Phamacy, Modinagar - India

Abstract

In Computer, Excel is probably the most commonly used spreadsheet. It is easily used to do variety of calculations, includes a collection of statistical functions, and a Data Analysis ToolPak. Excel is also widely used for the entry and management of data. Some points are given in this guide, but these topics are covered in more detail in a companion document, entitled "The Disciplined Use of Spreadsheets for Data Entry". As a result, if you suddenly find you need to do some statistical analysis, you may turn to it as the obvious choice. The current article deals with how well Excel would serve as a Data Analysis application.

Key-Words: Computer, Excel, Biostatics

Introduction

Statistics is an area that most newer statistician find difficult. The formulae are often complicated, the calculations tedious, degrees of freedom mysterious, and probability tables confusing. But in fact students need no longer grapple with any of these. In real life, statisticians rarely use calculation and tables these days, but instead use statistical packages such as Minitab or SPSS. But it isn't even necessary to buy an expensive statistics package, since spreadsheet soft-ware such as Excel has most of the common statistical tests built-in.

When using statistics, the first hurdle is to decide which statistical test to use. There are many other possible statistical tests, but this flow chart should be more than sufficient for users (Dytham, 1999).

It briefly summarises the Excel formulae and how to interpret the results, so it can be used as a handy guide on its own once the users is familiar with the tests. The present article describes in detail how to carry out these tests using Excel in biostatistics. It is divided into five sections.

*** Corresponding Author**
E-mail: s29sonia@gmail.com

S.No.	Types of Statistics	Parameters
1.	Descriptive statistics	mean, median, mode standard deviation, standard error, confidence interval
2.	Graphing data	scatter graphs, bar graphs error bars, lines
3.	Association statistics	P-coefficient, Spearman coefficient linear regression
4.	Comparative statistics	paired and unpaired t-test Mann-Whitney U-test ANOVA
5.	Frequency statistics	test test of association

Biostatistics

Biostatistics (or biometry) is the application of statistics to a wide range of topics in biology. The science of biostatistics encompasses the design of biological experiments, especially in medicine, pharmacy, agriculture and fishery; the collection, summarization, and analysis of data from those experiments; and the interpretation of, and inference from, the results. A major branch of this is medical biostatistics (Abhaya, 2012) which is exclusively concerned with medicine and health.

Scope and Application of Biostatistics

- Public health, including epidemiology, health services research, nutrition, environmental health and healthcare policy & management.
- Design and analysis of clinical trials in medicine.
- Population genetics, and statistical genetics in order to link variation in genotype with a variation in phenotype. This has been used in agriculture to improve crops and farm animals (animal breeding) (Helen et al., 2003).
- In biomedical research, this work can assist in finding candidates for gene alleles that can cause or influence predisposition to disease in human genetics (Jump up et al., 2003).
- Ecology, ecological forecasting
- Biological sequence analysis
- Systems biology for gene network inference or pathways analysis

Problems for working on Excel

Enable the Analysis ToolPak

The major problem is that the Data Analysis ToolPak is not installed with the standard Excel setup.for this please look in the Tools menu. If you do not have a Data Analysis item, you will need to install the Data Analysis tools. Search Help for "Data Analysis Tools" for instructions.

Missing Values

This is an important issue. For this a blank cell is the only way for Excel to deal with missing data. If you have any other missing value codes, you will need to change them to blanks.

Data Arrangement

Sometimes users not able to find the different analyses require the data to be arranged in various ways. If you plan on a variety of different tests, there may not be a single arrangement that will work. You will probably need to rearrange the data several ways to get everything you need.

Dialog Boxes

It has been observed that mostly users confused to choose tools/data analysis, and select the kind of analysis which he/she want to do. With the help of the following instruction he can get the proper tools.

Input Range: Type the upper left and lower right corner cells.

Labels - There is sometimes a box you can check off to indicate that the first row of your sheet contains labels.

Output location - New Sheet is the default. Or, type in the cell address of the upper left corner of where you want to place the output in the current sheet.

Biostatistics and Excel

Excel has a large range of statistical functions that are very useful. However before you use them make sure you understand what Excel is actually returning with each function. Summary statistics can be obtained directly from these functions or else from the Analysis Tool, available from the Tools menu.

The quickest way to get means and standard deviations for a entire group is using Descriptives in the Data Analysis tools. User can choose several adjacent columns for the Input Range (in this case the X and Y columns), and each column is analyzed separately. The labels in the first row are used to label the output, and the empty cells are ignored. If users have more, non-adjacent columns you need to analyze, you will have to repeat the process for each group of contiguous columns. The procedure is straightforward, can manage many columns reasonably efficiently, and empty cells are treated properly.

Important Parameters in Biostatistics

Excel can be used with confidence to obtain basic descriptive statistics, such as mean, median, mode, maximum, and minimum. All of these functions can be accessed through Excel's formula function (Dretzke, 2001)

Correlations

Using the Data Analysis tools, the dialog for correlations is much like the one for descriptives - you can choose several contiguous columns, and get an output matrix of all pairs of correlations. Empty cells are ignored appropriately. The output does NOT include the number of pairs of data points used to compute each correlation (which can vary, depending on where you have missing data), and does not indicate whether any of the correlations are statistically significant.

Two-Sample T-test

This test can be used to check whether the two treatment groups differ on the values of either X or Y. In order to do the test you need to enter a cell range for each group. Since the data were not entered by treatment group, we first need to sort the rows by treatment. Be sure to take all the other columns along with treatment, so that the data for each subject remains intact. After the data is sorted, you can enter the range of cells containing the X measurements for each treatment. Do not include the row with the labels, because the second group does not have a label row. Therefore your output will not be labeled to indicate that this output is for X. If you want the output labeled, you have to copy the cells corresponding to the second group to a separate column, and enter a row with a label for the second group. If you also want to do the t-

test for the Y measurements, you will need to repeat the process. The empty cells are ignored, and other than the problems with labeling the output, the results are correct.

Paired t-test

The paired t-test is a method for testing whether the difference between two measurements on the same subject is significantly different from 0. In this example, we wish to test the difference between X and Y measured on the same subject. The important feature of this test is that it compares the measurements within each subject. If you scan the X and Y columns separately, they do not look obviously different.

One-way ANOVA

Data must be arranged in separate and adjacent columns (or rows) for each group. Clearly, this is not conducive to doing 1-ways on more than one grouping. If you have labels in row 1, the output will use the labels.

Two-Factor ANOVA without Replication

This only does the case with one observation per cell (i.e. no Within Cell error term). The input range is a rectangular arrangement of cells, with rows representing levels of one factor, columns the levels of the other factor, and the cell contents the one value in that cell.

Two-Factor ANOVA with Replicates

This does a two-way ANOVA with equal cell sizes. Input must be a rectangular region with columns representing the levels of one factor, and rows representing replicates within levels of the other factor. The input range MUST also include an additional row at the top, and column on the left, with labels indicating the factors. However, these labels are not used to label the resulting ANOVA table. Click Help on the ANOVA dialog for a picture of what the input range must look like.

Probability Distributions

Excel's probability functions include all that would normally be found in a simple set of statistical tables.

Conclusion

Microsoft Excel is spreadsheet software that is used to store information in columns and rows, which can then be organized and/or processed. Excel is a powerful program with an intuitive user interface, and can be a great option for entering, organizing, and cleaning data. In addition to its spreadsheet functions, Excel provides a number of standard statistical andgraphing procedures. However, these should be approached with caution, as statisticians have found numerous errors in Excel's statistical routines and distributions.

References

1. Abhaya Indrayan (2012). *Medical Biostatistics*. CRC Press. ISBN 978-1-4398-8414-0.
2. Helen Causton, John Quackenbush and Alvis Brazma (2003). Statistical Analysis of Gene Expression Microarray Data. Wiley-Blackwell.
3. Jump up Terry Speed (2003). Microarray Gene Expression Data Analysis: A Beginner's Guide. Chapman & Hall/CRC.
4. Frank Emmert-Streib and Matthias Dehmer (2010). *Medical Biostatistics for Complex Diseases*. Wiley-Blackwell. ISBN 3-527-32585-9.
5. Jump up Warren J. Ewens and Gregory R. Grant (2004). *Statistical Methods in Bioinformatics: An Introduction*. Springer.
6. Jump up Matthias Dehmer, Frank Emmert-Streib, Armin Graber and Armindo Salvador (2011). *Applied Statistics for Network Biology: Methods in Systems Biology*. Wiley-Blackwell. ISBN 3-527-32750-9.
7. Dretzke, B. J. (2001) Statistics with Excel, 2nd ed. New York: Prentice Hall.
8. Dytham, C. (1999) Choosing and using statistics. A biologist's guide. Oxford: Blackwell Science.

How to cite this article

Sharma S., Kumar R., Vasishtha H. and Sharma D. (2014). Need of excel in Biostatics. *Int. J. Pharm. Life Sci.*, 5(7):3648-3650.

Source of Support: Nil; Conflict of Interest: None declared

Received: 25.06.14; Revised: 30.06.14; Accepted:05.07.14