IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

COMPARISON BETWEEN t-TEST AND F-TEST WITH EQUAL AND UNEQUAL VARIANCES USING EXCEL

Dr.Ch.Prasuna,
Assistant Professor on Contract
Department Of Statistics
Vikrama Simhapuri University

Abstract

The user of excel for performing t-test faces two sample t-test with two options viz.,Two sample with equal variances and Two sample with unequal variances. Non statistics user often selects one of the two methods without any basis. The correct method is to perform F-test for equality of variances before doing t-test. This project is aimed at stimulating a large number of t-tests under both methods and estimate the chance of committing an error. We are motivated to exploit Random Generation tools in Excel.

1. To stress the importance of homogeneity of variances in conducting t-test.
2. To assess the behaviour of the two-tailed t-value for independent samples with and without assuming equal variances
3. To simulate two independent normal populations and examine the behaviour of t-statistic by taking resampling.

Keywords: Ms-Excel, data analysis pack, t-test

1. ABOUT MS –EXCEL

Microsoft Excel is a spreadsheet developed by Microsoft for Windows, Mac OS X, Android and iOS. It features calculation, graphing tools, pivot tables and a macro programming language called Visual Basic for Applications. It has been a very widely applied spreadsheet for these platforms, especially since version 5 in 1993, and it has replaced by Lotus 1-2-3 as the industry standard for spreadsheets. Excel forms a part of Microsoft Office. By default, Excel will open a blank workbook that contains three worksheets (spreadsheets). Each box, located in both a column and a row, is called a cell.

Spread sheet

A type of application program which manipulates the numerical and string data in rows and columns of cells. The value in a cell can be calculated from a formula which can involve other cells. A value is recalculated automatically whenever a value on which it depends changes. Different cells may be displayed with different formats.

Workbook

- A booklet containing problems and exercises that a student may work directly on the pages.
- A manual containing operating instructions, as for an appliance or machine.
- A book in which a record is kept of work proposed or accomplished.

Worksheet

- A sheet of paper with multiple columns; used by an accountant to assemble figures for financial statements.
- A piece of paper recording work planned or done on a project.

Menu bar

The Menu Bar is directly below the Title bar and displays the menu. The menu begins with the word File and continues with the following:

Edit, View, Insert, Format, Tools, Data, Window, and Help.

You use the menu to give instructions to the software. Point with your mouse to a menu option and click the left mouse button. A drop-down menu will appear. You can now use the left and right arrow keys on your keyboard to move left and right across the Menu bar options. You can use the up and down arrow keys to move up and down the drop-down menu. To select an option, highlight the item on the drop-down menu and press Enter.

Format cells

Cells can be formatted to help handle various types of data. Right click on a single cell, or a group of cells, and select "Format Cells" from the drop down menu. Brief descriptions of format types can be seen at the bottom of the dialog box. Take a moment to look through the various formatting options. Clicks cancel when you have done.

2. ADVANTAGES OF EXCEL

- It is very important to know that Excel has the ability to organize large amounts of data into orderly spreadsheets and charts quickly. It is easy to enter and format the data. It has the ability to create the graphical or the visual representations of your data, and it is easy to integrate Excel with other business applications.
- It is very important to know that it allows you to let others view a vast amount of data by letting you to send it via email or print it out just as it appears on the computer. Excel can be sent through email and viewed by most smart phones which makes more convenient.
- You have to know that you can make easy, effective comparisons with the powerful analytical tools included within Microsoft Excel. You have the ability to analyze large amounts of data to discover trends and patterns that

will influence decisions. Microsoft Excel's graphing capabilities allows you to summarize your data enhancing your ability to organize and structure of your data.

3. DISADVANTAGES OF EXCEL

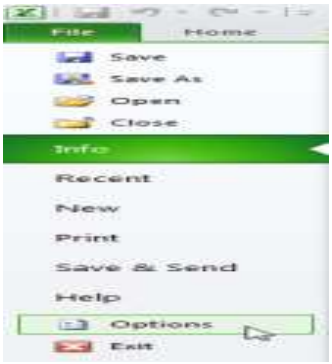
- It is very important to know that the viruses can be attached to an Excel file through macros which are mini programs that are written into an Excel spread sheet. Using only one file can make the file size very big and the program run slowly. So you might have to break it into smaller files, so that there is an increased risk in Excel data being lost.
- You should know that while the spreadsheets are ideal for creating one time analysis, they become problematic as the data grows and evolves over time. As new rows and columns get added, the summary ranges and formulas may need to be modified or new ones created, the data and the formulas are not consistently updated and these mistakes lead to bad results and decisions.
- It is very important to know that entering the data into Excel manually can take a very long time especially if you have a lot of data to enter. The amount of time it takes to enter the data can be inefficient and can lead to boredom which leads to potentially costly inattentiveness.

4. INSTALLATION OF DATA ANALYSIS PAK IN MS- EXCEL

The Analysis Tool Pak is an Excel Add-ins program that provides Data Analysis tools for Financial, Statistical and Engineering data analysis.

To load the Analysis Tool Pak Add-ins, execute the following steps.

1. Click on the green File tab. The File tab in Excel 2010 replaces the Office Button (or File Menu) in previous versions of Excel.
2. Click on Options.



3. Under Add-ins, select Analysis Tool Pak and click on the Go button.

4. Check Analysis Tool Pak and click on OK.

5. On the Data tab, you can now click on Data Analysis.

The following dialog box below appears.

6. For example, select Random Number Generation and click OK to generate random numbers in Excel.

5. STATISTICAL TOOLS USED UNDER STUDY

Population

Successful statistical practice is based on focused problem definition. In sampling, this includes defining the population from which our sample is drawn. A population can be defined as including all people or items with the characteristic one wish to understand. Because there is very rarely enough time or money to gather information from everyone or everything in a population, the goal becomes finding a representative sample (or subset) of that population.

Sampling

Sampling is concerned with the selection of a subset of individuals from within a statistical population to estimate characteristics of the whole population. Each observation measures one or more properties (such as weight, location, color) of observable bodies distinguished by an independent objects or individuals. In survey sampling, weights can be applied to the data to adjust for the sample design, particularly stratified sampling. Results from probability theory and statistical theory are employed to guide the practice. In business and medical research, sampling is widely used for gathering information about a population.

Random number generation

When a die is rolled, you get a random number between 1 and 6.

A random number generator (RNG) is a computational or physical device designed to generate a sequence of numbers or symbols that cannot be reasonably predicted better than by a random chance.

Various applications of randomness have led to the development of several different methods for generating random data, of which some have existed since ancient times, including dice, coin flipping, the shuffling of playing cards and many other techniques. Because of the mechanical nature of these techniques, generating large numbers of sufficiently random numbers (important in statistics) required a lot of work and/or time. Thus, results would sometimes be collected and distributed as random number tables. Nowadays, after the advent of computational random-number generators, a growing number of government-run lotteries and lottery games have started using RNGs instead of more traditional drawing-methods. RNGs are also used to determine the odds of modern slot machines.

Several computational methods for random number generation exist. Many fall short of the goal of true randomness, although they may meet, with varying success, some of the statistical tests for randomness intended to measure how unpredictable their results are (i.e, to what degree their patterns are discernible). However, carefully designed cryptographically secure computationally based methods of generating random numbers also exist, such as those based on the Yarrow algorithm, the Fortuna (PRNG), and others.

6. t-test

Definition

A t-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis. It can be used to determine if two sets of data are significantly different from each other, and is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistic (under certain conditions) follows a Student's t distribution.

t-Test: Two-Sample Assuming Equal Variances

This analysis tool performs a two-sample student's t-test. This t-test form assumes that the two data sets came from distributions with the same variances. It is referred to as a homoscedastic t-test. You can use this t-test to determine whether the two samples are likely to have come from distributions with equal population means.

The following formula is used to determine the statistic value *t*.

$$\frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

t-Test: Two-Sample Assuming Unequal Variances

This analysis tool performs a two-sample student's t-test. This t-test form assumes that the two data sets came from distributions with unequal variances. It is referred to as a heteroscedastic t-test. As with the preceding Equal Variances case, you can use this t-test to determine whether the two samples are likely to have come from distributions with equal population means. Use this test when there are distinct subjects in the two samples. Use the Paired test, described in the follow example, when there is a single set of subjects and the two samples represent measurements for each subject before and after a treatment.

The following formula is used to determine the statistic value *t*.

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

7. F-test

While using the t-test for comparing two means, we have assumed that the variances of the populations are equal. This assumption can be verified by testing the hypothesis $H0.\sigma_1^2 = \sigma_2^2$. This test is called the *variance ratio test.* Excel has a module in the Data Analysis Pak to do this job. If n_1, n_2 denote the sizes of two samples respectively and S_1, S_2 denote the corresponding estimates of the population variances based on the two samples, then the F-statistic is given by

$$F = S_1^2/S_2^2.$$

This statistics follows a distribution called Snedecor's F-distribution with (n_1 - 1, n_2 – 2) DOF. It is assumed that the populations from which the two samples have been drawn are normally distributed.

For the one-sided alternative hypothesis $H_1.\sigma_1^2 > \sigma_2^2$, the critical value is denoted by F_α. We reject H_0 and otherwise we accept H_1, if the calculated F value (F_{cal}) exceeds F_α.

If the alternative is $H_1 : \sigma_1^2 < \sigma_2^2$, the critical value is $F_{1-\alpha}$ with (n_2 - 1,n_1 – 1) DOF. We accept H_1 if $F_{cal} < F_{1-\alpha}$. The table of critical values are usually given at $\alpha = 0.05$ and $\alpha = 0.01$ levels. The DOF of the numerator in F is shown on the top of the statistical tables and the DOF of the denominator is shown along the rows. As a matter of convenience, while conducting this test, we keep the larger variance in the numerator so that the calculated value is always ≥1.

8. CONCLUSIONS

Student's t-test may be used

i. To test the significant difference between sample arithmetic mean and population arithmetic means.
ii. To test the significant difference between two sample means.
iii. To test the significant of the given treatment
iv. To test the significance of small sample correlation coefficient.
v. To test the significance of sample growth rate and so on.

The main use of F-distribution is to test whether two independent samples have been drawn for the normal populations with the variance, or if two independent estimates of the population variance are homogenous or not, since it is often desirable to compare two variances rather than two averages.

vi. To test the significant difference between two sample variances.
vii. To test the homogeneity of several sample means.
viii. To test the significance of multiple correlation coefficient.
ix. To test the linearity of regression and so on.

REFERENCES

- K.V.S. Sarma (2010), Statistics Made Simple, Do It Yourself on PC, Prentice Hall of India.
- Web resources on distributions.
- Excel help line.
- S.C.Gupta and V.K. Kapoor (2006), Fundamentals of Mathematical Statistics, Sulthan chand & sons

International Research Journal
IJNRD
Research Through Innovation