

DATA ANALYTICS FOR CORPORATE EMPLOYABILITY

M. Arun Shankar¹ & V. Ramanathan²

^{1,2}MBA, Anna University, Madurai

Abstract

Capture of large scale data from customers, social media and mobile gadgets is easily possible today. But analyzing this voluminous data in a short time span was posing a threat to the practical utility of such data for decision makers. Data analytics software and application packages makes this possible now and organizations can get information in a jiffy regarding customer behavior, purchase patterns, media habits etc. to form competitive strategies. This article explains the concepts of Big Data analytics and how this skill can become a competitive advantage for an engineer or MBA graduate at the entry level or advancement of career with an organization. It also illustrates certain typical situations where Big Data analytics can be used by an organization in its functional departments.

Key words: Data analytics, employability, hypotheses, t-tests, SPSS applications

Introduction

Big data analytics is the process of examining large data sets containing a variety of data types i.e., big data to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits.

The primary goal of big data analytics is to help companies make more informed business decisions by enabling data scientists, predictive modelers and other analytics professionals to analyze large volumes of transaction data, as well as other forms of data that may be untapped by conventional business intelligence (BI) programs. That could include Web server logs and Internet click-stream data, social media content and social network activity reports, text from customer emails and survey responses, mobile-phone call detail records and machine data captured by sensors connected to the Internet. Big data can be analyzed with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics, data mining, text analytics and statistical analysis. Mainstream BI software and data visualization tools can also play a role in the analysis process. As a starting point, the fresh graduates should have good knowledge about IBM SPSS software for statistical data analysis. This article illustrates the data analytics applications using SPSS only.

SPSS Data Analytics Skills for Beginners

SPSS is a Windows based program that can be used to perform data entry and analysis and to create tables and graphs. SPSS is capable of handling large amounts of data

and can perform all of the analyses covered in the text and much more. SPSS is commonly used in the Social Sciences and in the business world, so familiarity with this program should serve you well in the future.

On its surface, SPSS looks a lot like a typical spreadsheet application. When you open it, you see the familiar tabular grid and can enter values in cells. Spreadsheets, on the other hand, are capable of a lot of things that SPSS is good at, like generating graphs and statistics on a data set. The difference can be summed up by saying that spreadsheets are designed to be very flexible and broadly applicable to many different tasks, while SPSS was designed specifically for statistical processing of large amounts of data at an enterprise level. For example, unlike a spreadsheet, SPSS has the concepts of “case” and “variable” built in. The rows in SPSS always represent cases, for example survey responses or experimental subjects, and the columns always represent variables observed from those cases, like the specific values given by the survey respondent or measurements from the experimental subject. Because of this case/variable arrangement, when a calculation is performed over a set of data, the result does not get inserted into another cell on the table, like it would in a typical spreadsheet, but appears in a separate window. This is particularly advantageous when dealing with large sets of data, since it keeps calculated statistics and graphs separate from the raw data but still easily accessible.

It is also much more convenient to perform statistical tests in SPSS, even though many are possible using typical spreadsheets. For example, to perform a one-sample T-test with Excel, you’ll have to calculate the T value independently for the sample and use the “T.DIST” function to return the significance, while also selecting a cell for the results and labeling it in another cell. To perform the same test in SPSS you select a variable and supply the value to compare with your sample and, when you click “Ok,” SPSS generates a table with t, the degrees of freedom, the significance, and a confidence interval neatly calculated.

Research Situations where SPSS Data Analysis can be Employed

A few organizational or business situations are illustrated below, where use of data analytical techniques using SPSS can be implemented. This gives an idea to the new users of SPSS.

1. A researcher has information that the satisfaction level of customers has a mean value of say, 3.4 (on a scale of 1 to 5) from a previous study. He is interested to check if the satisfaction has changed at present. For this he takes a sample of customers now and measures the current satisfaction level as 3.7 (again on a 1-5 scale). Now the dilemma is, does the difference in the mean values of 3.4 and 3.7 (between the sample mean and known population mean) a really significant change or it can be attributed to the errors in the survey? In such a situation, One sample t- test can be used.

2. Suppose a researcher wants to test if male and female groups differ in customer satisfaction for a product. He has to calculate the means of the two groups separately and compare them. If he gets the mean satisfaction values for male as 3.5 and for female as 3.7, here again, the question is do the two group means significantly differ or it is due the chance error in the survey. Independent sample t-test (for 2 groups) is employed here.
3. A third situation is, a researcher wants to test if the performance of a student has changed after a training program. Or the patient's condition improves after a treatment. In other words, he wants to compare the mean scores before and after the treatment. In such situation, a paired sample t-test is used.
4. When a researcher wants to compare the means of more than 2 groups, ANOVA test is used. For example, he wants to study if the customer satisfaction is same or different among four mobile phone networks. In other words, is the difference in mean scores of satisfaction among the 4 groups significant?

Setting Hypothesis

For case 1 above, the Null Hypothesis will be, there is no difference between the sample mean and the known population mean. For case 2, the Null Hypothesis will be there is no significant difference between the means of Male and Female customers. For case 3, the Null Hypothesis will be, there is no significant difference in means before and after the treatment. For case 4, the Null Hypothesis will be, there is no significant difference between the means of all groups. Statistical tests corresponding to case 1 is, one sample t-test, for case 2 it is independent sample groups test, for case 3 it is paired sample t test and for case 4 it is one-way ANOVA test using SPSS. How the test is conducted and inference is drawn is illustrated for case 1.

One sample t-test using SPSS

In SPSS tool bar, select Analyse, Compare means, one sample t-test. Select the variable from the variables box and shift it in to test variable box. Enter test value in the test value box. Test value is the mean or known population mean for comparison. Click OK. SPSS generates an output table for one sample t test.

One-Sample Test						
	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Avg monthly minutes	53.706	249	.000	158.18562	152.3846	163.9867

If the significance value is less than the confidence level 0.05, Reject Null Hypothesis. So we are rejecting “there is no difference between test value and sample mean”. In other words, we are accepting that there is significant difference between the test value and sample mean. In case1, we are concluding that there is significant difference in satisfaction levels between previous study and present study. While analyzing using SPSS, the significance value or p-value for each test is directly given in the output based on which decision to accept or reject the null hypothesis can be taken. There is no need to compare the value of any test statistic with a table value, as we do in Chi square tests.

Conclusion

Employability of fresh graduates will improve to a great extent if they master the skill of data analytics using standard software like SPSS. The salary levels offered to such candidates who can analyse huge data and draw conclusions regarding business decisions will be also high. For teachers and skill trainers this should be an objective while designing placement training for the final year graduates. SPSS training can be given without much expense, in-house.

References

1. Sanjose University, School of Social work (notes) downloaded from <http://www.sjsu.edu/people/peter.a.lee/courses/ScWk242/s1/t%20Test%20Notes.pdf> on 26/12/2014.
2. McDonald, J.H., (2008). Handbook of Biological Statistics, Sparky House Publishing, Baltimore.
3. Bluman, Allan G. (2008). Elementary Statistics: A Step by Step Approach, A Brief Version, 4th ed. New York: McGraw Hill.
4. SPSS. 2007. SPSS 15.0 Syntax Reference Guide. Chicago, IL: SPSS Inc.