TECHNOVISION
Look Beyond the Obvious….

1119 Raritan Rd. Suite 2
Clark, NJ 07066
732 381 0200
www.qavision.com

Software Quality Assurance Methodology Series:

Test Case Design Using The Taguchi Method

Table of Contents

ABSTRACT

One of the major challenges of any software testing effort is to determine how many test cases does the system require to achieve the quality acceptance level. Considering today's distributed environments, advanced component level software development and the need to provide grater functionality, software products are getting inherently complex. Hence, generating more variables and more levels for those variable to be tested.

This document suggests the adaptation of a well know scientific method of Taguchi Experimental Design to the software testing industry to aid in determining the number of test cases to achieve optimum level coverage and quality.

The Taguchi method of design of experiments is a statistical tool based on the systematic approach of conducting minimal number of experiments using a mathematical instrument called orthogonal arrays. Traditionally, the method has been used to predict the significance of contribution of each design variables and their level to achieve optimum combination by conducting a real time experiment.

1. Background

The technique of laying out the conditions of experiments [6] involving multiple factors was first proposed by, Sir R.A.Fisher. The method is popularly known as the factorial design of experiments. A full factorial design will identify all possible combinations for a given set of factors. Since most industrial experiments usually involve a significant number of factors, a full factorial design results in a large number of experiments.

One can consider a simple logon window component of a software application. If the fields "username" and "password" require a minimum of 7 numeric characters each, than the full factorial design of variables for this component would require, 2×10^7, a total of 20,000,000 test cases. This result is practically impossible for a software QA project to achieve.

Through careful advanced planning experiments can be designed which require minimal number of test cases and still provide estimates of most important interaction effects.

To reduce the number of experiments to a practical level, only a small set from all the possibilities is selected. The method of selecting a limited number of experiments which produces the most information is known as a partial fraction experiment. Although this method is well known, there are no general guidelines for its application or the analysis of the results obtained by performing the experiments. Taguchi constructed a special set of general design guidelines for factorial experiments that cover many applications.

2. Basic concepts

2.1. Definition

Taguchi has envisaged a new method of conducting the design of experiments which are based on well defined guidelines. This method uses a special set of arrays called orthogonal arrays. These standard arrays stipulates the way of conducting the minimal number of experiments which could give the full information of all the factors that affect the performance parameter. The crux of the orthogonal arrays lies in choosing the level combinations of the design variables for each test input.

2.2. A typical orthogonal array

While there are many standard orthogonal arrays available, each of the arrays is meant for a specific number of independent design variables and levels . For example, if one wants to conduct an experiment to understand the influence of 4 different independent variables with each variable having 3 set values (level values), then an L9 orthogonal array might be the right choice. The L9 orthogonal array is meant for understanding the effect of 4 independent factors each having 3 factor level values. This array assumes that there is no interaction between any two factor. While in many cases, no interaction model assumption is valid, there are some cases where there is a clear evidence of interaction. A typical case of interaction might be the interaction between the price and volume of the purchase of a stock.

In our example of the "login" window, we can consider that there are 7 different variables in each field (field length) and 10 levels of each variable. However, knowing that in a second or third generation language application or object oriented approach, these fields will be handled in the same class for parsing the input. Therefore a subjective selection approach shall be used. This empirical (subjective) selection method comes from software testing experience. Industry approved selection criteria, which we will be adopting in our methodology uses:

1. Valid Data (data comes from normally occurring actions)
    a. Upper or lover acceptable level of numeric values
    b. Acceptable sequence of alphabetical characters
    c. Special alphanumeric characters where applicable (i.e. $ sign for currency, % sign for APR.)
2. Invalid data
    a. Entering alphabetical characters instead of numeric or vice versa
    b. Using data that is invalid by business description (i.e. invalid account number)
    c. Using incomplete or extraneous data, or omitting the field entirely

d. Using data that violates the limits established by the standards.

$L_9 (3^4)$ Orthogonal array					
	Independent Variables				Performance Parameter Value
Experiment #	Variable 1	Variable 2	Variable 3	Variable 4	
1	1	1	1	1	p1
2	1	2	2	2	p2
3	1	3	3	3	p3
4	2	1	2	3	p4
5	2	2	3	1	p5
6	2	3	1	2	p6
7	3	1	3	2	p7
8	3	2	1	3	p8
9	3	3	2	1	p9

Table 1 Layout of L_9 orthogonal array.

The Table 1 shows an L_9 orthogonal array. There are totally 9 experiments to be conducted and each experiment is based on the combination of level values as shown in the table. For example, the third experiment is conducted by keeping the independent design variable 1 at level 1, variable 2 at level 3, variable 3 at level 3, and variable 4 at level 3.

2.3. Properties of an orthogonal array

The orthogonal arrays has the following special properties that reduces the number of experiments to be conducted.

1. The vertical column under each independent variables of the above table has a special combination of level settings. All the level settings appears an equal number of times. For L9 array under variable 4 , level 1 , level 2 and level 3 appears thrice. This is called the balancing property of orthogonal arrays.
2. All the level values of independent variables are used for conducting the experiments.
3. The sequence of level values for conducting the experiments shall not be changed. This means one can not conduct experiment 1 with variable 1, level 2 setup and experiment 4 with variable 1 , level 1 setup. The reason for this is that the array of each factor columns are mutually orthogonal to any other column of level values. The inner product of vectors corresponding to weights is zero. If the above 3 levels are normalized between -1 and 1, then the weighing factors for level 1, level 2 , level 3 are -1 , 0 , 1 respectively. Hence the inner product of weighing factors of independent variable 1 and independent variable 3 would be

$$((-1\times-1)+(-1\times0)+(-1\times1))+((0\times0)+(0\times1)+(0\times-1))+((1\times0)+(1\times1)+(1\times-1))=0$$

2.4. Minimum number of experiments to be conducted

The design of experiments using the orthogonal array is, in most cases, efficient when compared to many other statistical designs. The minimum number of experiments that are required to conduct the Taguchi method can be calculated based on the degrees of freedom approach ($N_{Taguchi}$).

$$N_{Taguchi} = 1 + \sum_{i=1}^{NV} (L_i - 1) \qquad \text{Equation 2.4}$$

For example, in case of 8 independent variables study having 1 independent variable with 2 levels and remaining 7 independent variables with 3 levels (L18 orthogonal array) , the minimum number of experiments required based on the above equation is 16. Because of the balancing property of the orthogonal arrays, the total number of experiments shall be multiple of 2 and 3. Hence the number of experiments for the above case is 18.

The aim here is to investigate not only the effects of the individual variables (or factors) on the outcome, but also how the variables interact. Obviously, even with a moderate number of factors and a small number of levels for each factor, the number of possible level combinations for the factors increases rapidly. It may therefore not be feasible to make even one observation at each of the level combinations. In such cases observations are made at only some of the level combinations, and the purpose of the orthogonal array is to specify which level combinations are to be used. Such experiments are called "fractional factorial" experiments. While there are nowadays other applications of orthogonal arrays in statistics (for example in computer experiments and survey sampling), the principal application is in the selection of level combinations for fractional factorial experiments.

Applying this method to our "login" window and using the software test data selection criteria for valid data, we will consider:

1. "Username" as a single (1) independent variable with 2 levels (no data or a string length of 7 alphabetical characters)..
2. "Password" as 7 independent variable with 3 levels (valid boundary values and a random valid middle value. i.e. 0, 5, 9).
3. Using orthogonal array properties, we will construct a test suite with 18 test cases.

The resulting test suite could be represented in table below (Table 2):

$L_{18}(2^1/3^7)$ Orthogonal array								
	Independent Variables							
Test Case #	U_name	Pwd_1	Pwd_2	Pwd_3	Pwd_4	Pwd_5	Pwd_6	Pwd_7
1	1	0	0	0	0	0	0	0
2	0	1	1	1	1	1	1	0
3	1	2	2	2	2	2	2	0
4	1	0	0	1	2	1	2	0
5	1	1	1	2	0	2	0	0
6	1	2	2	0	1	0	1	0
7	1	0	1	0	2	2	1	1
8	1	1	2	1	0	0	2	1
9	1	2	0	2	1	1	0	1
10	1	0	2	2	0	1	1	1
11	1	1	0	0	1	2	2	1
12	1	2	1	1	2	0	0	1
13	1	0	1	2	1	0	2	2
14	1	1	2	0	2	1	0	2
15	1	2	0	1	0	2	1	2
16	1	0	2	1	1	2	0	2
17	1	1	0	2	2	0	1	2
18	1	2	1	0	0	1	2	2

Table 2 Layout of L_{18} orthogonal array - Addelman-Kempthorne construction.

U_name represents the two levels of the "username" field.
U_name 1=7character string, 0= no input.
As a logical approach, no input for either "username" or "password" should generate an error message.

Pwd_1 -Pwd_7 represent the seven numeric digits of the "password" field.
0 Value = 0 lower boundary and 3 Value = 9 upper boundary levels
1 Value = 5 valid data between the boundaries

As a result, the Taguchi method of designing test cases using an orthogonal arrays, software testing can achieve effective testing in 18 test cases instead of the impossible 20 Million alternative.

3. Assumptions of the Taguchi method

The additive assumption implies that the individual or main effects of the independent variables on performance parameter are separable. Under this assumption, the effect of each factor can be linear, quadratic or of higher order, but the model assumes that there exists no cross product effects (interactions) among the individual factors. That means the effect of independent variable 1 on performance parameter does not depend on the different level settings of any other independent variables and vice versa. If at anytime, this assumption is violated, then the additivity of the main effects does not hold, and the variables interact.

4. Designing an experiment

The design of an experiment involves the following steps

1. Selection of independent variables
2. Selection of number of level settings for each independent variable
3. Selection of orthogonal array
4. Assigning the independent variables to each column
5. Conducting the experiments
6. Analyzing the data
7. Inference

4.1. Selection of the independent variables

Before conducting the experiment, the knowledge of the product/process under investigation is of prime importance for identifying the factors likely to influence the outcome. In order to compile a comprehensive list of factors, the input to the experiment is generally obtained from all the people involved in the project.

4.2. Deciding the number of levels

Once the independent variables are decided, the number of levels for each variable is decided. The selection of number of levels depends on how the performance parameter is affected due to different level settings. If the performance parameter is a linear function of the independent variable, then the number of level setting shall be 2. However, if the independent variable is not linearly related, then one could go for 3, 4 or higher levels depending on whether the relationship is quadratic, cubic or higher order.
In the absence of exact nature of relationship between the independent variable and the performance parameter, one could choose 2 level settings. After analyzing the experimental data, one can decide whether the assumption of level setting is right or not based on the percent contribution and the error calculations.

4.3. Selection of an orthogonal array

Before selecting the orthogonal array, the minimum number of experiments to be conducted shall be fixed based on the total number of degrees of freedom [5] present in the study. The minimum number of experiments that must be run to study the factors shall be more than the total degrees of freedom available. In counting the total degrees of freedom the investigator commits 1 degree of freedom to the overall mean of the response under study. The number of degrees of freedom associated with each factor under study equals one less than the number of levels available for that factor. Hence the total degrees of freedom without interaction effect is 1 + as already given by equation 2.4. For example, in case of 11 independent variables, each having 2 levels, the total degrees of freedom is 12. Hence the selected orthogonal array shall have at least 12 experiments. An L12 orthogonal satisfies this requirement.

Once the minimum number of experiments is decided, the further selection of orthogonal array is based on the number of independent variables and number of factor levels for each independent variable.

4.4. Assigning the independent variables to columns

The order in which the independent variables are assigned to the vertical column is very essential. In case of mixed level variables and interaction between variables, the variables are to be assigned at right columns as stipulated by the orthogonal array [3].

Finally, before conducting the experiment, the actual level values of each design variable shall be decided. It shall be noted that the significance and the percent contribution of the independent variables changes depending on the level values assigned. It is the designers responsibility to set proper level values.

4.5. Conducting the experiment

Once the orthogonal array is selected, the experiments are conducted as per the level combinations. It is necessary that all the experiments be conducted. The interaction columns and dummy variable columns shall not be considered for conducting the experiment, but are needed while analyzing the data to understand the interaction effect. The performance parameter under study is noted down for each experiment to conduct the sensitivity analysis.

4.6. Analysis of the data

Since each experiment is the combination of different factor levels, it is essential to segregate the individual effect of independent variables. This can be done by summing up the performance parameter values for the corresponding level settings. For example, in order to find out the main effect of level 1 setting of the independent variable 2 (Table 1), sum the performance parameter values of the experiments 1, 4 and 7. Similarly for level 2, sum the experimental results of 2, 5 and 7 and so on.

Once the mean value of each level of a particular independent variable is calculated, the sum of square of deviation of each of the mean value from the grand mean value is calculated. This sum of square deviation of a particular variable indicates whether the performance parameter is sensitive to the change in level setting. If the sum of square deviation is close to zero or insignificant, one may conclude that the design variables is not influencing the performance of the process. In other words, by conducting the sensitivity analysis, and performing analysis of variance (ANOVA), one can decide which independent factor dominates over other and the percentage contribution of that particular independent variable. The details of analysis of variance is dealt in chapter 5.

4.7. Inference

From the above experimental analysis, it is clear that the higher the value of sum of square of an independent variable, the more it has influence on the performance parameter. One can also calculate the ratio of individual sum of square of a particular independent variable to the total sum of squares of all the variables. This ratio gives the percent contribution of the independent variable on the performance parameter.
In addition to above, one could find the near optimal solution to the problem. This near optimum value may not be the global optimal solution. However, the solution can be used as an initial / starting value for the standard optimization technique.

5. Robust Design

A main cause of poor yield in manufacturing processes is the manufacturing variation. These manufacturing variations include variation in temperature or humidity, variation in raw materials, and drift of process parameters. These source of noise / variation are the variables that are impossible or expensive to control.
The objective of the robust design [4] is to find the controllable process parameter settings for which noise or variation has a minimal effect on the product's or process's functional characteristics. It is to be noted that the aim is not to find the parameter settings for the uncontrollable noise variables, but the controllable design variables. To attain this objective, the control parameters, also known as inner array variables, are systematically varied as stipulated by the inner orthogonal array. For each experiment of the inner array, a series of new experiments are conducted by varying the level settings of the uncontrollable noise variables. The level combinations of noise variables are done using the outer orthogonal array.

The influence of noise on the performance characteristics can be found using the ratio [4] where S is the standard deviation of the performance parameters for each inner array experiment and N is the total number of experiment in the outer orthogonal array. This ratio indicates the functional variation due to noise. Using this result, it is possible to predict which control parameter settings will make the process insensitive to noise. However, when the functional characteristics are not affected by the external noises, there is no need to conduct the experiments using the outer orthogonal arrays. This is true in case of experiments which are conducted using the computer simulation as the repeatability of a computer simulated experiments is very high.

BIBLIOGRAPHY

[0] Sankar Nelian, "Integration of Taguchi Design of Experiments and Finite Method for Robust Design", master's project for Graduate School of the University of Massachusetts, 1996

[1] Douglas E. Montgomery, "Design and Analysis of Experiments", Third Edition, John wiley and sons, 1991

[2] Dr.Genichi Taguchi,"Taguchi Methods, Design of Experiments",ASI Press

[3] Dr.Genichi Taguchi and S. Konishi, "Taguchi methods: Orthogonal arrays and linear graphs", ASI Press

[4] Khosrow Dehnad, "Quality control, Robust Design, and the Taguchi method", Wadsworth & Brooks/Cole

[5] Tapan P.Bagchi,"Taguchi Methods Explained",Prentice-Hall of India, 1993

[6] Ranjit K. Roy,"A Primer on the Taguchi Method",Van Nostrand Reinhold, Newyork

[7] Anthony R. Rizzon, "Quality engineering with FEA and DOE", Mechanical Engineering, May 1994

[8] A library of over 200 orthogonal arrays. Maintained by N. J. A. Sloane (email address: njas@research.att.com; home page: www.research.att.com/~njas/), July 2003