Proceedings of the 2016 International Conference on Industrial Engineering and Operations Management
Detroit, Michigan, USA, September 23-25, 2016

Design of Experiments and Web Page Designs: Theories and Applications

Lihui Shi
Centerfield Media
El Segundo, CA, USA
shilihui@uw.edu

Bo Li
School of Arts
Wuhan Sports University
Wuhan, Hubei, China
82bobo@163.com

Abstract

Design of experiments (DOE) is a systematic method to determine the relationship between factors affecting a process and the output of that process. While it is frequently applied into many industries such as agriculture, clinical trials, chemical engineering, manufacturing, service, etc, only the simplest concepts in DOE are usually applied in information technology as such A/B and multivariate testing using full factorial designs. We apply some more advanced methodologies in DOE into our web page designs, mainly the fractional factorial design ideas to design, test and optimize the websites more accurately, economically and efficiently. Some of those ideas and methodologies are very similar to the Taguchi method (robust design method) developed by Genichi Taguchi decades ago in manufacturing industry. Some different criteria will be discussed for the optimal design selection in fractional factorial designs, and examples will be given for illustration and evaluation.

Keywords
Design of Experiments, Fractional Factorial Design, Web Analytics, Minimum Aberration Designs, Taguchi Method

1. Introduction
1.1 A/B Test and Multivariate Test
Website design and development is a very important task for every company who operates its own website, and it can be very competitive among the company and its competitors on the market. After building a website, and website optimization is especially crucial for the success of the websites. Conversion is usually the most interested metric for the pay-per-click (PPC) industry and search engine optimization (SEO) campaigns, and a well-designed web page will be very helpful for the business owners to achieve a decent number of conversions, which directly lead to the high revenue and profit.

With the development of the information technology and especially in digital marketing, more sophisticated and rigorous statistical methodologies are required for the success of website designs, testing and analysis. A/B test and multivariate test are two basic web testing methods applied by many companies nowadays.

A/B test is comparing two versions of a web page to see which one performs better. It is a randomized experiment with two variants, A and B, in a controlled experiment. Instead, multivariate test is for testing a scenario in which multiple variables are modified. In some sense, the multivariate test can be considered as multiple A/B tests. The goal of multivariate test is to determine which combination of variations performs the best out of all of the possible combinations.

Multivariate test is usually implemented by a design of experiment method called full factorial design, which is an experiment whose design consists of two or more factors, each with discrete possible values or levels, and whose experimental units take on all possible combinations of these levels across all such factors. For example, if there are n factors A, B, C…, and each factor has 2 levels, then a full factorial design would require 2^n runs.

Multivariate test has some clear advantages over A/B test: if conducted properly, a multivariate test can eliminate the need to run several sequential A/B tests on the same page with the same goal. Instead, the multivariate tests are run concurrently with a greater number of variations in a shorter period of time.

However, there are also some challenges for the multivariate test. The biggest challenge is, the number of variations in a multivariate test can add up quickly due to the fully factorial nature of these tests, which might make the design phase really time consuming. Also, another challenge is, since there are so many variations, the allocated traffic to each variation might be lower, and this can lead to the insufficient amount of visitor traffic required to reach meaningful results. Furthermore, one more challenge of multivariate testing is that, possibly one or more variables being tested do not have a measurable effect on the conversion goal, which would make the design tricky and not effective enough.

1.2 Some Basic Design of Experiments Methods

In order to solve all these challenges, we introduce a more efficient way to conduct the multivariate test, called the fractional factorial design in design of experiments. It is an experimental design consisting of a carefully chosen subset (fraction) of the experimental runs of a full factorial design. The design generators determine how the subset (fraction) is selected from the full set of runs in a fractional factorial design.

Consider the full factorial design 2^6 where –1 and +1 represent the low and high factor levels, respectively. Assume we want to have a smaller design with 2^{6-2} runs instead of the full one, with a fraction $2^{-2}= ¼$, we need 2 design generators to achieve such a small design. Suppose those 6 factors are called A, B, C, D, E and F, then there are different choices for the selection of the design: for example, we can have design generators E = AB, F = ACD, called design D_1; and another choice is E = ABC, F = ABD, called design D_2. We can see that by using those design generators, we will be unable to differentiate the effects E and AB, as well as effects F and ACD, in design D_1. This is called aliased, or confounding, which means that we have lost the ability to estimate some main and/or interaction effects.

Here we call the effect of E as main effect, the effect of AB as two-factor interaction effect, and the effect of ACD as three-factor interaction effect. If we are able to estimate a main and/or interaction effect, then it is related to the concept of clear effect. We call a main effect or two-factor interaction effect clear if none of its aliases are main effects or two-factor interaction effects. For design D_1, since we have E = AB and F = ACD, so we will have I = ABE and I = ACDF, where I denotes the column of all +'s. These equations with such identity element I are called the defining contrast subgroups.

When we design an experiment, obviously a clear effect is what we expect and the alias is what we try to avoid. In full factorial design, it is evident that we will be able to achieve all effects to be clear. However, an important fact about the fractional factorial design is that, aliasing of effects is a price one must pay for choosing a smaller design, which can be shown by the previous example. Therefore, a basic question for the fractional factorial design would be, which fraction should be selected to be the optimal design.

Due to the effect hierarchy principle in design of experiments, which suggests the lower-order effects are more likely to be important than the higher-order effects, and same order effects are likely to be equally important, we usually try to achieve all the main effects and second-order interaction effects to be clear in our design of experiments. Since there are main effects aliased with two-factor interaction effects in resolution III designs, and there are no main effects aliased with two-factor interaction effects in resolution IV designs, only resolution IV designs are considered when searching for the optimal design.

A regular 2^{n-k} design has n factors each at two levels and 2^{n-k} runs, and is completely determined by k defining contrast subgroups. Let $A_i(D)$ be the number of words of length i in the defining contrast subgroup of design D, where the length of a word is the number of letters in the word. The vector $W = (A_3,...,A_k)$ is called the wordlength pattern of the design. The resolution of design D is the smallest integer R such that $A_R(D) > 0$. For the above examples, the defining contrast subgroup of design D_1 is I = ABE = ACDF = BCDEF, so it has resolution III; the defining contrast subgroup of design D_2 is I = ABCE = ABDF, so it has resolution IV. As can be seen, all the main effects are clear in the resolution IV design, and there are main effects aliased with two-factor interaction effects in resolution III design. Therefore, design D_2 is superior than design D_1 since all its main effects are clear.

There are different optimal criteria for the selection of the fractional factorial design, and two important criteria are maximum resolution design and minimum aberration design. The maximum resolution design chooses the design with maximum resolution. For any two 2^{n-k} designs D_1 and D_2, let r be the smallest integer such that $A_r(D_1) \neq A_r(D_2)$. Then D_1 is said to have less aberration than D_2 if $A_r(D_1) < A_r(D_2)$. If there is no design with less aberration than D_1, then D_1 has minimum aberration. In the above example, since $A_3(D_2) = 0$ but $A_3(D_1) > 0$, so design D_2 has less aberration than D_1. The minimum aberration criterion selects designs that sequentially minimize $A_3(D),..., A_n(D)$.

There are systematic methods to generate the maximum resolution design and minimum aberration design for any given n and k, and some results are also listed in [1]. One basic procedure is, based on the given n and k, first find out all the possible designs with each possible design generators combination, then calculate the wordlength patterns for each design, then compare their wordlength patterns to find out the maximum resolution design and the minimum aberration design based on the definition. For given n and k, its maximum resolution design and minimum aberration designs are both not necessarily unique, and those multiple optimal designs are often isomorphic. It is interesting that some of those ideas and methodologies in design of experiments are very similar to the Taguchi method (robust design method) developed by Genichi Taguchi decades ago in manufacturing industry [2].

2. An Motivated Example

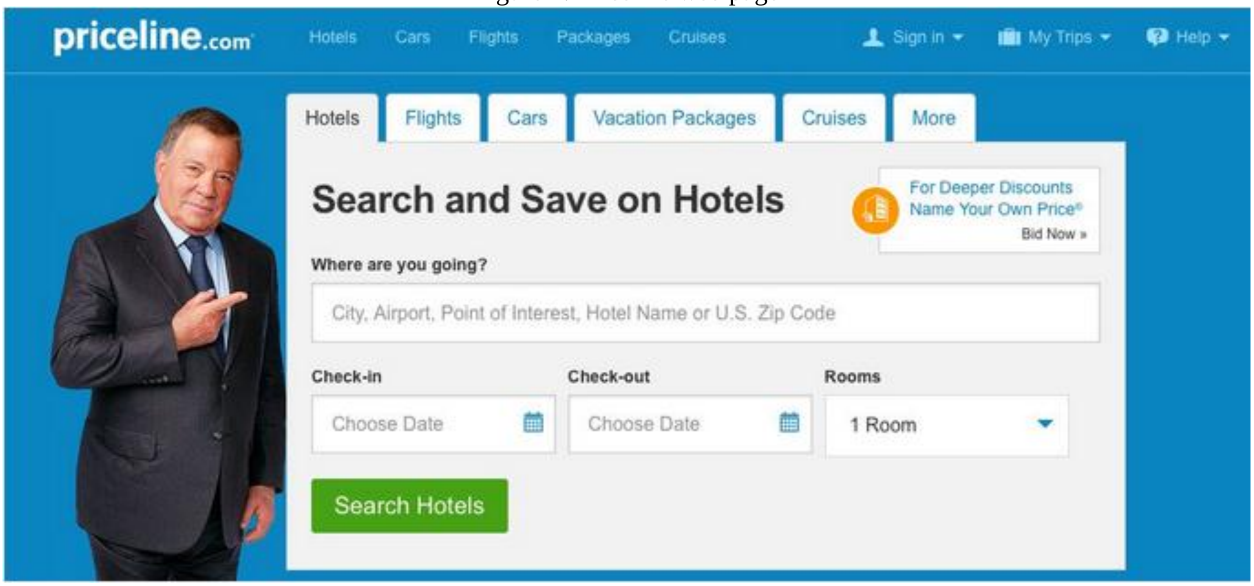
Here an artificial example is given on how to apply the design of experiments procedure into a web page design problem. Assume that we have the task to design and optimize the web page of Priceline, and there are 4 factors that can be tested under 3 different variants:

(1) The image: man hand in pocket -- 0, man points finger -- 1, man with fists -- 2.

(2) The text line: no change or cancel fees -- 0, hotel deals you won't find anywhere else -- 1, search and save on hotels -- 2.
(3) The background color of the text line: no fill -- 0, yellow -- 1, blue -- 2.
(4) The position of the image on the web page: bottom part -- 0, left side -- 1, right side -- 2.

Then it can be seen easily that the following web page in Figure 1 corresponds to the factor level combinations of (A, B, C, D) = (1, 2, 0, 1). And Our goal is to find out the optimal web page layout with the above factors. In the next sections, design and analysis procedures will be provided for this example.

Figure 1. Priceline web page



3. Basic Procedures of Applying Design of Experiments for Web Page Designs

In this section, how the fractional factorial design method can be applied in the web page design will be illustrated, and the design and analysis procedures will be explained step by step as well.

The goal of web page design and testing is to try various types of variants in order to identify those factors that have a significant effect on a measured result, and then identify the particular levels of significant factors that positively contribute to desired results. To get the effects of different factors on the KPI (key performance indicator), which is often the conversion rate on the web page, i.e., the proportion of visitors to a website who take action, a regression model is required usually. The basic steps for the web page design and testing are proposed as follows:

(1) Based on the business objective and requirements from the web page owner, we know how many factors will be tested, and how many levels for each factor exist.
(2) Also get the indication of the interaction effects that might be of interest. Avoid the designs that might aliased with those interaction effects in the following experiments design selection.
(3) After we get those basic information, find out its minimum aberration design, which includes the design generators, experiment run size, etc.
(4) Create the corresponding web page variants and ready for the multi-factor/multi-level experiments.
(5) Conduct the experiments.
(6) With the data collected for each web page variant, calculate the least squares estimator of the coefficients for effects.
(7) Using an analysis of variance (ANOVA) table to find out all the significant effects.
(8) Based on the estimated coefficients for all the factors, choose the optimal levels for the significant factors, which corresponds to the optimal web page variant.

4. Evaluations and Analysis for the Experiments

In this section, the specific steps will be discussed in detail for the Priceline web page design problem in section 2.

For simplicity, assume that no interaction effects in this case. The full factorial 3^4 design would require 81 web-page variants, which is too large to be implemented. Since the interaction effects do not exist, we can assign design generators using a fractional factorial design to make the design much smaller. Table 1 shows the design matrix for a 3^{4-2} design, with design generators C = AB and D = AB^2, which is both the maximum resolution design and the minimum aberration design for n = 4 and k = 2 [1]. The first two columns in the design matrix makes a full factorial design 3^2 for factors A and B, with 9 experiments, then the factors C and D are generated by the given design generators. Notice that the modulo operation is used for the design generators calculation, for example, in experiment 9, the level of factor C is defined as C = AB = (2 + 2) mod 3 = 4 mod 3 = 1, and the level of factor D is defined as D = AB^2 = (2 + 2*2) mod 3 = 6 mod 3 = 0.

Table 1. The 3^{4-2} design with C = AB and D = AB^2

	Factor A	Factor B	Factor C	Factor D
Experiment 1	0	0	0	0
Experiment 2	0	1	1	2
Experiment 3	0	2	2	1
Experiment 4	1	0	1	1
Experiment 5	1	1	2	0
Experiment 6	1	2	0	2
Experiment 7	2	0	2	2
Experiment 8	2	1	0	1
Experiment 9	2	2	1	0

In Table 1, each row in the design matrix represents a different experiment, i.e., one web-page variant. Thus, each row is labeled with an experiment name, such as "Experiment 1" for the first row in the design matrix, and each column in the design matrix corresponds to a single factor A, B, C, B. Each numeric value in the design matrix represents a particular level for the factor in the experiment. In the design matrix in Table 1, three different levels for 4 different factors are specified. For example, the experiment, or web-page variant, with all factors having the zero level is represented by the first row in the design matrix. The theory of fractional factorial design provides design methodology for testing multiple levels of multiple factors systematically without needing to test all possible web-page variants. In this case, statistically meaningful results are obtained for 4 three-level factors using only 9 page-variants rather than 3^4 = 81 possible web-page variants.

After the 9 web-page variants are designed, each one will be tested for the online users for data collection for a certain period of time, and then the data will be analyzed using the following regression model:

$$y_1 = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1^0 + \boldsymbol{\beta}_2^0 + \boldsymbol{\beta}_3^0 + \boldsymbol{\beta}_4^0 + \boldsymbol{\epsilon}_1$$
$$y_2 = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1^0 + \boldsymbol{\beta}_2^1 + \boldsymbol{\beta}_3^1 + \boldsymbol{\beta}_4^2 + \boldsymbol{\epsilon}_2$$
...
$$y_8 = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1^2 + \boldsymbol{\beta}_2^1 + \boldsymbol{\beta}_3^0 + \boldsymbol{\beta}_4^1 + \boldsymbol{\epsilon}_8$$
$$y_9 = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1^2 + \boldsymbol{\beta}_2^2 + \boldsymbol{\beta}_3^1 + \boldsymbol{\beta}_4^0 + \boldsymbol{\epsilon}_9$$

In the above equations, y_i is the conversion rate for the ith web-page variant, the values β_k^l represent the coefficients associated with the lth level or levels of the kth factor, when l and k are single-digit quantities. Instead, when l and k are multiple-digits, it is associated with the interaction effects. For example, $\boldsymbol{\beta}_{\mathbf{k1,k2}}^{\mathbf{l1,l2}}$ represents the coefficient associated with the two-factor interaction effect for the k_1th and k_2th factors, with factor k_1 on the l_1th level and factor k_2 on the l_2th level. The above equations can usually be written in the matrix form $Y = X\beta + \epsilon$, i.e.,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 1&1&0&0&1&0&0&1&0&0&1&0&0 \\ 1&1&0&0&0&1&0&0&1&0&0&0&1 \\ 1&1&0&0&0&0&1&0&0&1&0&1&0 \\ 1&0&1&0&1&0&0&0&1&0&0&1&0 \\ 1&0&1&0&0&1&0&0&0&1&1&0&0 \\ 1&0&1&0&0&0&1&1&0&0&0&0&1 \\ 1&0&0&1&1&0&0&0&0&1&0&0&1 \\ 1&0&0&1&0&1&0&1&0&0&0&1&0 \\ 1&0&0&1&0&0&1&0&1&0&1&0&0 \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix},$$

where $\beta = [\beta_0, \beta_1^0, \beta_1^1, \beta_1^2, \beta_2^0, \beta_2^1, \beta_2^2, \beta_3^0, \beta_3^1, \beta_3^2, \beta_4^0, \beta_4^1, \beta_4^2]^T$, while Y, X and ϵ are the corresponding vectors and matrix. Notice that there is a one-to-one mapping between the design matrix in Table 1 and the X matrix in the regression model here.

From the linear regression theory, we have the lease squares estimator (LSE) of β is $\hat{\beta} = (X^TX)^{-1}X^TY$. After 14 days of experiments, the data collected for the 9 web-page variants are:

Table 2. The 3^{4-2} design with C = AB and D = AB^2

	Number of visits	Number of conversions	Conversion Rates
Experiment 1	5142	121	2.35%
Experiment 2	5085	105	2.06%
Experiment 3	4876	78	1.60%
Experiment 4	4989	45	0.90%
Experiment 5	4760	101	2.12%
Experiment 6	4943	89	1.80%
Experiment 7	5206	70	1.34%
Experiment 8	5181	131	2.53%
Experiment 9	4995	92	1.84%

Replace the Y vector by the conversion rates column in Table 2, we can get the LSE of β is:
$\hat{\beta}$ = (0.0079, 0.0043, 0.0003, 0.0033, -0.0005, 0.0066, 0.0017, 0.0065, 0.0002, 0.0011, 0.0053, 0.0010, 0.0016)^T

Then the estimated coefficients for factors A, B, C, D are (0.0043, 0.0003, 0.0033), (-0.0005, 0.0066, 0.0017), (0.0065, 0.0002, 0.0011) and (0.0053, 0.0010, 0.0016) respectively, in order to maximize the conversion rate, we should choose 0.0043, 0.0066, 0.0065 and 0.0053 for factors A, B, C, D respectively, i.e., the optimal level for the 4 factors are: factor A on level 0, factor B on level 1, factor C on level 0 and factor D on level 0. More specifically, the optimal web page should have the following setups:

(1) The image: man hand in pocket.
(2) The text line: hotel deals you won't find anywhere else.
(3) The background color of the text line: no fill.
(4) The position of the image on the web page: bottom part.

It can be easily seen that the optimal factor level combinations are not among the 9 experiments in Table 1. This is a great advantage for applying fractional factorial design in web page design problem.

5. Conclusions and Future Work

In this paper, the website testing problem is formulated into a fractional factorial design problem, and basic procedures have been proposed. A Priceline website design and optimization problem is applied here as a motivated example, and the maximum resolution design and minimum aberration design are selected to be the optimal criteria for the fractional factorial design selections. For simplicity, we assume there is no significant interaction effects.

Future work includes the application of more complicated web page design problems, such as mixed-level factor design experiments, and adaptive online experiments for web page design and testing, etc.

References

[1] Wu, J., Hamada, M., *Experiments: Planning, Analysis, and Optimization*, 2nd Edition, Wiley, 2000.

[2] Phadke, M., *Quality Engineering Using Robust Design*, 1st Edition, Prentice Hall, 1989.

Biography

Lihui Shi is a senior data scientist in Centerfield Media at El Segundo, CA, USA. He earned B.S. in Information and Computational Science from Hebei University of Technology, Tianjin, China, Masters in Statistics from Nankai University, Tianjin, China and University of Washington, Seattle, WA, USA and PhD in Industrial and Systems Engineering from University of Washington, Seattle, WA, USA. He has published papers on journals and conferences, such as Quality and Reliability Engineering International, Quality Technology and Quantitative Management, International Journal of Performability Engineering, Industrial Engineering Research Conference, International Conference on Pattern Recognition, etc, on different research areas such as statistical process control, process adjustment, parametric and nonparametric statistics, supervised and unsupervised learning, etc. Dr Shi has been working in data science team of several IT companies and completed research projects on A/B & Multivariate Testing on web analytics, paid search (Google/Bing/Yahoo adwords, PLA and display ads), and bidding models using machine learning methods. His research interests include quality and reliability, applied statistics and machine learning. He is a member of IIE, INFORMS, ASQ (American Society for Quality) and ASA (American Statistical Association).

Bo Li is an associate professor in School of Arts at Wuhan Sports University, Hubei, China. He holds a Bachelor degree in Physical Education, a Master degree in Physical Education and Sports Training, at Wuhan Sports University, Hubei, China. He received his PhD from the National University of Physical Education and Sport in Ukraine. He taught courses in Latin dance, Jazz and physical education. He has research interests in physical education, dance kinesiology and data analysis in dancing performance. He has published over 10 papers in academic journals and international conferences, in English, Russian and Chinese.