UNDERSTANDING STATISTICAL DESIGN AND ANALYSIS OF EXPERIMENTS IN THE CLASSROOM

Ho Siong Lin, Ngee Ann Polytechnic, Republic of Singapore

This paper illustrates the use of the Catapult device for enhancing the understanding of design of experiments in the classroom. It is an effective training tool that simulates the design of a real-life product. 2^K Factorial design is used to analyse the experimental data. Key input factors that significantly affect the output response are determined.

INTRODUCTION

Statistical Design of Experiments (SDOE) involves understanding how different factors to a system, whether concerning product or process, affect the output(s) and subsequently identify the critical factors and their optimal settings. It has emerged as the premier technique in the product and process development cycle for quality improvement and is widely being used by practitioners in the industries. If these methodologies are properly applied, significant benefits in terms of effectiveness, cost, optimization and efficiency can be achieved. This gives them an added edge over their competitors.

However, when teaching SDOE to the students, it is a challenge to the lecturers to bring across not only the theoretical knowledge but more importantly the environmental setting that can be related so that the techniques can be applied appropriately. Real life case studies do help in illustrating its applicability but often they are difficult to visualize and understand. Alternate means are sought and the use of a mechanical device such as the Catapult can be a very effective teaching aid for demonstrating the power of blending engineering theories with statistically designed experiments. It is an excellent tool to simulate the design of a product and through design of experiments, a better product (catapult) can be designed with increased productivity and reduced time to market (need to complete assignment within two weeks). In this paper, the Catapult design from *Lighting Calculator* is presented. Experiments are planned using the 2^K factorial design approach. Techniques such as the analysis of variance, normal probability plots are presented. Significant factors and the optimal settings are identified. Finally, a regression model can be formulated from the results.

DESCRIPTION OF THE CATAPULT

The Catapult is a device for launching a small projectile, for example a ball, towards a targeted impact site. The design used in our experiments is depicted in Figure 1. Six input factors such as Arm-tension position, Upright-tension position, Ball-seat position, Ball type, Turntable and Elevator height influence the distance traveled by the ball, which defines the output response of the system. In this diagram, the coded levels (+, -) of each factor are indicated clearly. As for the Ball type, one is the ping-pong ball (-) and the other is the rubber ball (+). Experimental design techniques can be employed to determine the optimal setting of each factor to maximize or minimize the projected distance. Other Catapult designs reported in Launsby and Weese (1995), Schbert *et al.* (1992) and Luner (1994) have demonstrated how a good blend of engineering knowledge and statistical techniques can be successfully applied to modeling the catapult from experimental data, as compared to the traditional mathematical/mechanical approach.

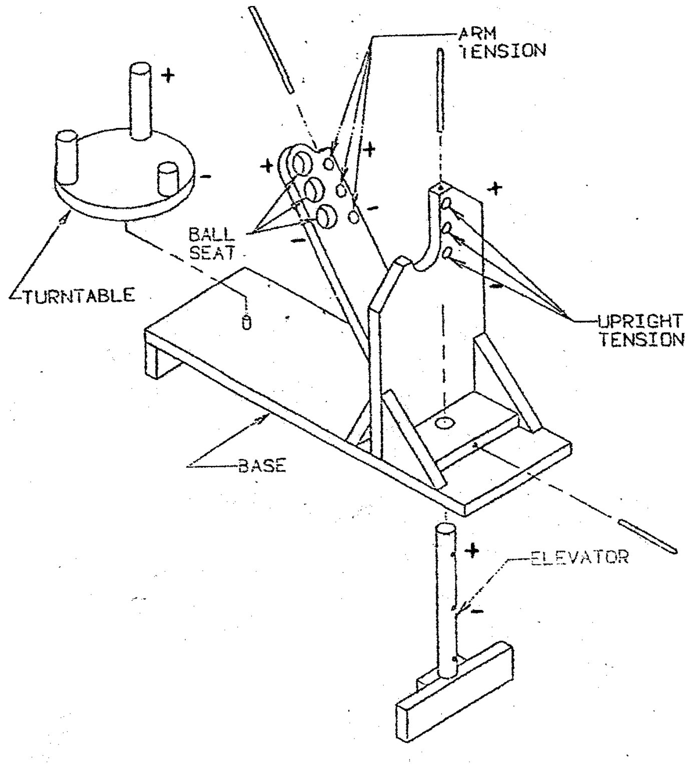

Figure 1. The Catapult Design

THE 2^K FACTORIAL DESIGN

Factorial designs are widely adopted in industries to investigate several factors and their joint effects on the response. One of the most popular techniques that form the basis of other designs of considerable practice value is the full 2^k Factorial Design. It is used to study k factors, each at two levels. The levels can be qualitative or quantitative. An example of the design matrix for four factors (A, B, C, D) with single replicate is shown in Table 1. The estimates of the main and interaction effects can be obtained using Yates' algorithm by Yates (1937). Analysis of Variance (ANOVA) is used to determine whether the effects are significant, at a specified level of confidence. In general, for a single replicate 2^k Factorial Design, if we are interested in the first factor "*A*" under the first column, then

$$Effect_{(A)} = \{ -y_1 + y_2 - y_3 + y_4 \ldots\ldots -y_{2^k-1} + y_{2^k} \} / (2^{k-1})$$

If there are *n* replicates and $y_{i1}, y_{i2}, y_{i3}, \ldots, y_{in}$ are the response data at the i^{th} run, the overall variance estimate is

$$s^2 = \frac{\sum_{i=1}^{2k}\sum_{j=1}^{n}(y_{ij} - \overline{y}_i)^2}{(n-1)2^k}$$

Therefore the standard error of each effect estimate will be

$$std.error = \sqrt{\frac{s^2}{(n)2^{k-2}}}$$

and the mean square error (MSE) in the ANOVA table can be used to estimate s^2.

The 95% confidence interval for the effect of factor "*A*" can thus be established as follows :

$$Effect_{(A)} \pm 2\ (std.error)$$

AN EXPERIMENTAL STUDY

A 2^4 Factorial experiment with two replicates was conducted and the observations are recorded in Table 2. The factors chosen are Ball Type (A), Elevator height (B), Turntable (C) and Ball-seat position (D). Both Arm-tension position and Upright-tension position are assumed fixed at the normal position. Assuming that three-factor interactions are negligible, then analysis in Tables 3-4 indicate that main effects B, C, D and interaction effects BC are significant at α = 5%. This is also evident from the effects' 95% confidence interval, which do not include zero, and the normal probability plots of effects presented in Figure 2. Hence, an empirical model can be built from the experimental results. However, it is important to validate the assumptions of normality and equality of variance in ANOVA through the normal probability plot of residuals in Figure 3. The results are satisfactory. Further refinement of the model is possible by removing the non-significant variables in the analysis and the analysis are shown in Tables 5-6. The final regression model has the form

$$Y = 231.56 - 13.87B - 44.87C + 13.75D + 7.68BC$$

As our objective is to maximize the distance traveled by the ball, the effect plots in Figures 4-7 suggest that the optimal level combination of factors is $B^{(-)}C^{(-)}D^{(+)}$. It is also concluded that the material of the ball has the least impact on the distance traveled.

CONCLUSION

This Catapult has created the right environmental setting for the students and also enhanced their understanding of experimental design. Key factors and the optimal levels are identified from the analysis of experimental data. However, since the levels are qualitative in this design, we can only postulate a regression model representation of the Catapult based on coded variables. Hence, further improvements can be made in this respect. Nevertheless, students are now more motivated in learning SDOE and also eager to get involved in different stages of the experiments such as planning the design strategy, brainstorming, data collection and analysis. As evident from this study, we have seen an excellent training tool for demonstrating experimental design in the classroom and every student is beginning to LOVE "SDOE" now!

REFERENCES

Luansby, R. G. and Weese, D. L. (1995). *Straight Talk on Designing Experiments*, 2nd ed., Launsby Consulting, Colorado.

Luner. Continuous Improvement with Dual Response Approach, *Quality Engineering*, 692-705.

Schubert, K., Kerber, M. W., Schmit, S. R. and Jones, S. E.(1992). "The Catapult Problem: Enhanced Engineering Modeling Using Experimental Design", *Quality Engineering*, *4(4)*, 463-473.

Yates, F. (1937). *Design and Analysis of Factorial Experiments*, Tech. Comm. No. 35, Imperial Bureau of Soil Sciences, London.