**Basketball Hits Rate Design of Experiment**

**Instructor**: Dr. Chihoon Lee

**Authors**: Rumeng Zuo, Qian Lu, Yang Wu, Haodong Zhao

**Date**: May 13, 2019

**BIA-654B-Experimental Design II**

# Abstract

Basketball is a kind of quite popular sport, which is suitable to every citizen all around the world. Especially for those energic young people, they always spend lots of time to watch video to learn and go to the gym to practice. In order to make the practice more effective and efficient, most significant factors, which affect basketball practice, should be considered first. In this situation, players can get maximal benefits from practice.

# *Keywords*

Fractional factorial design, Design of experiment, Significant factors

# Project Motivation

In this project, we played basketball and counted the successful hits in 20 shoots of different patterns to find out what factors may affect the hit rate.

This topic is selected as it is not boring at all when repeating experiments. Besides, it’s healthy and relaxing. Basketball is popular among students, and we would like to apply what we learned in Design of Experiment to do some field research and improve the way we play the basketball.

Some of the conditions we have here in this project might seem to be a little bit impractical, for example, shooting with a football. However, we think it would be fun to add these factors and we might get surprising results.

# Description of the Experiment

We mainly follow 3 steps to conduct the experiment: data collection from gym​, data processing in JMP​ and results analysis​.

## Response variable

The response variable is successful hits of 20 shots.

## Factors

We design 5 factors, each in 2 levels to manipulate:

|  |  |  |
| --- | --- | --- |
| Factors​ | +​ | -​ |
| Posture​ | Jump​ | No Jump​ |
| Defense​ | With defense​ | Without defense​ |
| Location​ | Penalty line​ | Low post​ |
| Hand(s)​ | One hand​ | Two hands​ |
| Ball​ | Basketball​ | Football​ |
|  |  |  |

**Table 1**. Factors

1. Posture when shoot the goal: Jump/Do Not Jump
2. Whether there is one on defense: With Defense/Without Defense
3. Where to shoot the goal: at Penalty Line/at Low Post
4. Shoot with one hand or both hands
5. The ball: basketball/football

## Blocks

With 4 team members, we set 4 persons as 4 blocks to get rid of the disturbance. Each player is randomly assigned to treatment conditions. Because this design reduces variability and potential confounding, it produces a better estimate of treatment effects.

## Design method

We intend to use fractional factorial design because one full replicate of the full factorial design is hard to afford. Factorial factional design was selected because each factor has two levels (+ and -) and we tried to analyze the interaction between different factors. In the experiment, we use N = 2^(k-p), where k=5, and p=3, therefore, N=4.

## Replication

Replication is the repetition of an experimental condition so that the variability associated with the phenomenon can be estimated. We replicated once in this experiment to insure less bias. For each condition, each person has another 20 hits, counted the successful hits and repeated this action again.

## Randomization

Randomized experiment is the experiment that allows the greatest reliability and validity of statistical estimates of treatment effects. We list the five factors and their positive levels and negative levels in JMP. The patterns were designed in JMP in advance which provided randomization before conducting the field test in gym.

# Experiment Conduct

We went to the basketball court at Stevens Gym on April 23. The playground is indoor so there is no need to consider the weather. We had 4 different players with different experience in basketball to take a shot. We took a basketball and borrowed a football from gym as our experiment equipment. Following the patterns designed earlier, we shoot the goals and recorded the successful hits in JMP as the response values.

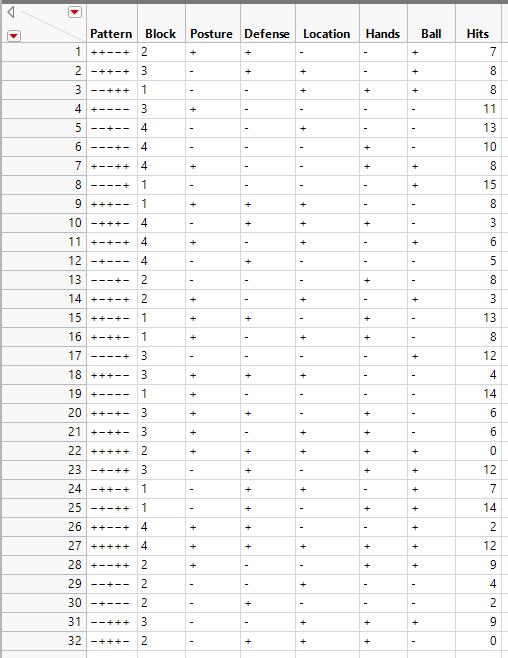
# Result Analysis

We used the following methods to analyze our experimental results.

## Factorial Factor Design

After collecting data from the court, we designed a Factorial Factor Design for our experiment. Different from Full Factor Design, Factorial Factor Design used the subsets of the former one, ignored the compound effects of the interaction of 3 factors, and focused effects of primary elements and two factors interaction.

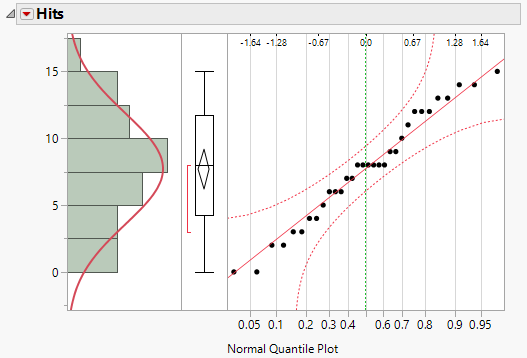
We set a 2 to the power of 3 times design pattern for each block generated in JMP, and we did one replication. The following figure is part of our experiment design. It is noticeable that we did random factors design for each experiment, meaning that each factor was not tested by order. For instance, row 2 means, for block 3, he/she should shoot the ball without jumping (-posture), with defense (+defense), in the penalty line (+location), with the left hand (- hand), and lastly, with basketball (+ball). And the response, which is the hit rate, is 8 out of 20.

****

**Fig 1**. Factorial Factor Design Sample Data

## Normal Quantile Plot

The normal quantile plot reflects an intuition of whether a set of data is normally distributed. For the X-axis, it shows the theoretical quantiles where a standardized data would be plotted. As we could see, all the points in the plots represented a hit rate. Those points were all located almost close to the normal distribution line. From the left graph, each green bar chart represents the group of hit rate: Above 15 out of 20 only weighs less than the rate 8 out of 20. Besides, the distribution of green bar chart followed the normal distribution line (red line). Concluded from the above, we believed that our response data (hit rate) is normally distributed at 95% confidence level.

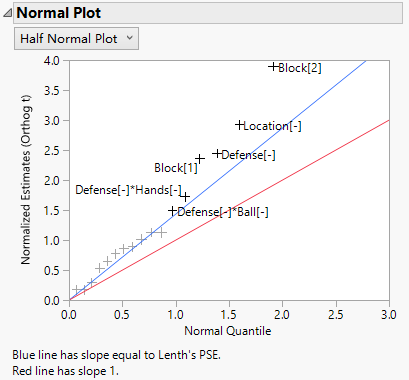
****

**Fig 2.** Normal Quantile Plot

## Half Normal Plot

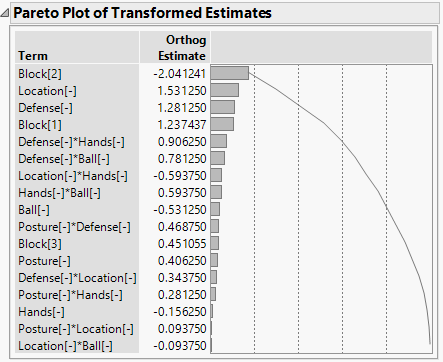
To test which effects would be significant in our model, we first plotted a half normal plot. The half-normal plot represents the absolute value of the comparison derived from the quantile absolute value of a distribution. Different from the normal plot, the significance factor appears above the lines towards the upper right corner of the graph.

In our case, we could see that any labeled effects were considered as the statistical significance that far from the blue line. The other factors would be either no or little effect. As shown in the figure, we guessed that the effects of location, defense, the interaction effect of defense and ball, and the interaction effect of defense and hand, would have significant influence. So far, we still need more analysis to explore the significances of factors.

****

**Fig 3.** Half Normal Plot

## Pareto Plot of Transformed Estimates



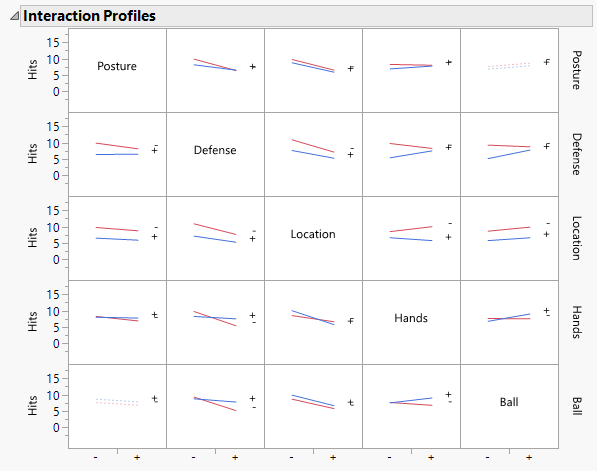
**Fig 4.** Pareto Plot of Transformed Estimates

The Pareto Plot is a vertical bar chart representing the significant influence of terms in order. The idea of Pareto plot was from the Pareto principle (80/20 rule): about 80% percent of effect contribution is derived from the 20% factors, while the rest of 80% factors only represent a little effect.

Based on this plot, we aimed to find the top significant factor. We could see that location has the most effects among all factors, followed by the defense, the interaction effect of defense and hand, and the interaction effect of defense and ball. Therefore, we can double confirm the significance above four factors.

## Interaction Plot

Interaction plot shows two factors’ interaction effect on each other. As the plot presents following, when the two lines are parallel to each other, which means the two factors doesn’t have much effect on each other. Otherwise, there will be a significant effect. For instance, according to observation, “Defense” and “Hands” have strong effect on each other to some extent.

****

**Fig 5.** Interaction Plot

## Sorted parameter estimates

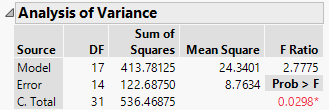
Sorted parameter estimates is a simple tool to observe the significance of each factor and its interaction with other factors. The length of bars is sorted by the absolute value of t-ratio. As the absolute value of t-ratio decreases, the significance to the experiment is decreasing. Besides, the two blue lines are auxiliary tools to indicate the significant factors. Those bars, which are longer than the interval of blue lines, are relatively significant in the experiment.

## Summary of Fit

Summary of Fit shows a summary for a one-way analysis of variance. R-square measures the proportion of the variation accounted for by fitting means to each other factor level and the rest variation is attributed to random error. In our experiment, R-square is 0.77, which means that 77% points are fitted well to the regression line. Adjusted R-square is 0.49, which is calculated as 1 - (Mean Square of error/Mean Square of C Total).

## Analysis of Variance

Analysis of variance has four main aspects to analysis, DF, Sum of Squares, Mean Square and F ratio. Generally speaking, we hypothesis that the group means are equal as H0. Otherwise, the group means are not equal as H1. In this case, F ratio is 2.7775, and obtaining significant probability is 0.0298, which is smaller than the confidence interval, 0.05. Thus, we can conclude that there are no differences in the group means.

****

**Fig 6.** Analysis of Variance

## Residual Analysis

Residuals are estimates of experimental error obtained by subtracting the observed responses from the predicted responses. When residuals are approach to the regression line, it means that the variance of estimate errors is small. Examining residuals is a key part of all statistical modeling, including DOE's. By looking at residuals, we can conclude whether our assumptions are reasonable, and our choice of model is appropriate.



**Fig 7.** Formula of Residual

There are two plots to show residuals distribution of our data. In Residual Normal Quantile Plot, most of the Hits residuals are around the red line. When the normal quantile is around 0.3 to 0.8, points distribution is relatively dense. In the other plot, Studentized Residuals, more than half of points are on the blue line and about 85% points are constrained in the two green lines. Based on the plot, only four residuals are outliers, which are not in the green line interval. The four points exist obvious difference between the actual values, so we can exclude the four points. Both of the two residual plots prove that other data’s residual is relatively normal distributed, and the model selection is relative appropriate.

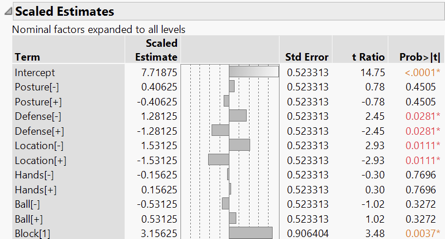
# Conclusion and Recommendation

In this project, we designed a basketball hitting experiment. The method we use is factorial factor design, and in this experiment, there are five features, each with two levels: ‘Posture’ (jump / no jump), ‘Defense’ (with defense / without defense), ‘Location’ (on penalty line / on low post), ‘Hands’ (one hand / two hands), ‘Ball’ (basketball / football). And the experiment response is the hitting rate. We also make the experimental results more realistic by using the ‘randomization’ and ‘replication’.

We hope to find out the relationship between the five features we selected and the hitting rate through experiments, but each person’s own basketball level is different, which leads to a big difference in the shooting percentage of each person, so we set people as blocks. Ignore the difference in shooting levels between people.

Through the analysis of our experimental data, we found that ‘Location’ and ‘Defense’ have a significant impact on the field goal percentage. In the case of no one to defend, or to shoot closer to the basket, the hit rate will be significantly improved. ‘Posture’ and ‘Ball’ have a slightly smaller impact on hitting rate percentage, shooting no jump shot or using a basketball to make a shot will increase the hit rate, and vice versa. And ‘Hands’ have the least impact on shooting percentage, using one hand to shoot a shot will be higher than using two hands.

In addition, we also performed regression analysis based on our experimental data, hoping to use regression prediction to get the combination of features which we selected that can achieve the maximum hitting rate.



**Fig 8.** Regression Prediction Parameters

Through regression analysis, we get the parameters of each factor and get the regression prediction equation:

Because each feature has two levels (+1 / -1), in order to get the maximum response, we should choose negative level (-1) for ‘Posture’, ‘Location’ and ‘Defense’, and choose positive level (+1) for ‘Ball’ and ‘Hands’. Therefore, the maximum response we predict is 11.625, which is about 12. To achieve this maximum response, the way we should choose to shoot is: ‘Location -’ (close shot), ‘Defense -’ (avoid defenders), ‘Ball +’ (use basketball to shoot), ‘Posture –’ (No jump shot) and ‘Hands +’ (shoot with one hand).

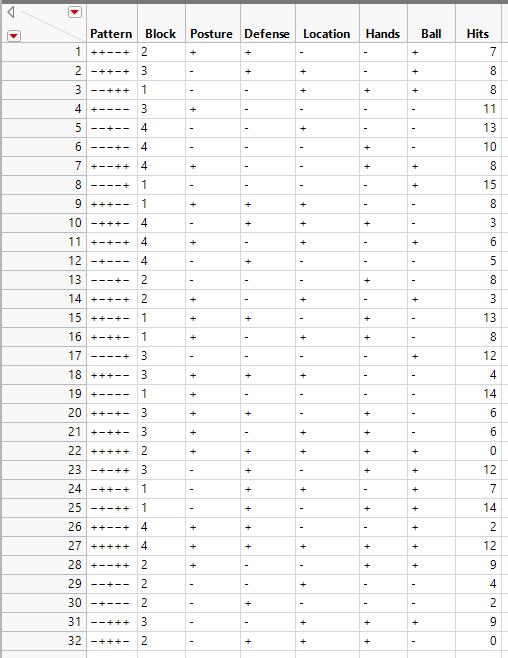
We also found that the experimental results we got could correspond to the actual basketball game, such as NBA. In the NBA game, there are three-pointers, for most players, long-range shooting is more difficult than close-range shooting, it is more difficult to score, so long-range shooting will get higher scores. And in the basketball game, there are skills ‘crossover’ and ‘pick and roll’ to avoid defense (‘jump shot’ is also a skill to avoid defense). These techniques, which avoid the direct confrontation between the shooter and the defensive player, also show that the defensive player has a great influence on the shooting percentage. Regular basketball games have also adopted a unified ball, avoiding the impact of using different balls. Moreover, most players use one-handed shooting, and the other hand only plays a role in assisting shooting. These associations with the NBA game show that our experimental results are valid. We believe that if basketball novices can play attention to these factors when they are just in contact with basketball, they can better participate in the sport and have better performance.

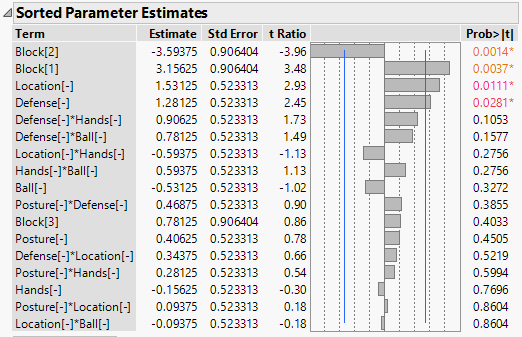
In our future work, we will add more influencing features to the experiment, and hope to analyze the impact of different features and their interaction on the basketball hitting rate, then provide basketball enthusiasts with a data reference that can effectively improve the shooting percentage.

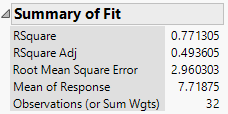
Finally, we hope that more people can actively participate in basketball. Basketball is not just a game, a kind of competition, but also a sport that helps us regulate our body and mind and health.

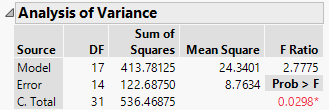
# Appendix

Our Design Table and Response

****

****

****

****

