Week 5 Lecture 1: Order effects, Counterbalancing and Latin Squares

Instructor: Eakta Jain CIS 6930, Research Methods for Human-centered Computing
Scribe: Chris(Yunhao) Wan, 1677-3116 University of Florida, Spring 2015

The within-subjects design has several advantages: it requires fewer participants; variance due to participants predispositions will be approximately the same across test conditions; it is not necessary to balance groups of participants. Because of these advantages, experiments in HCI tend to favor within-subjects designs over between-subjects designs[Mac12].

1 Order Effects

When the test conditions are assigned within-subjects, participants are tested with one condition, then another condition, and so on. In such a design, participants' performance will change due to several reasons:

- **Practice**. In most within-subjects designs, it is possible that participants' performance will improve as they progress from one test condition to the next.Through practice from one condition to another condition, they become familiar with the apparatus and procedure, and they are learning to do the task more effectively.
- **Fatigue**. In some cases, it is also possible that the participants' performance will become worse on conditions that follow other conditions. This may because of *mental or physical fatigue*.
- **Order Effect**. The participants' performance may either improved or worsen due to the order of the test conditions.

The goal of the experiment is to evaluate the performance through the test conditions. These confounding influence will greatly affect the accuracy of the experiments results. When there's problems, there always be a solution. The most common solution for the order effect is to divide all participants into groups and arrange the test conditions in a different order for each group. This is called *counterbalancing*, which is also the topic for next section.

2 Counterbalancing and Latin Square

Let's start with an example: In a design there are two levels(test conditions) A and B. The total number of participants is N. If we divide all participants into two groups, each group has $\frac{N}{2}$ participants, and the design will looks like the following table:

Group	Test Level Sequence	
1	A	B
2	B	A

Furthermore, if we increase both test level and group number, e.g. how to deal with the design of 3 groups?

2.1 Latin Square

Let's continue with the example for 3 test levels. If we equally divide the the participants into three groups, then each group will have $\frac{N}{3}$ participants. And we will distribute the test condition sequence as following table:

Group	Test Level Sequence		
1	A	B	C
2	B	C	A
3	C	A	B

Through further inspection from the previous two cases, we can find out that each test condition appeared once in every possible position in the sequence. e.g. in the example of three test conditions, condition A is in the first place in group one; in third place of group 2 and in second place of group 3.

Further more, matrix like this type can be expand and genaralized to a $n \times n$ matrix, and in experimental design, it has a formal name: *Latin Squares*, which was inspired by mathematical papers by Leonhard Euler, who used Latin characters as symbols.[WG11]

$n \times n$				
1	2	...	n-1	n
2	3	...	n	1
⋮	⋮	...	⋮	⋮
n-1	n	...	n-3	n-2
n	1	...	n-2	n-1

A simple algorithm to generate the *Latin Square* talked previously is to use circular. The details on how to construct such an *Latin Square* and the proof was talked in [Bra58].

2.2 Balanced Latin Square

The *Latin Square* provides an equally appearance for each condition in each possible position. however, there is still an order effect exists. Recall the example of the 3×3 matrix: B follows A twice, but A follows B only once. That's not a coincidence. Actually a deficiency in Latin squares of order 3 and higher is that conditions precede and follow other conditions an unequal number of times. So an A-B sequence effect is not fully compensated for. We introduce one solution to such deficiency, which is called: *Balanced Latin Squares.*

If we reconstruct the 4×4 square as the following matrix:

Group	Test Condition Sequence			
1	A	B	D	C
2	B	C	A	D
3	C	D	B	A
4	D	A	C	B

In the above balanced Latin Square, the A-B sequence effect as been solved. If we want to construct a $n \times n$ balanced Latin Square, we can follow the matrix below

$$\begin{array}{ccccccc}
1 & 2 & n & 3 & n\text{-}1 & 4 & \dots \\
2 & 3 & 1 & 4 & n & 5 & \dots \\
3 & 4 & 2 & 5 & 1 & 6 & \dots \\
\vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\
n\text{-}1 & n & n\text{-}2 & 1 & n\text{-}3 & 2 & \dots \\
n & 1 & n\text{-}1 & 2 & n\text{-}2 & 3 & \dots
\end{array}$$

Before use the above square to construct the balanced latin squares, one thing must be noticed: balanced latin squares can only be generated for even number conditions. And for a n condition design, the number of participants needed is a $k \times n$, which k is a positive integer.

3 Example

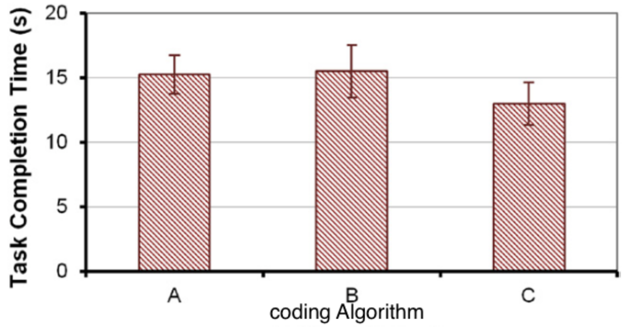
At the end, let's go through a detail example for what we've talked about previously.

In an experiment, there are 3 algorithms to be tested: A. Video frame coded with average optical flow; B. Video frame coded with average sound level; C. don't code the video. And there are also 3 tasks: 1. Birthday Party; 2. Group Skiing; 3. Basketball Game. The experiment was designed and average data for each condition(average value for 3 tasks/condition) is shown as following:

Participant	Test Condition			Group	Mean
	A	B	C		
1	12.5	16.9	12.2	1	μ_A^1
2	14.4	16.3	14.1		
3	...	...	...	2	μ_A^2
4					
5	...	...	...	3	μ_A^3
6					
mean	μ_A	μ_B	μ_C		

If we present the results in a chart:

In this example, if the values of μ_A^1, μ_A^2 and μ_A^3 are all close enough. Thus we can infer there is no order effect in this design on condition A. This conclusion can also apply to condition B and C.



References

[Bra58] James V Bradley. Complete counterbalancing of immediate sequential effects in a latin square design. *Journal of the American Statistical Association*, 53(282):525–528, 1958.

[Mac12] I Scott MacKenzie. *Human-computer interaction: An empirical research perspective.* Newnes, 2012.

[WG11] Walter D Wallis and John George. *Introduction to combinatorics.* CRC Press, 2011.