830 Phase1

Yuqing Zhang

2021/8/16

```
##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##     lowess
```

PHASE I - Factor Screening

Objective:

In the first stage of experiment, our goal is to determine which factors (Tile.Size, Match.Score, Prev.Length, Prev.Type) significantly influence the response (browse time).

Design

We chose 2^{4-1} fractional factorial experiment to save the number of experimental conditions. Each factor is represented by a binary variable {-1, +1} by the default setting in the project. Each experimental condition can be identified by a unique combination of plus and minus ones. The defining relation is $I = ABCD$ if we represent the factor as A, B, C, D. The corresponding x values also satisfy $x_1 x_2 x_3 x_4 = 1$.

We chose 2^{4-1} fractional factorial experiment to save the number of experimental conditions (8), compared to the full factorial experiment (16 experimental conditions). This design has Resolution = IV. This is the maximum resolution we can achieve to mitigate the negative effect of aliasing. It can use the advantage of Principle of effect sparsity.

Data:

Build the design matrix as follows.

By the default low and high setting, we can convert it into the natural unit. Import the matrix into the simulator, we can get the data.

x1	x2	x3	x4
+1	+1	+1	+1
-1	+1	+1	-1
+1	-1	+1	-1
-1	-1	+1	+1
+1	+1	-1	-1
-1	+1	-1	+1
+1	-1	-1	+1
-1	-1	-1	-1

Factor	Low (-1)	High (+1)
Tile.Size (x1)	0.1	0.3
Match.Score (x2)	80	100
Prev.Length (x3)	100	120
Prev.Type (x4)	TT	AC

After obtaining the data, we can convert the natural unit into coded unit and build the X and y matrix.

Then we do the linear regression with the following linear predictor

$$\begin{aligned}
&\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\
&\quad + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 \\
&\quad + \beta_{123} x_1 x_2 x_3 + \beta_{124} x_1 x_2 x_4 + \beta_{134} x_1 x_3 x_4 + \beta_{234} x_2 x_3 x_4 \\
&\quad + \beta_{1234} x_1 x_2 x_3 x_4
\end{aligned}$$

Analysis:

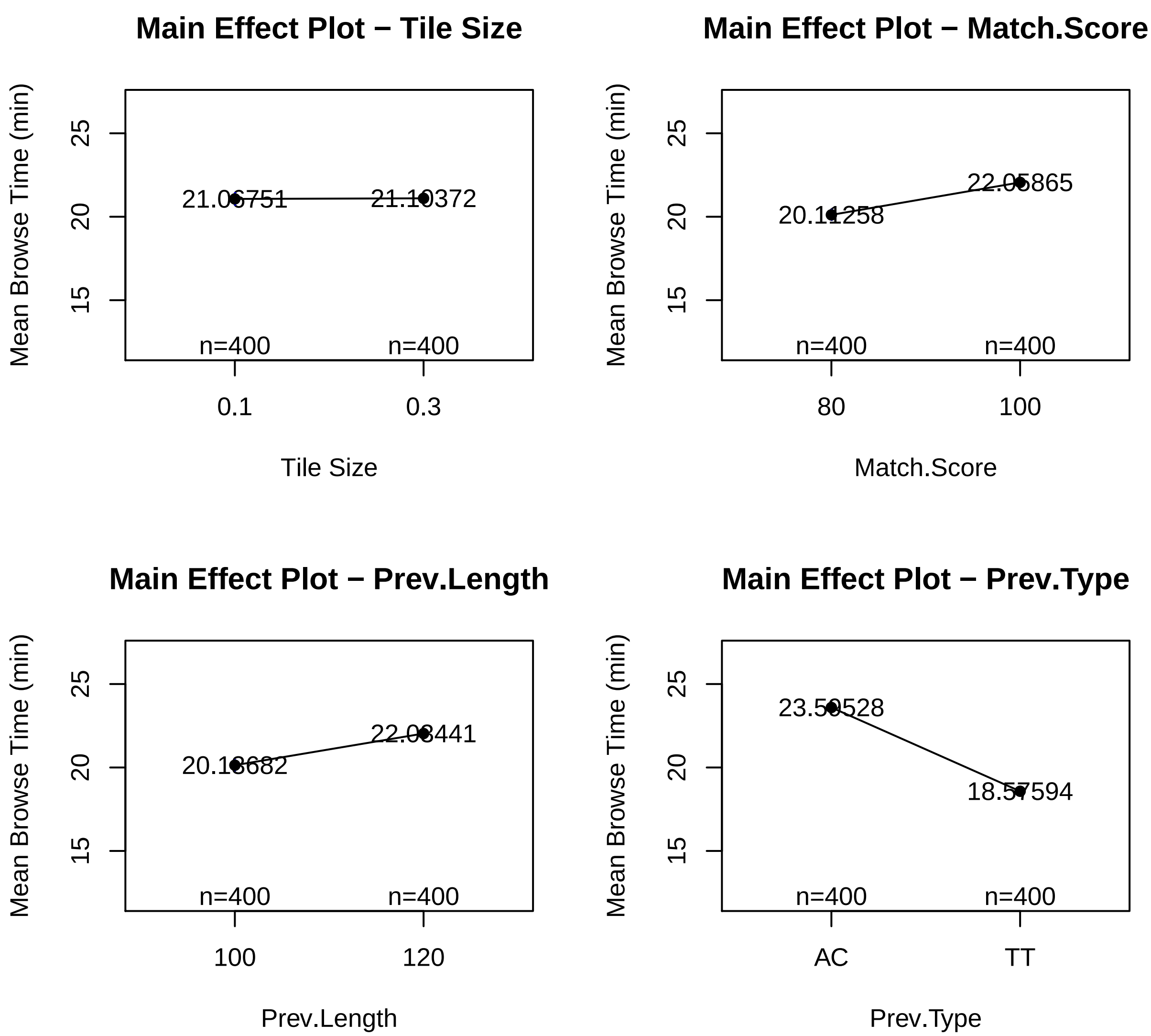
The regression output shows that the following terms are significant:

- main effect: x2 (Match.Score), x3 (Prev.Length), x4 (Prev.Type)
- interaction effect: x2:x3

In addition, we fitted the reduced model with the significant terms $\beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{23} x_2 x_3$

Compare the full model with the reduced model by the anova test. The result shows a P-value with 0.635. We can conclude this reduced model fits as well as the full one.

Here we use main effect plots to help us interpret these effects:



Conclusion:

Through the main effect plots, we can see expect the first plot, the main effect is significant. We can conclude that Match.Score, Prev.Length, Prev.Type can significantly influence the response. The factor (Tile Size) can be considered insignificant, so we ignore it in all subsequent phases of experimentation.