Statapult

E. Croffoot-Suede, S. Kim

4/6/2022

Data

We began by collecting our data. The data were collected on April 7, 2022 using a "statapult." Out of six possible factors, we chose to analyze five of them, leaving the last factor (POST) at the low level for all runs. The factors are as follows.

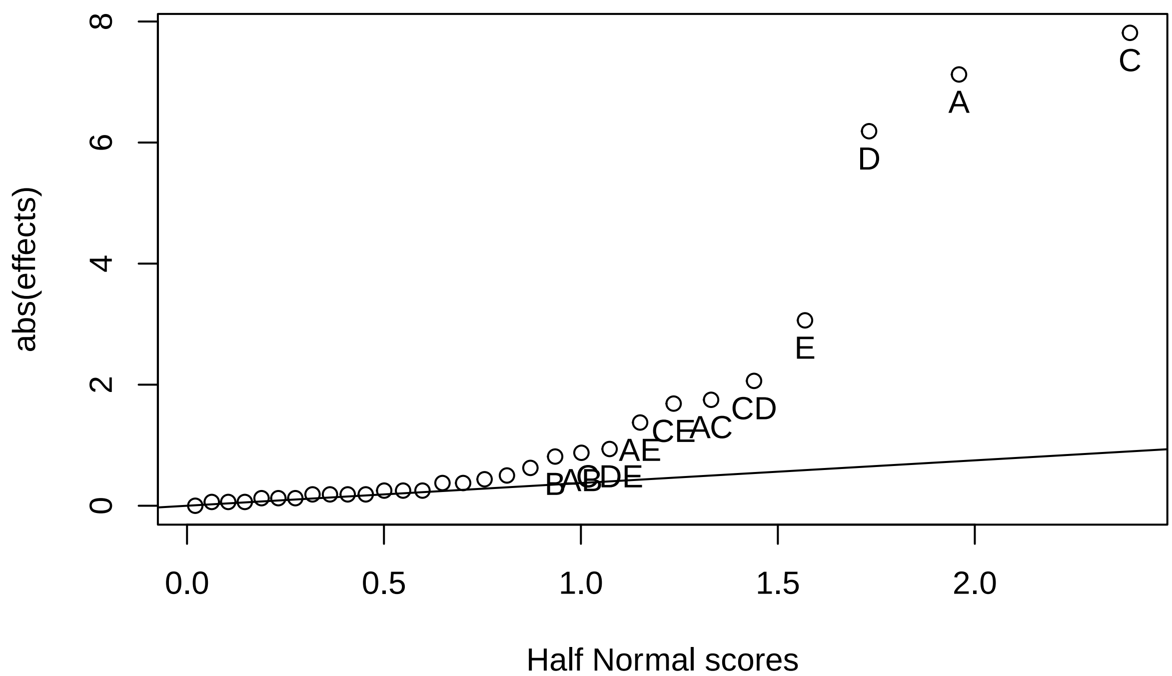
- `arm` - rubber band attachment point on the arm: 1, 3
- `ball` - ball type: ping-pong ball (low), wiffle ball (high)
- `position` - ball position: 1, 3
- `height` - arm draw back height: 1, 3
- `tilt` - catapult tilt: 1, 3

The last factor (`post`) was the rubber band attachment point on the post. It was left at the high level (1) for all runs.

Analysis

Full Fractional Design

We started by analyzing the full 2^5 fractional design. To determine significant effects, we used a half-normal plot on the full dataset with no replicates.




```
## zscore= 0.02089009 0.06270678 0.1046335 0.146745 0.1891184 0.2318344 0.2749777
   0.3186394 0.3629173 0.4079187 0.4537622 0.5005801 0.5485223 0.5977601
   0.6484922 0.7009514 0.755415 0.8122178 0.871771 0.9345893 1.001331 1.072861
   1.150349 1.23544 1.330562 1.439531 1.56892 1.731664 1.959964 2.39398effp=
   1.251781e-15 0.0625 0.0625 0.0625 0.125 0.125 0.125 0.1875 0.1875 0.1875
   0.1875 0.25 0.25 0.25 0.375 0.375 0.4375 0.5 0.625 0.8125 0.875 0.9375 1.375
   1.6875 1.75 2.0625 3.0625 6.1875 7.125 7.8125
```

Based on the half-normal plot, we determined that all main effects A, B, C, D, and E ought to be put into the model, along with the two-way interactions AB, AC, AE, CD, and CE, and the three-way interaction CDE. The three-way interaction term includes the main effects C, D, and E, and the two-way interactions AC, AE, CD, and CE. The two-way interaction term AB includes the main effects A, and B.

The resulting full model with all terms is

$$Y_{\text{Distance}} = C * D * E + A * B + A * E + A * C + C * E$$

To analyze the data, we start by testing for overall significance using a F-test for the model. To do so, we fit two linear models: one with the response (`distance`) by the full model and one with the response by 1 to make our reduced model. Then, we tested the full model against the reduced model using analysis of variance.

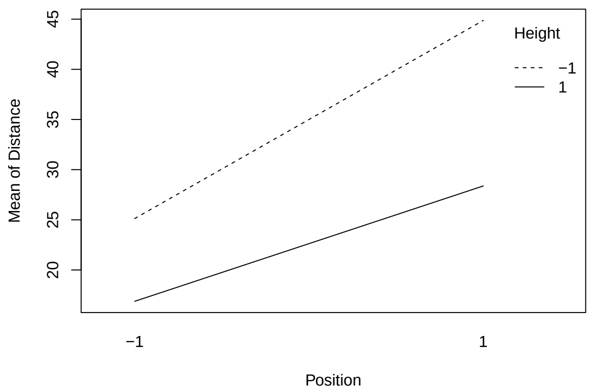
We found that the full model is significantly different with a p-value of at most 2.2×10^{-16} and a F-value of 185.94 on (12, 19).

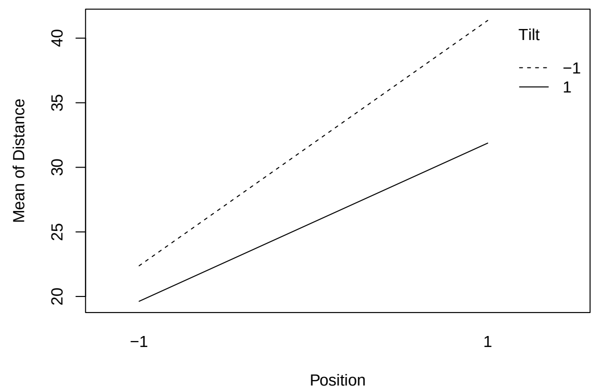
Based on this, we ran a F-test for three-way interactions. We found that the three-way interaction $C * D * E$ was significantly different with a F-value of 11.2797 on (1, 19) degree of freedom and a p-value of $p = 0.003300 < 0.05 = \alpha$.

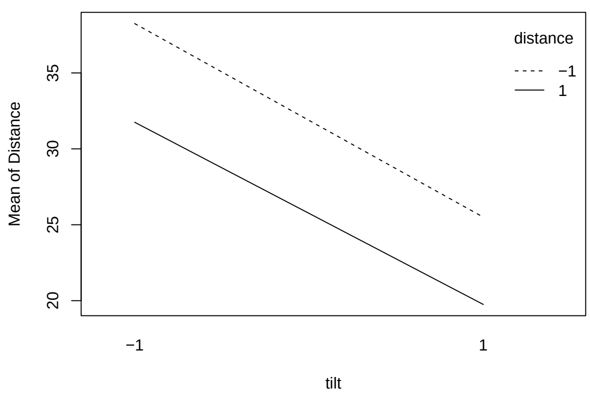
We found that the two-way interaction $A * B$ was significantly different with a F-value of 9.8259 on (1, 19) degree of freedom and a p-value of $p = 0.005457 < 0.05 = \alpha$.

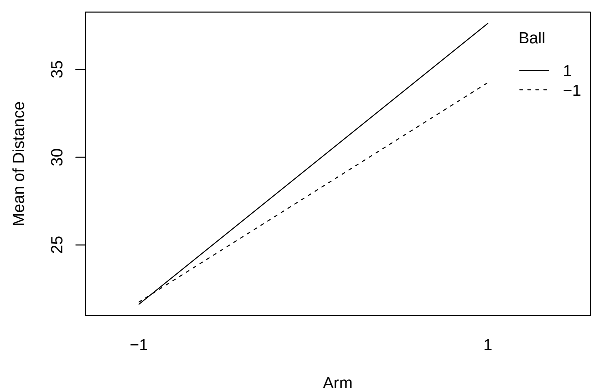
Based on this, we generated contrasts on the three-way interaction cells. We chose to use Bonferroni's correction method on the contrasts because of its higher power on smaller numbers of means, especially compared to Tukey's correction method.

Based on this, we found that to maximize distance, we ought to keep position (C) high, tilt (E) low, and height (D) low, based on the interaction plots between position, tilt, and height.









Since all of the p-values for these contrasts are below our $\alpha = 0.05$, we also determined that this combination of position at the high-level, and tilt and height at the low-level were significantly different from the other combinations.

We also generated contrasts for the two-way $A * B$ cell. Similarly, we chose to use Bonferroni's correction method on the contrasts over Tukey's correction method. Based on this, we found that to maximize distance, we ought to keep ball and arm at the high level to achieve the highest distance based on the interaction plot.

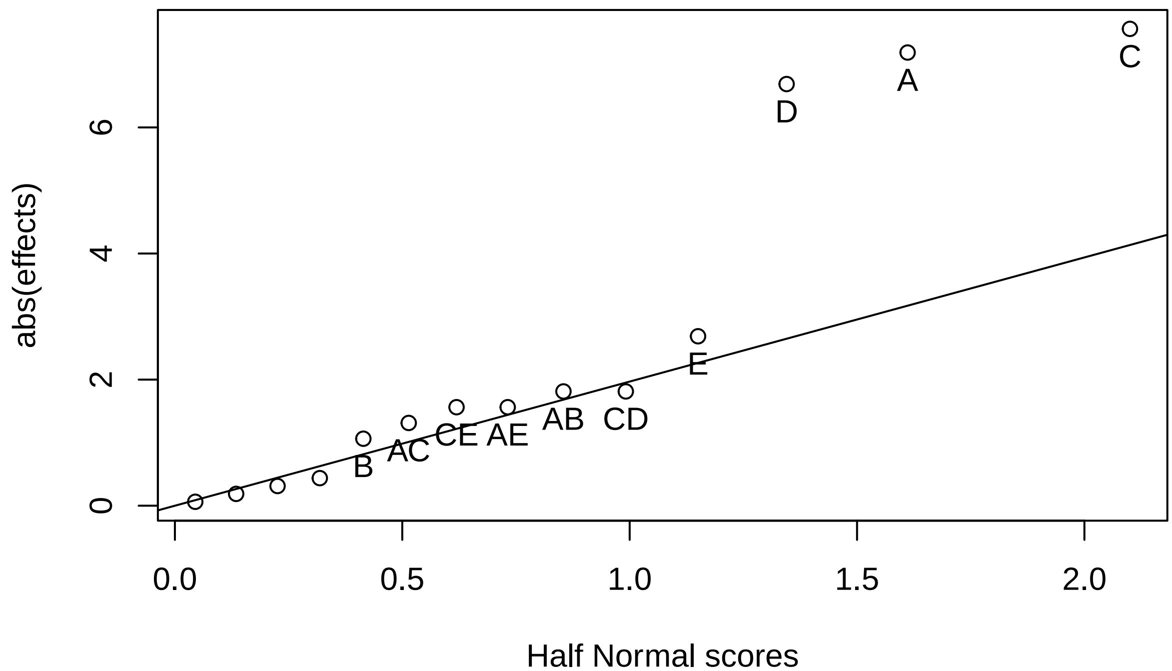
Since all of the p-values for these contrasts are below our $\alpha = 0.05$, we also determined that this combination of ball and arm at the high-level were significantly different from the other combinations.

Half Fractional Design

We then analyzed the same dataset, but subset it to analyze the data using a $\frac{1}{2}$-fractional design. That is, we used a 2^{5-1}_{V} design to analyze a subset of our data.

To subset the data, we found the sixteen runs in the half-fractional design and identified them by what factor they were. We then got the index number of that observation to pull those observations into a new dataset.

We then used a half-normal plot on the fractional dataset with no replicates.




```
## zscore= 0.04477618 0.1346898 0.225708 0.3186394 0.4144133 0.5141561 0.6193068
   0.7318081 0.8544474 0.9915265 1.150349 1.345167 1.611169 2.100165effp= 0.0625
   0.1875 0.3125 0.4375 1.0625 1.3125 1.5625 1.5625 1.8125 1.8125 2.6875 6.6875
   7.1875 7.5625
```

Based on the half-normal plot, we determined that all main effects A, B, C, D, and E, ought to be put into the model, along with the two-way interactions AB, AC, AE, CD, and CE. The two-way interaction terms include all of the main effects.

The resulting full model with all terms is

$$Y_{\text{Distance}} = A * B + A * C + A * E + C * D + C * E$$

To analyze the data, we start by testing for overall significance using a F-test for the model. To do so, we fit two linear models, one with the response (`distance`) by the full model and one with the response by 1 to make our reduced model. We then tested the full model against the null model using analysis of variance.

We found that the full model is significantly different with a p-value of 4.549×10^{-6}. We then found that the two-way interaction $A * B$ was significantly different with a F-value of 45.215 on (1, 5) degrees of freedom and a p-value of $p = 0.0011024 < 0.05 = \alpha$. The two-way interaction $A * C$ was significantly different with a F-value of 23.710 on (1, 5) degrees of freedom and a p-value of $p = 0.0045958 < 0.05 = \alpha$. The two-way interaction $A * E$ was significantly different with a F-value of 33.602 on (1, 5) degrees of freedom and a p-value of $p = 0.0021528 < 0.05 = \alpha$. The two-way interaction $C * E$ was significantly different with a

F-value of 33.602 on (1, 5) degrees of freedom and a p-value of $p = 0.0021528 < 0.05 = \alpha$. The two-way interaction $C * D$ was significantly different with a F-value of 45.215 on (1, 5) degrees of freedom and a p-value of $p = 0.0011024 < 0.05 = \alpha$.

Based on this, we generated contrasts on the two-way interaction cells. Again, we chose to use Bonferroni's correction method on the contrasts because of its higher power on smaller numbers of means, especially compared to Tukey's correction method.

Based on the contrasts and the interaction plots, we determined that to maximize distance, we ought to keep arm (A) high, ball (B) high, position (C) high, and height (D) low. There is no significant difference between having the tilt at the high or low level.

Discussion

Full versus Half Fractional Designs

Randomization

To minimize the effect of rubber band fatigue on the force that the statapult pulled with, we chose to randomize the runs. To do so, we generated a vector of 32 random numbers using the default seed based on system time in Python. After writing each of the factor levels out, we sequentially assigned the each level combination to the next random number in the vector.

Conclusion

We believe that the best way to maximize distance from the statapult is to use a wiffle ball at arm position 3 position at position 3, height at position 1, and tilt at position 1. This