Design of Experiment to Maximize Temple Run Distance

Team Members:
Liye Pan
Chenxin Wang
Tingting Jiang
Mohit Ravi Ghatikar

Contents:

Introduction:

Temple Run is one of the most iconic games in the modern smartphone era. It has been downloaded over a billion times till date on both Android and iPhone devices. The premise of the game is simple. It is an endless running game where the character runs along a defined path, collecting points along the way. The only way the game ends is when the character falls off the path or if the character stumbles into a trap.

Motivation for the Project:

Our objective in this experiment is to determine factors that impact the final score. Besides applying experimental design principles on a real world gaming application, we also hope that this experience will be useful in dealing with other business problems especially in the area of marketing.

Choice of Factors and Levels:

After playing the game for some time and researching online, we came up with a bunch of factors that are relevant in obtaining high scores. The eight factors chosen are:

1) **Device**: We chose the two most popular devices for the game.

Levels: iPhone (+) and iPad (-)

2) **Sound**: We would like to know if the soundtrack from the game makes any difference on the final score.

Levels: On (+) and Off (-)

3) **Gender:** We would like to find out if gender is a significant factor.

Levels: Female (+) and Male (-)

4) **Screen Brightness:** The brightness level of the device may impact the final score. We invariably chose low brightness and high brightness levels.

Levels: 100% (+) and 30% (-)

5) **New player**: Like any other mobile game, the more you play the better you get. More experienced gamers usually figure out the idiosyncrasies in the game which might be unknown to new gamers. We wanted to test this hypothesis by choosing two levels, a person who has already played this game frequently and a person who has never played the game before. Ultimately we want to figure out if experienced gamers have an edge over new players.

Level: Yes (+) and No (-)

6) **Location**: External factors can also have an influence over the final score. We chose two levels that indicate noise levels. The level ‘library’ indicates negligible noise whereas the level ‘outside’ indicates noisy atmosphere.

Level: Library (+) and Outside (-)

7) **Posture:** The player can play the game on the device in different postures. We considered sitting and standing as our two levels.

Level: Sitting (+) and Standing (-)

8) **Hand Position:** The game of Temple run is designed in such a way such that the player has to tilt their device left or right, in order to move the character to the left or right of the screen to collect coins or avoid objects. This can done by using only one of your hands or both the hands. These are the two chosen levels.

Level: One (+) and Two (-)

Selection of the Response Variable:

Our objective is to maximize the high score given a set of variables. High scores are determined by various aspects including the distance ran by the character, coins collected, special utilities etc. Since distance is the only factor that we can control, we chose the total distance ran in meters as our response variable.

Figure – 1: The final score table

Design of Experiment:

There are eight main factors to be considered in this experiment as described above. Each of the factor has two levels. Hence, we decided to go ahead with 2^{8-4} fractional factorial design to reduce the complexity of the experiment. This experiment has a resolution of IV, with 16 runs. We can estimate the main factors and some two factor interactions. The experiment run was randomized, where each run had four replications. The final response variable was the average of the four replicates.

Our defining relation is as follows:
I = Posture * Hand Position * Device * Sound

The entire randomized experimental log is given below:

	No	Pattern	Posture	Device	Sound	Screen Brightness	New Player	Gender	Location	Hand	No 1	No 2	No 3	No 4	response
1	1	++++++++	Sit	Iphone	On	Bright	Yes	Female	Library	One	821	876	1057	1066	955
2	2	+++----+	Sit	Iphone	On	Dark	No	Male	Outside	One	2748	3190	4086	2706	3182.5
3	3	-+-+-+-+	Stand	Iphone	Off	Bright	No	Female	Outside	One	2039	2323	1821	2407	2147.5
4	4	---++++-	Stand	Ipad	Off	Bright	Yes	Female	Library	Two	450	94	662	300	376.5
5	5	-+--+-++	Stand	Iphone	Off	Dark	Yes	Male	Library	One	1457	2259	3176	2420	2328
6	6	-++--++-	Stand	Iphone	On	Dark	No	Female	Library	Two	4060	3653	3568	3339	3655
7	7	++-+--+-	Sit	Iphone	Off	Bright	No	Male	Library	Two	6748	5456	3748	4785	5184.25
8	8	+-+-+-+-	Sit	Ipad	On	Dark	Yes	Male	Library	Two	980	591	977	1030	894.5
9	9	+--++--+	Sit	Ipad	Off	Bright	Yes	Male	Outside	One	345	536	427	392	425
10	10	--------	Stand	Ipad	Off	Dark	No	Male	Outside	Two	1465	1715	2978	2693	2212.75
11	11	++--++--	Sit	Iphone	Off	Dark	Yes	Female	Outside	Two	1063	1135	1182	1305	1171.25
12	12	--+-++-+	Stand	Ipad	On	Dark	Yes	Female	Outside	One	733	542	420	495	547.5
13	13	+----+++	Sit	Ipad	Off	Dark	No	Female	Library	One	793	454	1263	1131	910.25
14	14	-++++---	Stand	Iphone	On	Bright	Yes	Male	Outside	Two	1734	1701	998	593	1256.5
15	15	+-++-+--	Sit	Ipad	On	Bright	No	Female	Outside	Two	1838	1587	1689	1650	1691
16	16	--++--++	Stand	Ipad	On	Bright	No	Male	Library	One	935	1186	2347	1263	1432.75

Figure – 2: Experimental log

Statistical Analysis of the Data:

1) **Normality Test:**

Before analyzing the results, we verified the normality of each run using Shapiro – Wilk test. We found out that all runs are normally distributed since their p-values were greater than 0.05.

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	955	755.99741	1154.0026
Dispersion	σ	125.06265	70.846684	466.30201

-2log(Likelihood) = 48.9820268213846

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.839626	0.1943

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Figure – 3: Shapiro – Wilk test conducted for the first run.

2) **Half – Normal Plot:**

To get a sense of significant variables, we analyzed the half-normal plot. It is shown below:

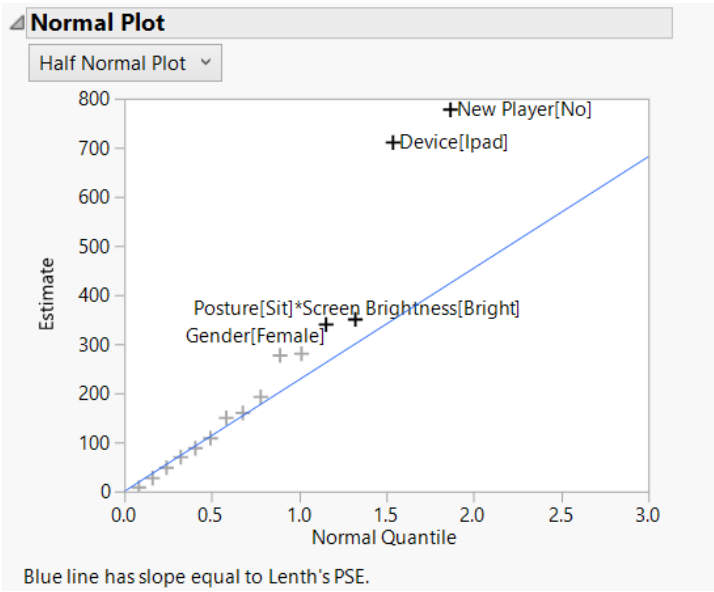


Figure – 4: Half – Normal Plot.

We can notice that the factors ‘Device’ and ‘New Player’ are significantly far away from the straight line. This suggests that these factors must be significant. We can also see that the factor ‘Gender’ and the two-way interaction between Posture and Screen Brightness are not located on the straight line. This suggest that they may be significant and further analysis is required to validate these results.

3) **Pareto Plot:**

A Pareto plot is useful in highlighting the important factors in descending order. They are plotted below:

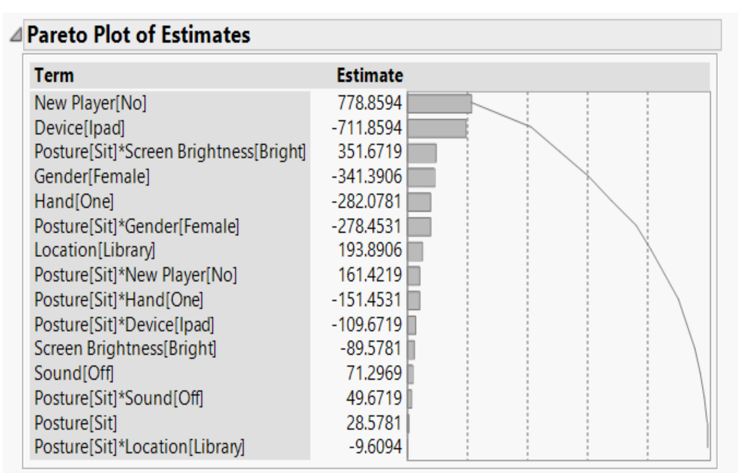


Figure – 5: Pareto Plot

New player and Device are clearly the most important factors. The other main factors and two way interactions are insignificant.

4) **Parameter Estimates:**

Through the half-normal plot and Pareto plot, we have an intuition of what factors are significant in describing the response variable. We can now finally look at the parameter estimates table to corroborate our hypothesis. The parameter estimates table is given below:

Sorted Parameter Estimates

Term		Estimate	Relative Std Error	Pseudo t-Ratio	Pseudo p-Value
New Player[No]		778.85938	0.25	3.43	0.0050*
Device[Ipad]		-711.8594	0.25	-3.13	0.0086*
Posture[Sit]*Screen Brightness[Bright]	Biased	351.67188	0.25	1.55	0.1476
Gender[Female]		-341.3906	0.25	-1.50	0.1588
Hand[One]		-282.0781	0.25	-1.24	0.2381
Posture[Sit]*Gender[Female]	Biased	-278.4531	0.25	-1.23	0.2438
Location[Library]		193.89063	0.25	0.85	0.4101
Posture[Sit]*New Player[No]	Biased	161.42188	0.25	0.71	0.4909
Posture[Sit]*Hand[One]	Biased	-151.4531	0.25	-0.67	0.5176
Posture[Sit]*Device[Ipad]	Biased	-109.6719	0.25	-0.48	0.6380
Screen Brightness[Bright]		-89.57813	0.25	-0.39	0.7003
Sound[Off]		71.296875	0.25	0.31	0.7590
Posture[Sit]*Sound[Off]	Biased	49.671875	0.25	0.22	0.8306
Posture[Sit]		28.578125	0.25	0.13	0.9020
Posture[Sit]*Location[Library]	Biased	-9.609375	0.25	-0.04	0.9670
Device[Ipad]*Sound[Off]	Zeroed	0	0	.	.
Device[Ipad]*Screen Brightness[Bright]	Zeroed	0	0	.	.
Device[Ipad]*New Player[No]	Zeroed	0	0	.	.
Device[Ipad]*Gender[Female]	Zeroed	0	0	.	.
Device[Ipad]*Location[Library]	Zeroed	0	0	.	.
Device[Ipad]*Hand[One]	Zeroed	0	0	.	.
Sound[Off]*Screen Brightness[Bright]	Zeroed	0	0	.	.
Sound[Off]*New Player[No]	Zeroed	0	0	.	.
Sound[Off]*Gender[Female]	Zeroed	0	0	.	.
Sound[Off]*Location[Library]	Zeroed	0	0	.	.
Sound[Off]*Hand[One]	Zeroed	0	0	.	.
Screen Brightness[Bright]*New Player[No]	Zeroed	0	0	.	.
Screen Brightness[Bright]*Gender[Female]	Zeroed	0	0	.	.
Screen Brightness[Bright]*Location[Library]	Zeroed	0	0	.	.
Screen Brightness[Bright]*Hand[One]	Zeroed	0	0	.	.
New Player[No]*Gender[Female]	Zeroed	0	0	.	.
New Player[No]*Location[Library]	Zeroed	0	0	.	.
New Player[No]*Hand[One]	Zeroed	0	0	.	.
Gender[Female]*Location[Library]	Zeroed	0	0	.	.
Gender[Female]*Hand[One]	Zeroed	0	0	.	.
Location[Library]*Hand[One]	Zeroed	0	0	.	.

Figure – 6: Parameter estimates table

As expected, the main factors 'New player' and 'Device' with p-values of 0.005 and 0.0086 respectively are significant at a 95% confidence level. The main factor 'Gender' and two-way interaction of Posture and Screen brightness would have been significant if we had set an alpha level of 15%. Ultimately, we don't consider them because we are interested in dealing with factors that satisfy an alpha level of 5%.

5) **Regression Model:**

Based on these two factors, we can build a regression model to make a prediction of how far one can go in this game:

$$Distance = 1773.14 + 711.86 \times Device - 778.86 \times NewPlayer$$

The statistical summary of the model is shown below. Since we picked only significant factors, all the variables have p values less than 0.05. The overall model is also significant since the p-value for its F-ratio is also less than 0.05. The adjusted R square is close to 0.64. It implies that our model can explain 64% of the response variable.

Summary of Fit

RSquare	0.684815
RSquare Adj	0.636325
Root Mean Square Error	794.1519
Mean of Response	1773.141
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	17813851	8906926	14.1228
Error	13	8198804	630677	**Prob > F**
C. Total	15	26012655		0.0006*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	1	1240578.3	1240578	2.1395
Pure Error	12	6958226.1	579852	**Prob > F**
Total Error	13	8198804.4		0.1692
				Max RSq
				0.7325

Parameter Estimates

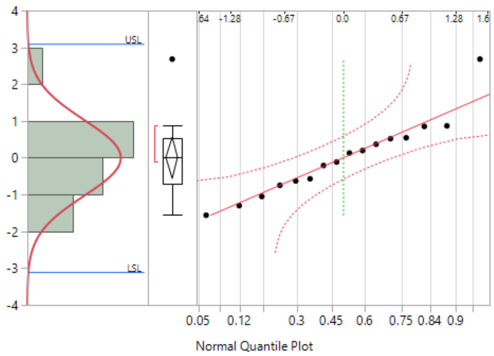
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1773.1406	198.538	8.93	<.0001*
Device[Ipad]	-711.8594	198.538	-3.59	0.0033*
New Player[No]	778.85938	198.538	3.92	0.0017*

Figure – 7: ANOVA table

a) **Analysis of Residuals:**

The residual analysis is essential to a regression model because it is based on the assumption that the error of the model is randomized and normally distributed. We analyzed the Studentized residual which is basically a division of the residual by an estimate of its standard deviation. This is useful to detect outliers. This analysis is present in the fit model section of SAS JMP.

Normal Quantile Plot and **Normal Test of Studentized Residual** results are shown below. The plot shows that the residuals are generally normally distributed. Additionally, the residual distribution satisfies the Shapiro-Wilk Test with a P value of 0.35. So we accept the hypothesis that the residuals follow normal distribution.



Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	-2.78e-17	-0.550338	0.5503379
Dispersion	σ	1.0327956	0.7629311	1.5984486

-2log(Likelihood) = 45.4386494007507

Capability Analysis

Quantiles: Uncentered and Unscaled

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.940223	0.3518

Figure – 8: Normal Q-Q plot and Shapiro-Wilk test for Studentized residuals.

b) **Outlier Detection:**

The following graph below describes the relationship between **predicted distances and actual distances Vs the run order**. Similarly, we also a graph which describes **the studentized residual Vs run order**. Clearly, we can find an outlier in our sample. For run no – 7, the error term of 2000 is pretty high. Correspondingly, the point marked in red for run no – 7 has a studentized residual larger than 2. It may be possible that players in run no – 7 are all experienced gamers which might explain the variation in predictions.

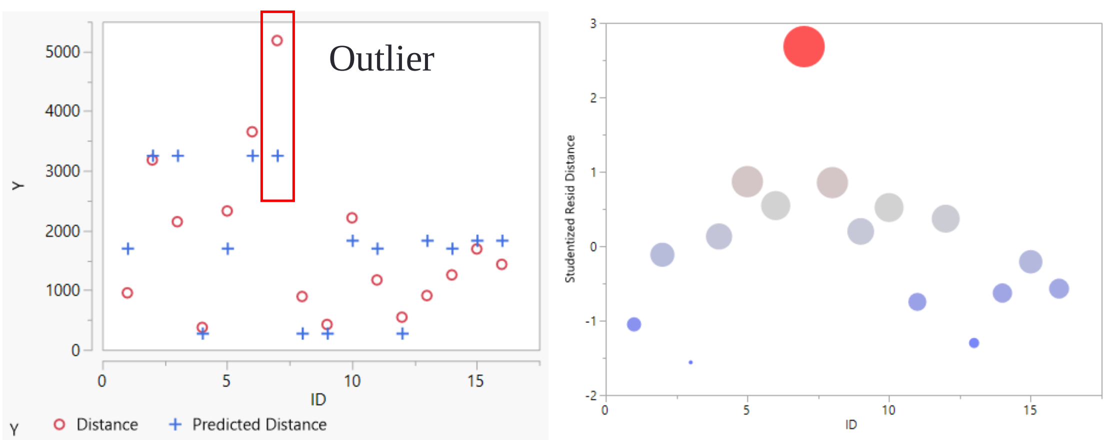


Figure – 9: 1) Left graph - Predicted distances and actual distances Vs the run order.
2) Right graph - Studentized residual Vs run order.

The plot of **Studentized residual Vs actual distance** is shown below. Instead of being randomly distributed, the points seem to have a linear relationship with distance, especially for those points that have distances between 1000 and 2500. There may be two main reasons for this unexpected result. Firstly, we only have 16 samples to fit the model in total. Thus the sample size is relatively too small for fitting a regression model. Secondly, we dropped most of the factors and only retained two variables in our final model. Although the other factors are not significant, they might still contribute to the final score in some way. As a result, some part of the response variable cannot be explained by the two significant factors selected. This kind of error is also shown in the R square value. So, there is a kind of trade-off between the feature selection and prediction accuracy in the regression model.

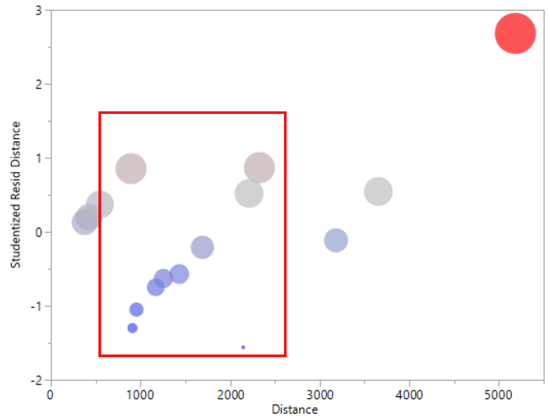


Figure – 10: A plot of studentized residual Vs actual distance

Studentized residual Vs significant factors are also plotted below. Besides excluding the outlier, all the other residuals seem distributed randomly around zero.

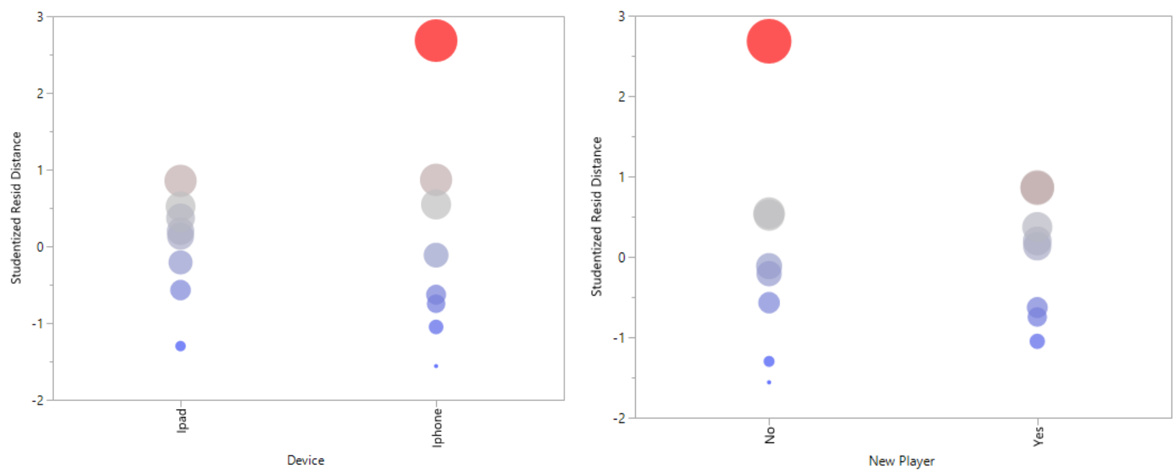


Figure – 11: A plot of studentized residual Vs significant factors

Results and Conclusion:

From our regression equation, we can vary the levels for the significant factors to predict the maximum distance. The calculations are shown below:

Factor	(-) Level	(+) Level
Device	Ipad	Iphone
New Player	No	Yes

Figure – 12: Significant factors and its levels.

Levels		Distance
-	-	1840.14
+	+	1706.14
-	+	282.42
+	-	3263.86

Figure – 13: Maximum Distance Prediction

From the above table, we can see that, when the device is iPhone and if the player has played this game before, the predicted distance value of 3263.86 is the highest. This has several implications. Since other factors are insignificant, we can choose any levels for these factors.

For example, the brightness can be low (30%) to save battery life or the sound can be turned off as it has no major impact on the final score. This also implies that the player can choose whichever posture or hand positions that they are most comfortable with.

References:

1) https://en.wikipedia.org/wiki/Temple_Run

2) https://en.wikipedia.org/wiki/Studentized_residual

3) Testing 1-2-3 Experimental Design with Applications in Marketing and Service Operations, Ledolter and Swersey, Stanford Business Books, 2007