## E1

```
BG_Vars_Summary = subset(ux1_set1, select = Age:Games)
summary(BG_Vars_Summary)
sd(BG_Vars_Summary$Age)
```

Age:
Min = 9.00; Max = 54.00; SD = 9.64; Mean = 28.02; Median = 28.00.

Gender:
Man = 53; Woman = 34; Other = 3.

Playing habits:
Seldom or never = 16; Monthly = 3; Weekly = 15; A few times a week = 44; Daily = 12.

Out of 90 participants, there are 53 men, 34 women and 3 other. Their playing habits varies from Seldom or never – 16 to Daily – 12 with most participants playing A few times a week – 44 and least playing Monthly – 3. There also were 15 participants who are playing weekly. The age of participants varies from 9 at the minimum to 54 at the maximum, with the mean = 28.02, median = 28 and standard deviation = approx. 9.64

## E2

```
Competence <- rowMeans(ux1_set1[,8:11])
Fear <- rowMeans(ux1_set1[,12:14])
ux1_set1 <- cbind(ux1_set1, Competence)
ux1_set1 <- cbind(ux1_set1, Fear)


head(ux1_set1[c("C1", "C2", "C3", "C4", "Competence", "F1", "F2", "F3",
"Fear")])
```

```
  C1 C2 C3 C4 Competence F1 F2 F3     Fear
1  4  5  5  4       4.50  5  5  3 4.333333
2  5  6  3  4       4.50  3  5  3 3.666667
3  6  6  3  3       4.50  6  6  5 5.666667
4  2  2  3  4       2.75  3  3  2 2.666667
5  5  5  6  6       5.50  6  2  6 4.666667
6  4  2  5  5       4.00  4  6  4 4.666667
```

## E3

```
T_TestSubsetA = subset(ux1_set1, Version != 3)
T_TestSubsetB = subset(ux1_set1, Version != 2)
t.test(Game.overs ~ Version, data = T_TestSubsetA)
t.test(Game.overs ~ Version, data = T_TestSubsetB)
```

Data from 60 players was analyzed, 30 players per game version 1 and 2.

There have been significantly more game overs ($t(47) = 4.605$, $p < 0.001$) encountered by the players who have been playing version 1 (mean = 2.83) than version 2 (mean = 1.07).

Further data from 60 players was analyzed, 30 players per game version 1 and 3.

There have been significantly more game overs ($t(41) = 6.133$, $p < 0.001$) encountered by the players who have been playing version 1 (mean = 2.83) than version 3 (mean = 0.60).

## E4

```
cor.test(ux1_set1$Competence, as.numeric(ux1_set1$Games))
cor.test(ux1_set1$Game.overs, as.numeric(ux1_set1$Games))
cor.test(ux1_set1$Competence, ux1_set1$Game.overs)
```

Data from 90 players was analyzed to see if there were any correlation between a competence level and the amount of games played, where a moderately positive correlation was found ($r(88) = 0.389$, $p < 0.001$). Further analysis of whether there was any correlation between a number of game overs and the amount of games played, as well as if there was any correlation between a level of competence and the number of game overs revealed that there was no significant correlation.

## E5

```
Aov_Test_Results_GameVersions <- aov(Game.overs ~ Version, ux1_set1)
Aov_Test_Results_Fear <- aov(Fear ~ Version, ux1_set1)
summary(Aov_Test_Results_GameVersions)
summary(Aov_Test_Results_Fear)
summary(glht(Aov_Test_Results_GameVersions, linfct = mcp(Version = "Tukey")))
summary(glht(Aov_Test_Results_Fear, linfct = mcp(Version = "Tukey")))
```

There was a statistically significant difference of the number of game overs between versions as determined by one-way ANOVA ($F(2,87) = 24.27$, $p < 0.001$)). Post-hoc test further revealed that the significant difference was between version 2 and 1 ($p < 0.001$) as well as 3 and 1 ($p < 0.001$). Difference between group 1 and 3 was statistically insignificant.

There was a statistically significant difference of the fear level of the participants between versions as determined by one-way ANOVA ($F(2,87) = 54$, $p < 0.001$)). Post-hoc test further revealed that the significant difference was between all version ($p < 0.001$).

## E6

```
pwr.r.test(power=.8,sig.level=0.05, r=0.25)
pwr.r.test(power=.8,sig.level=0.05, r=0.1)
pwr.t.test(d=.1,power=.8,sig.level=0.05)
pwr.t.test(d=.8,power=.8,sig.level=0.05)
pwr.anova.test(f=0.15,power=.8,sig.level=0.05, k=3)
pwr.anova.test(f=1,power=.8,sig.level=0.05, k=3)
```

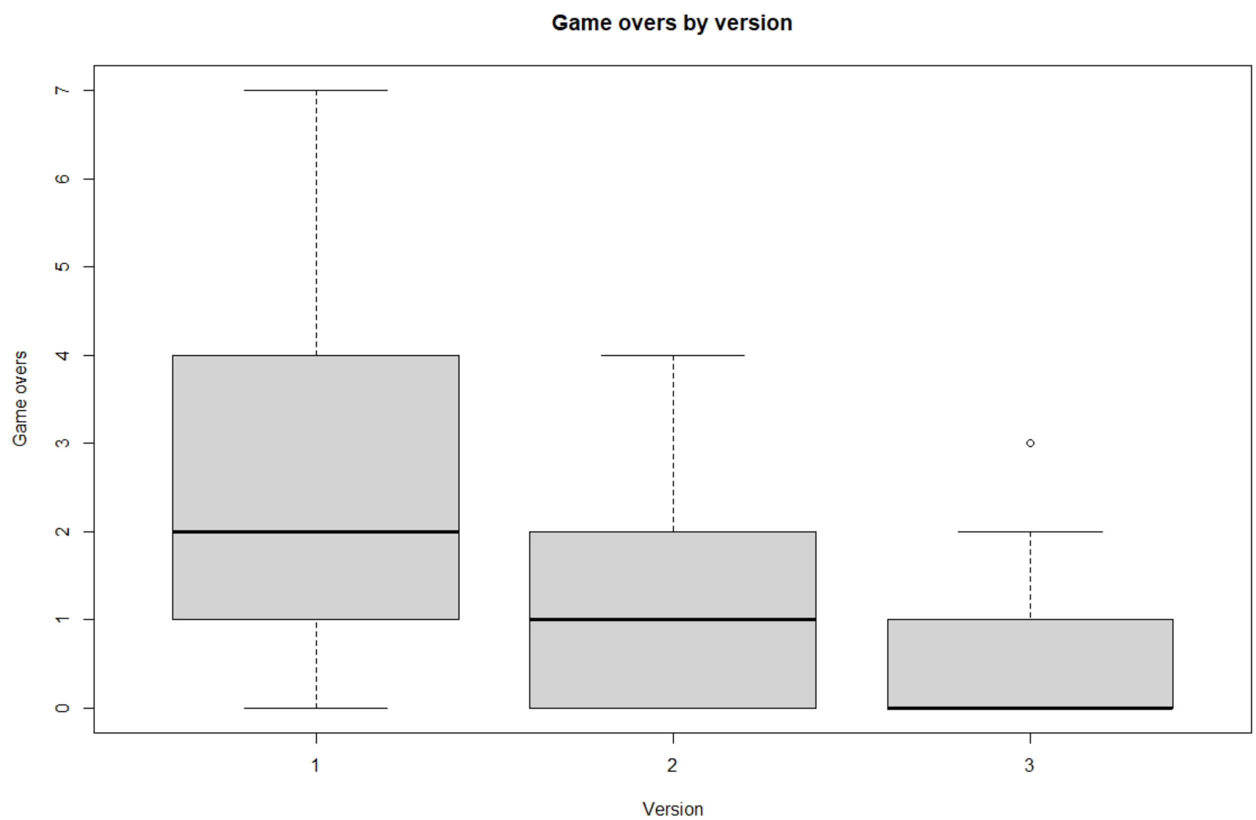
1. If $r = 0.25$ then n should be > 122; If $r = 0.1$ then n should be > 781
2. If $d = 0.1$ then n should be > 1570 * Groups; If $d = 0.8$ then n should be > 25 * Groups
3. If $f = 0.15$ and $k = 3$ then n should be > 143; If $f = 0.15$ and $k = 3$ then n should be > 4

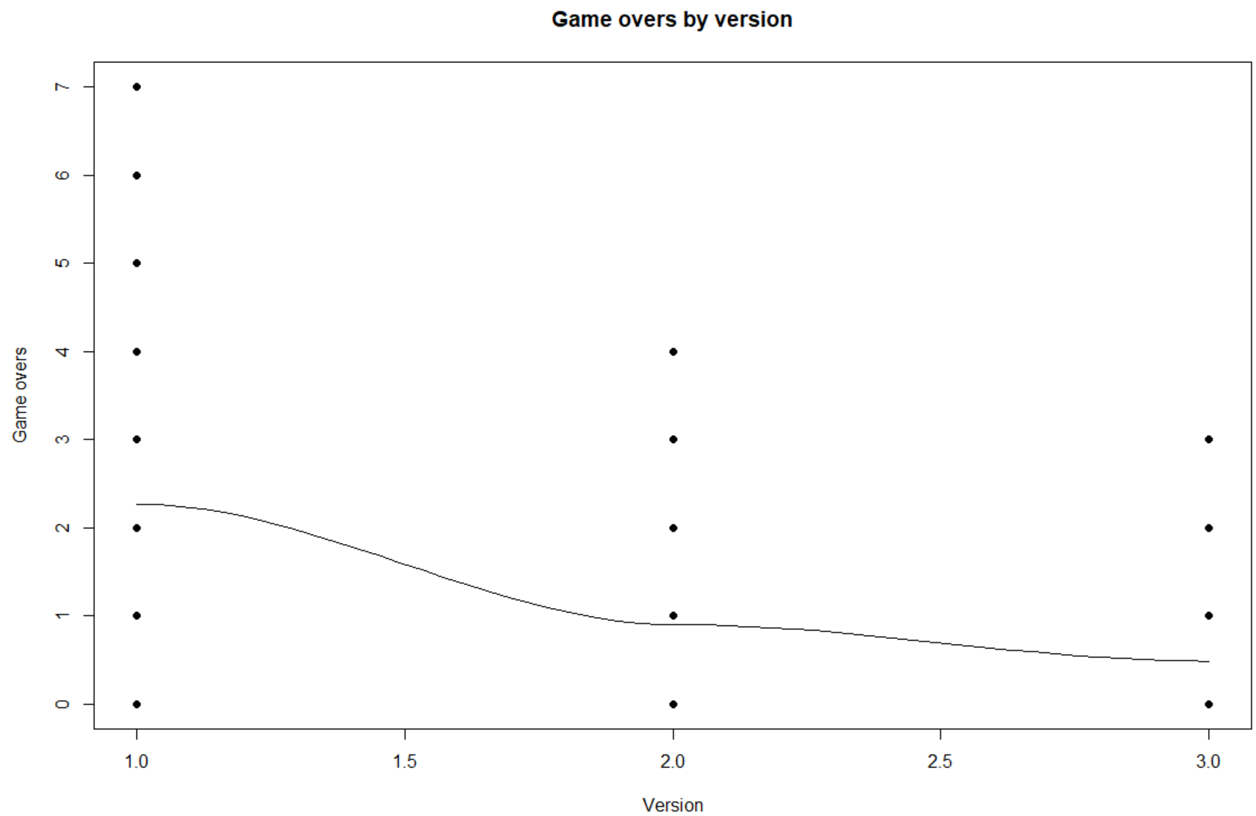
## E7

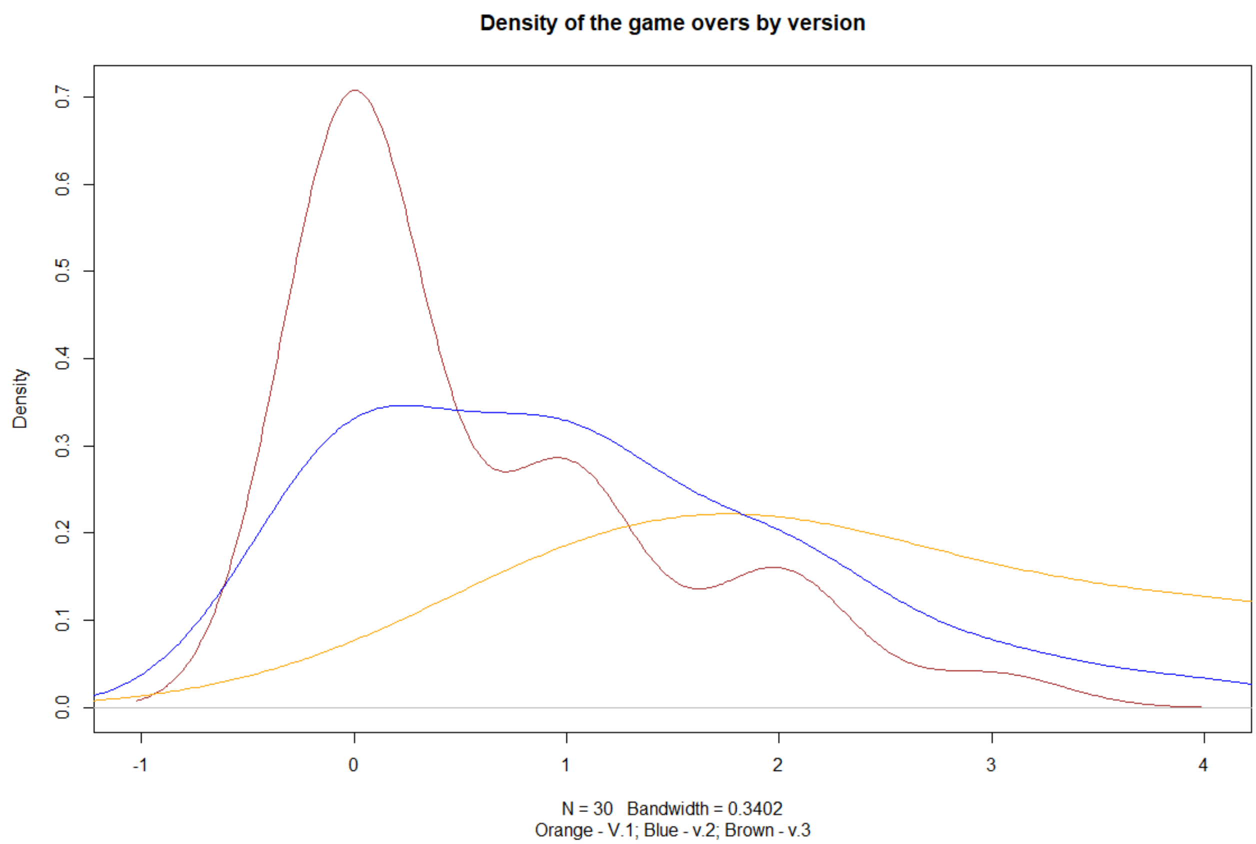
```
Chisq_Subset_AB = table(droplevels(subset(ux1_set1, select = c("Version",
"Completed"), Version != 3)))
Chisq_Subset_AC = table(droplevels(subset(ux1_set1, select = c("Version",
"Completed"), Version != 2)))
chisq.test(Chisq_Subset_AB)
chisq.test(Chisq_Subset_AC)
```

A chi-square test showed that there was no significant difference between completed games and the game version when testing between both version 1 and 2, as well as version 1 and 3.

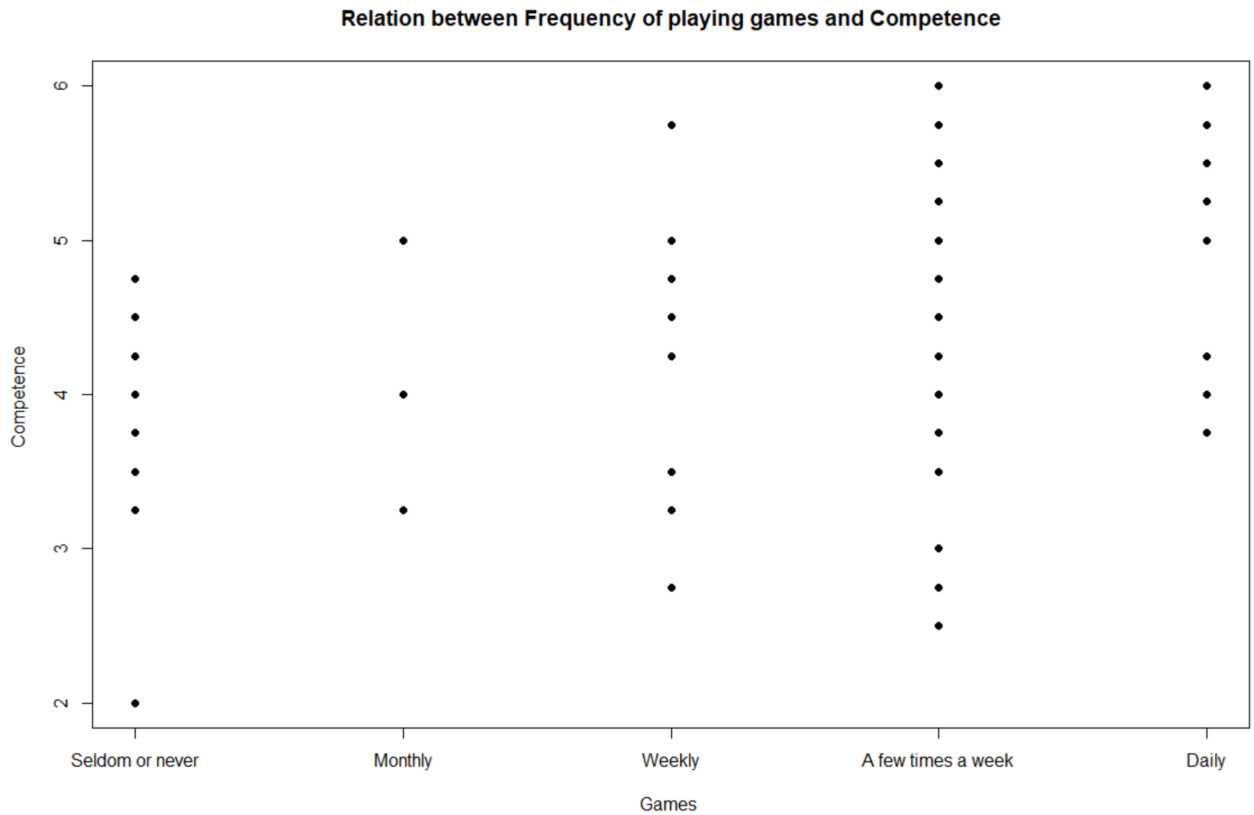
## E8




Game overs by version
Game overs
Version
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
6
7


Density of the game overs by version
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
-1
0
1
2
3
4
N = 30 Bandwidth = 0.3402
Orange - V.1; Blue - v.2; Brown - v.3

E (




## E9

In short, there are five main practices a researcher needs to adhere to when doing any research.

These are:

- Obtain an informed consent.
- Minimize the risk of harm.
- Protect anonymity and confidentiality.
- Avoid using deceptive practices.
- Provide the right to withdraw from the study.

While obvious, these practices are not universally applicable to each and every study. It can be difficult or even impossible to obtain an informed consent from the participants during certain studies. Reason being that in some cases, the act of obtaining the consent would pose a risk of altering the outcome of the study. In other cases, such as when observing the participants in a naturalistic setting – for example going about their daily routine – it can be impossible to ask every single person that is being observed for their consent. It also important to note that while obtaining the consent itself, the researcher should strive to provide as much information about the study as required by the participants, such as what is expected from the participants (without telling too much, so that not to affect the outcome – more on that later) but also try to leave out any material information as not to coerce the participants into partaking in the study.

The researcher should always try to minimize the risk of harm. Whenever harm could be a potential outcome during the course of the study – it should be clearly justified and the participants should be informed. Some types of harm are: physical harm, psychological distress or discomfort, social disadvantage, harm to participants' financial status, privacy or anonymity.

As it can be seen from the summary above – among other types of harms, participants can suffer financially or their privacy could be jeopardized. Because of this a great care should be taken when gathering, storing and working with the data, especially that of a sensitive character. While not always possible, the study should try to protect the identity of its participants. That includes, but is not limited to: changing the names and geolocations of the participants. It is also crucial to think if the data that cannot be altered, i.e. age groups can be used to identify the persons behind it. For example, if the researchers are sending out the questionnaire (for a quantitative research) to a small company and later they group up the questions by age, there might be only one employee in a certain age bracket (i.e. 70+ years), which could make this person identifiable.

On the topic of the deceptive practices – they seem to be controversial. On one hand, the researchers should always strive to get a consent and do no harm. On the other – some research requires the researchers to act covertly, making up a false story or misrepresenting the purpose of the study as not to degrade the quality of the findings.

Lastly, the participants should always be provided a right to withdraw when applicable. This would ensure that even if participants come in a harm's way, they still retain agency to avoid it.

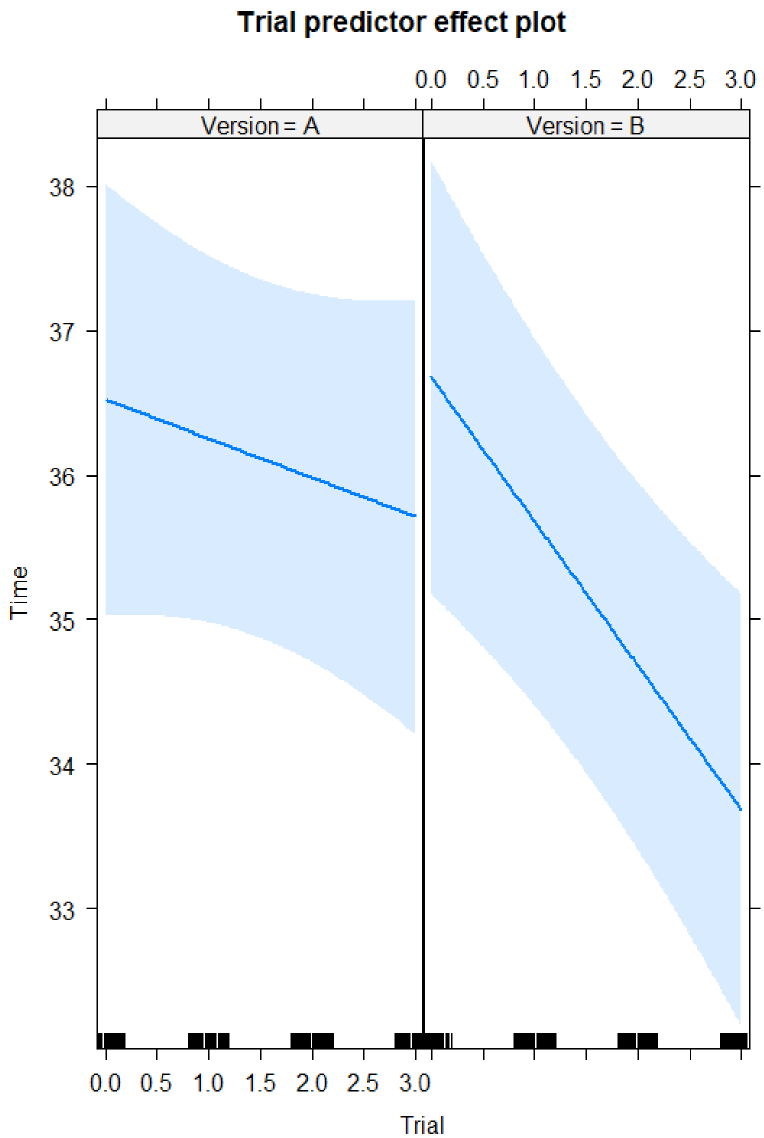
As for the quantitative research, where the data gathering is usually done using a questionnaire, the researchers could firstly incorporate a consent form into the questionnaire and then sort out the data, disregarding any, where the form was not properly accepted. A unique ID could be used for every questionnaire, making it possible to find the data set of any particular user that wishes to withdraw from the study. In the questionnaire itself, the research staff can leave out the questions that would reveal the persons' identity and in doing so – will protect their anonymity. As for minimizing the risk of harm and deceptive practices – they are the same for both qualitative and quantitative research.

## C1

```
mixed.lmer <- lmer(Time ~ Version * Trial + (1|id), data = ux1_set2)
summary(mixed.lmer)
confint(mixed.lmer)
```

Data from 320 observations of 40 participants was used in the analysis, powered by linear mixed effects model. Every participant played two versions of the game, four times each. Although clear trends were visible when looking at the data visually (Trial predictor effect plot) no statistically significant difference was found between the versions as the confidence interval crossed zero ($\beta$ = -0.727, $CI_{95}$ = -1.520 – 0.065).

Based on the data provided the original hypothesis that the players improve their skill more while playing game version two cannot be confirmed.




## C2

The research design used in the C1 assignment is called mixed effects linear model, done with repeated measurements. In difference to a linear regression model, the mixed effects model uses not only fixed effects, but also takes into account the random effects. These random effects represent a variation in observations that are not independent. For example, by assigning the ID of the participant as a random effect in the model, we are introducing a grouping factor, by which all the other factors are going to be grouped. Later, by assuming that all the random effects come from a larger general population we can bring the values closer to that common mean. This would have a greater effect on the outliers, while not affecting the general values by much. By taking the random effects into account we are eliminating a possibility for pseudo replication. Take for instance the data that was used in C1 – if we were not to include the ID as a random effect, we would treat each trial time as a separate entity and would have seen much more variation in the results as the completion time would vary greatly from one participant to another.

```
Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 11.83    3.439
 Residual             16.24    4.030
Number of obs: 320, groups:  id, 40
```

We can see how much of the variation we get from a participant to another if we look at the table above. If we were to divide the variance provided by id and divide it by a total amount

11.83/ (11.83+16.24) = 0.42, giving us a variation of 42% -  data that would have been lost if the linear regression model was used, possibly leading us to a wrong conclusion (Hajduk, 2020).

# References

Hajduk, G., 2020. *Introduction to linear mixed models*. [online] Available at: <https://ourcodingclub.github.io/tutorials/mixed-models/> [Accessed 1 Dec. 2020].