The Effect of Sound on Player Performance in Mario Kart 8 Deluxe

Olusegun Omotunde

1 Data Collection

My collected data can be found in appendix 3.2.

2 Analysis

2.1 Introduction of the model

I designed my experiment based off of a three-factor factorial randomized block design with fixed effects using the number of players playing at one time, the difficulty of the game, and the presence or lack of sound (referred to simply as number of players, difficulty, and sound respectively) as my three factors and used the participants as my blocking factor. My reasons for including these as factors can be referenced in my proposal in section **Error! Reference source not found.**. The equation of my model is:

$$y_{ijkl} = \mu + \tau_i + \gamma_j + \delta_k + (\tau\gamma)_{ij} + (\tau\delta)_{ik} + (\gamma\delta)_{jk} + (\tau\gamma\delta)_{ijk} + \beta_l + \varepsilon_{ijkl}$$

$$i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, c, \quad l = 1, \dots, n$$

$$\varepsilon_{ijkl} \overset{iid}{\sim} N(0, \sigma^2)$$

2.2 ANOVA table

Table 2-1 Analysis of Variance Table

Source of Variation	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Number of Players	1	14.29	14.29	0.4835	0.49
Game Difficulty	2	3512	1756	59.39	4.793e-14
Sound	2	89.04	44.52	1.506	0.2315
Participants (Block)	3	1661	553.6	18.73	2.539e-08
Number of Players:Difficulty	2	21.69	10.84	0.3668	0.6948
Number of Players:Sound	2	38.67	19.34	0.654	0.5243
Difficulty:Sound	4	172	43.01	1.455	0.2297
Number of Players:Difficulty:Sound	4	167.2	41.8	1.414	0.2427
Residuals	51	1508	29.56	NA	NA

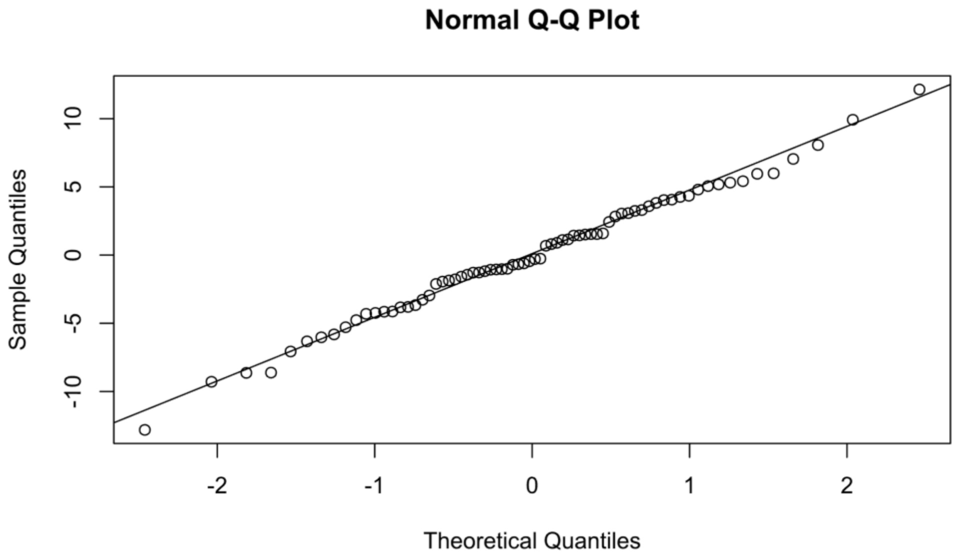
As seen in Table 2-1, neither sound nor the number of players has a significant effect on the length of time it takes a player to complete one seven-lap race on the Baby Park Circuit in Mario Kart 8 Deluxe, but game difficulty does.

2.3 Assumptions

Blocked three-factor factorial models require that the error terms are independent and identically distributed and follow a normal distribution with a mean of 0 and a constant variance. The factorial model used in the design of this experiment is robust to deviations away from normality for small sample sizes. I also assume that the residuals have equal variance among the levels of each factor and no interaction between my blocking factor and my treatments.

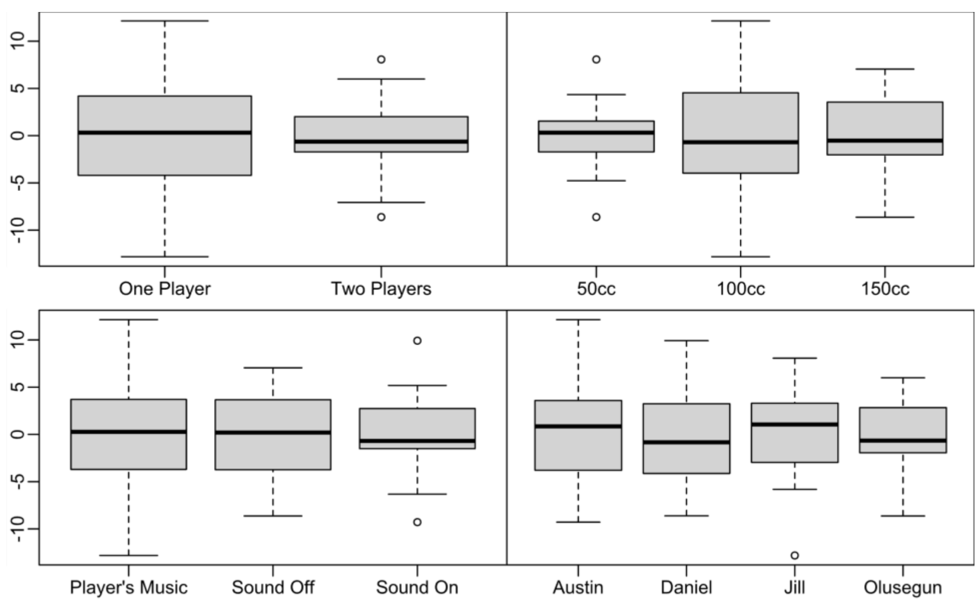
2.3.1 Normality

The assumption of normality can be assessed using the normal q-q plot below. Deviations away from normality can be seen as points deviating from the diagonal line. There is little to no deviation from the normal line, so I can assume that the data is normally distributed.

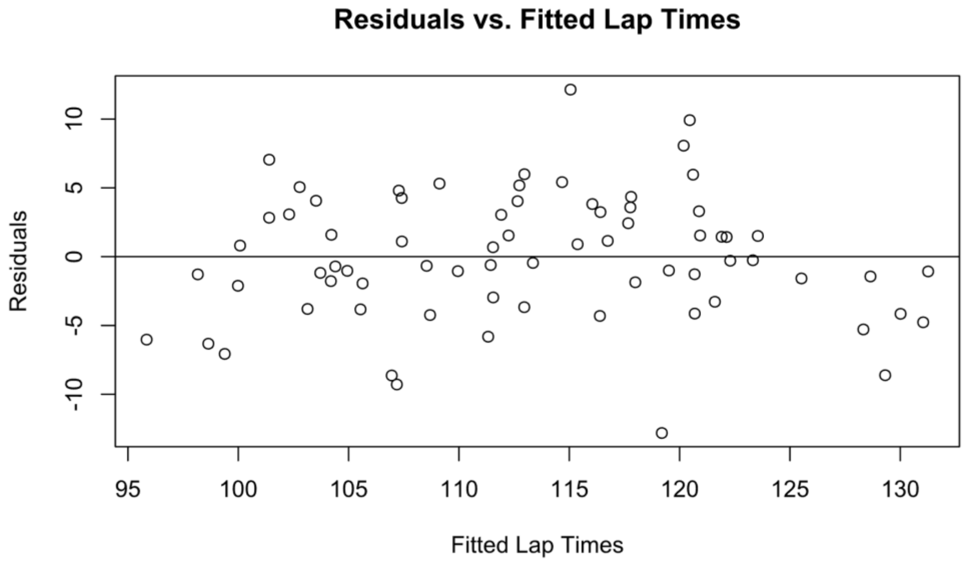


2.3.2 Constant Variance

First, I want to check if the residuals have equal variance among the levels of each factor. The graphs below display the spread of the levels of my three treatment factors and my blocking factor against the residuals. I can see that the variance is approximately equal across the levels of my treatments and block so I will state that this assumption holds.

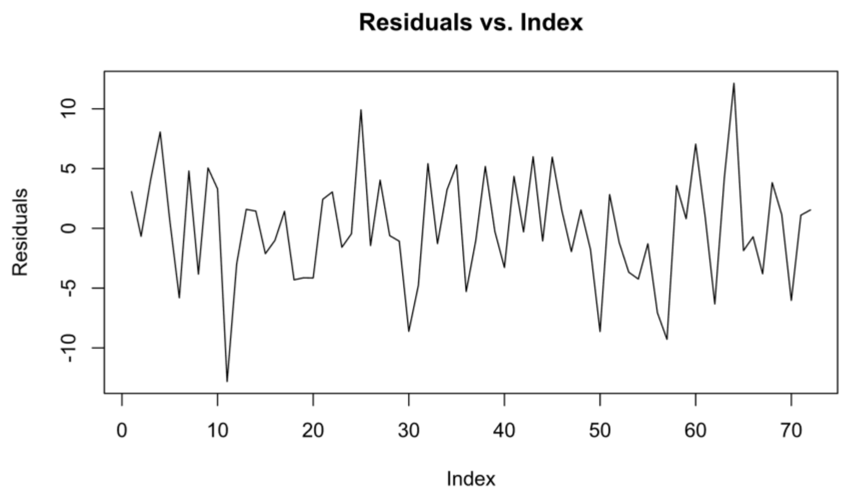


To check whether the variation of my residuals is constant and centered around zero, I refer to the plot below displaying my fitted values from the model against my residuals. I can see that the residuals are centered around zero and mostly equally distributed both along, and above and below the zero line. I will assume that the residuals have constant variance and a mean of 0.

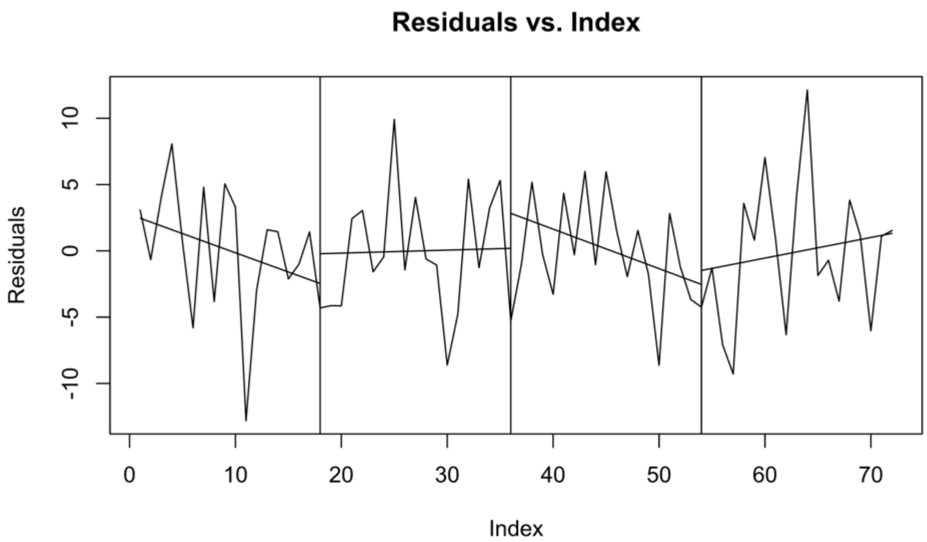


2.3.3 Independence

Dependence is present in the data when there is a trend or heteroscedasticity in the residual versus time plot. As I can see in the plot below, there does not appear to be any trend present, and the variance is constant.



Due to the nature of my design, I also want to check for independence within my blocks. The vertical lines in the plot below represents the last treatment in every block. A regression line has also been fit for each block to help distinguish any trend that may not be obvious at first glance. I can see that although there is a slight positive or negative trend in each block, the trends are not consistent, nor large enough to constitute a declaration of dependence. The variance across each block also appears to be constant so I can accept the assumption of independence.



2.3.4 No Interaction Between Treatments and Blocks

Interaction between my treatment factors and blocking factor can be detected by inspecting the fitted lap times versus residuals plot in section 2.3.2 for a curvilinear shape. Since there is no such pattern present, I can assume that there is no interaction between my block and treatment factors.

2.3.5 Outlier Check

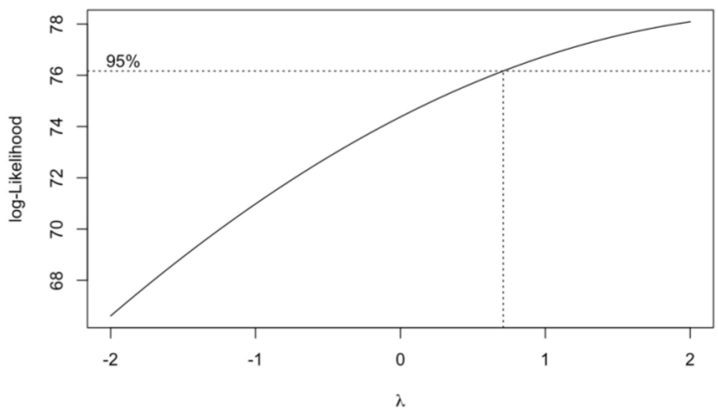
There is no hard rule that constitutes whether a point is an outlier or not, but the rule of thumb is if an observation has a residual satisfying the following condition, then it is an outlier:

$$\left|\frac{e_{ij}}{\sqrt{MS_E}}\right| > 3.$$

No points meet this criterion so I will keep the assumption that there are no outliers.

2.3.6 Box-Cox Transformation

Although all my assumptions have been met, it may be worth checking if a Box-Cox transformation is necessary. From the Box-Cox plot below, I can clearly see that the 95% confidence interval includes one; therefore, no transformation is recommended or necessary for my data.

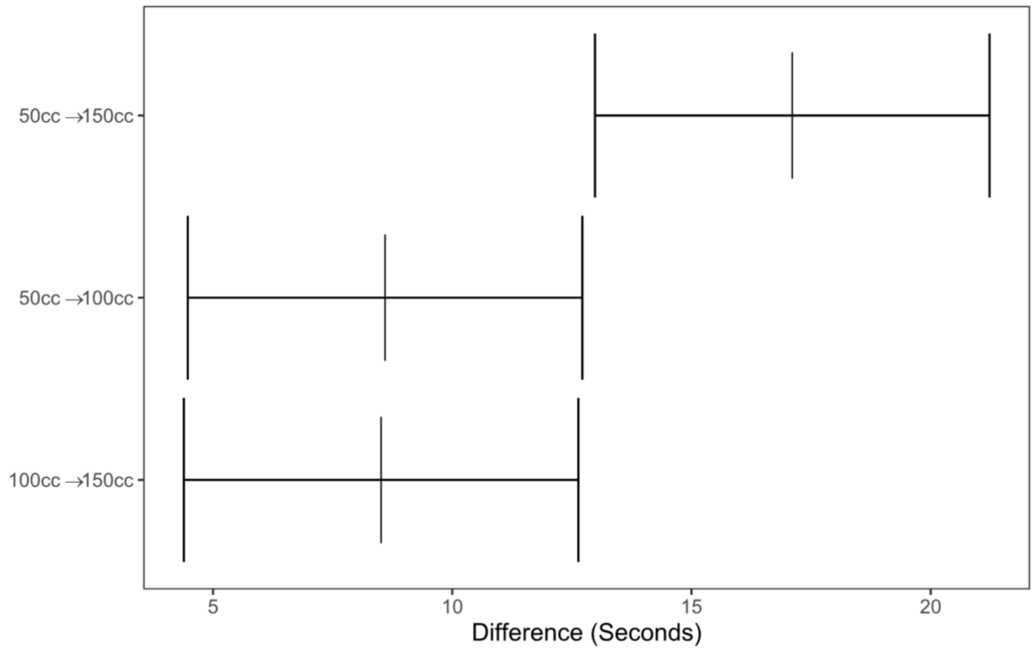


2.4 Results

My only significant treatment factor is the difficulty of the game which, as stated in my final proposal, was expected because as the difficulty is increased, the speed of the karts increases, decreasing the time it takes to complete a race. To check whether my intuition is correct, I will be doing both pairwise Tukey and polynomial contrasts, keeping a familywise α of 0.05. While polynomial contrasts are not normally associated with categorical variables, I have learned that this variable can also be numeric if one belongs to the modding community. Modding in the gaming community involves the practice of editing the game files to change an aspect of the game, in this case, the game's difficulty. Modders have found that they can edit the difficulty value to be anything above, below, and in between the levels of my difficulty treatment; therefore, I will be choosing to look at this factor with both points of view in mind and do pairwise Tukey comparisons for those without the knowledge of how to modify the game files in Mario Kart 8 Deluxe and do linear and quadratic, orthogonal, polynomial contrasts for those who do.

Below, I show the results of all pairwise Tukey comparisons for the difficulty treatment factor using a familywise α of 0.03. As I can see, all comparisons are significant with p-values less than 0.03. A better display of the results can be seen in the graph below where the middle lines are the differences, and the outer lines are the 95% confidence intervals. The difference in time to complete a race when going from 50cc to 100cc is very similar to the difference in time when going from 100cc to 150cc.

	Difference	Lower CI	Upper CI	p-Value
100cc-150cc	8.513	4.389	12.64	4.789e-06
50cc-150cc	17.11	12.98	21.23	0
50cc-100cc	8.595	4.471	12.72	3.98e-06

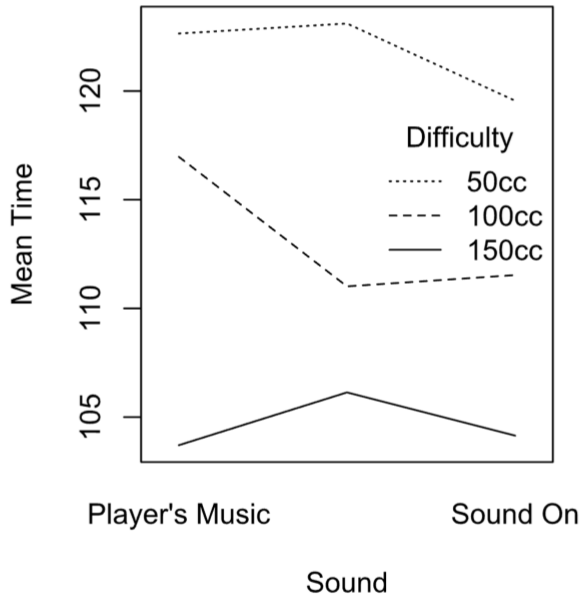


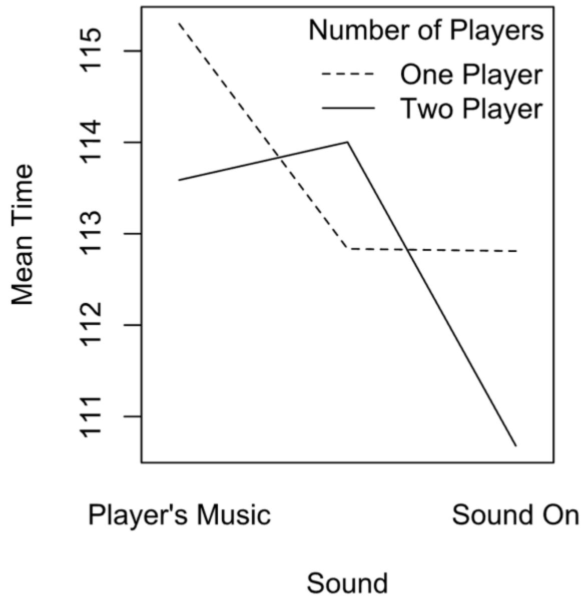
Below, I show the results of the polynomial contrasts where I can see that the confidence interval for the linear contrast does not include zero; therefore, there is a significant linear affect. I can also see that confidence interval for the quadratic contrast does include zero; therefore, difficulty does not have a quadratic effect.

	Lower CI	Contrast	Upper CI
Linear	2.558	17.11	31.66
Quadratic	-14.47	0.08	14.63

2.4.1 Interactions

My main factors of interest were the interactions between sound and difficulty and sound and number of players. I want to look at each of these interactions to see if there is a possibility of an interaction that I were not able to detect due to noise or error. I can see from the graphs below that the sound and difficulty may have an interaction but this may be due to error because of the insignificance of the interactions seen in section 2.2, and because there interactions are not very intuitive, for example, no sound changed from the best , to worst, to the best treatment as the difficulty increased. This is also the case for the interaction between sound and the number of players, as seen in the graph below.





2.5 Sources of error

I were not able to say that My effects of interest were significant so I would like to explore possible sources of error that could have buried my results.

2.5.1 Accuracy of my $\hat{\sigma}^2$

I estimated $\hat{\sigma}^2$ to be 33.14 using data from my pre-trial and the variation in the experiment was found to be 29.56 I found that my experiment results in a power of 0.9791, which is larger than my predicted power of 0.9640. Even though I underpredicted power, leading to a smaller type II error, I still want to test whether my estimated variance is statistically similar to the variance I got from my experiment using a F-test with an $\alpha = 0.05$: $H_0\colon \sigma_1^2 = \sigma_2^2$ vs. $H_1\colon \sigma_1^2 > \sigma_2^2$, where σ_1^2 represents my estimated pre-trial variance and σ_2^2 represents the variance from my experiment. My F-statistic is about $1.12 \sim F_{19,\,54}$ resulting in a p-value of about $0.358 > \alpha = 0.05$; therefore, I fail to reject my null hypothesis and can state that my variance estimated from my pre-trial was a good estimate of the variance in my experiment.

2.5.2 Blocking Efficiency

The question of whether I should have blocked or not can be answered simply by looking at the relatively large amount of variance explained by my blocking factor in section 2.2. I can also quantify this by calculating the relative efficiency of my randomized block design to my completely randomized design using the equation.

$$RE = cf\,\frac{\hat{\sigma}^2_{CRD}}{\hat{\sigma}^2_{RBD}}.$$

cf stands for correction factor and represents the equation;

$$cf = \frac{(dfe_{RBD}+1)(dfe_{CRD}+3)}{(dfe_{RBD}+3)(dfe_{CRD}+1)},$$

where dfe_{RBD} are the error degrees of freedom in my experiment and dfe_{CRD} are the error degrees of freedom if I were to run my experiment using a completely randomized design. dfe_{CRD} is calculated as $a(b-1)$, where a represents the number of treatments, and b represents the number of blocks. $\hat{\sigma}^2_{RBD}$ is my experiment variance estimated by the mean square error (MSE) and $\hat{\sigma}^2_{CRD}$ is my variance if I were to run my experiment with a completely randomized design and is estimated by;

$$\hat{\sigma}^2_{CRD} = \frac{(b-1)MS_{block} + b(a-1)SE}{(ab-1)},$$

where MS_{block} is the estimated variance explained by my blocking factor. I calculated my relative efficiency to be 5055.452 meaning that using the same universe of factors, I would have needed to increase my sample size by a factor of 5055.425 to achieve the same power as I did under the randomized block design. This means that there is not an unreasonable amount of error resulting from the use of blocking.

2.5.3 Data Collection

Out of all my sources of error, this may be the most significant. In my proposal I allowed the participants to have casual conversation and did not put any restrictions on the type of conversation. Some of my participants' competitive nature took over and the friendly banter may have created pressure to perform well which I did not account for. Another possible source of error in the data collection was the measurement error caused by inconsistent timing. A common complaint during the experiment came from the timers who found it difficult to pay attention to when they needed start and stop the stopwatches as the treatments within each block progressed. While any treatment that had a botched start or end time was scrapped and redone, I cannot rule out the possibility that some stopwatch operators may have been caught off guard by the beginning or end of the race and added or subtracted a few seconds to participants' time and not notify anyone.

3 Conclusions and Summary

The goal of my experiment was to check whether the main effect of sound and the interaction between sound and the number of players and the interaction between sound and difficulty, influenced the time it took to complete a seven-lap race on the Baby Park circuit in Mario Kart 8 Deluxe. After running my experiment, I discovered that the main factor and the two interactions I were interested in did not have a significant effect on a participant's race time; however, I did find that difficulty had a significant effect, but this was expected as the level of difficulty is increased, the speed of the karts in the game gets faster, reducing the length of time required to complete one seven-lap race. I explored the possibility that there was an interaction between sound and difficulty, and sound and the number of players buried beneath the error in my experiment. I cannot confirm or deny the possibility of an interaction effect between my factors of interest but I do have evidence that points towards the possibility that the effect of sound and it's interaction with difficulty and the number of players is less than what I were able to detect with my number of blocks.

3.1 Recommendations for Future Studies

Over the course of this study, I have come up with a few suggestions for future repeat studies. My first suggestion is to increase the number of blocks from 4 to 35 to be able to detect a difference of 2 seconds with a power of about 0.93039 as I believe the effect of sound to be much lower than I anticipated. The code I used to calculate this power can be found in the code section of my appendix. I also recommend that future studies find a more consistent way to time the races as this can become quite taxing on the individual who is given this task. One last recommendation is to limit the participants communication with each other. I believe that these recommendations will reduce the error and give clearer results.

Appendix

3.2 Data

Participant	Number of Players	Difficulty	Sound	Time (seconds)
Jill	Two Players	150cc	Sound On	105.38
Jill	Two Players	100cc	Sound On	107.87
Jill	Two Players	150cc	Sound Off	107.59
Jill	Two Players	50cc	Player's Music	128.25
Jill	Two Players	100cc	Sound Off	112.23
Jill	One Player	100cc	Sound On	105.51
Jill	One Player	100cc	Sound Off	112.08
Jill	One Player	150cc	Sound Off	101.71
Jill	One Player	150cc	Sound On	107.84
Jill	One Player	50cc	Sound Off	124.18
Jill	One Player	100cc	Player's Music	106.38
Jill	Two Players	100cc	Player's Music	108.59
Jill	Two Players	150cc	Player's Music	105.81
Jill	One Player	50cc	Player's Music	123.35
Jill	One Player	150cc	Player's Music	97.87
Jill	One Player	50cc	Sound On	118.51
Jill	Two Players	50cc	Sound Off	123.56
Jill	Two Players	50cc	Sound On	112.08
Daniel	Two Players	100cc	Player's Music	116.55
Daniel	One Player	50cc	Sound Off	125.86
Daniel	Two Players	100cc	Sound On	120.1
Daniel	One Player	150cc	Sound On	114.96
Daniel	Two Players	50cc	Sound On	123.94

Daniel	Two Players	150cc	Player's Music	112.9
Daniel	One Player	100cc	Sound On	130.38
Daniel	One Player	50cc	Sound On	127.21
Daniel	Two Players	150cc	Sound Off	116.69
Daniel	Two Players	150cc	Sound On	110.83
Daniel	Two Players	50cc	Sound Off	130.19
Daniel	Two Players	50cc	Player's Music	120.7
Daniel	One Player	50cc	Player's Music	126.27
Daniel	One Player	150cc	Sound Off	120.09
Daniel	Two Players	100cc	Sound Off	119.4
Daniel	One Player	100cc	Sound Off	119.65
Daniel	One Player	150cc	Player's Music	114.43
Daniel	One Player	100cc	Player's Music	123.04
Olusegun	Two Players	150cc	Sound Off	103.91
Olusegun	One Player	100cc	Sound On	117.92
Olusegun	One Player	50cc	Player's Music	123.06
Olusegun	Two Players	50cc	Player's Music	118.32
Olusegun	Two Players	50cc	Sound On	122.15
Olusegun	One Player	50cc	Sound Off	122
Olusegun	Two Players	100cc	Player's Music	118.96
Olusegun	Two Players	100cc	Sound On	108.9
Olusegun	One Player	100cc	Player's Music	126.57
Olusegun	Two Players	50cc	Sound Off	125.05
Olusegun	Two Players	150cc	Player's Music	103.69
Olusegun	One Player	50cc	Sound On	122.47
Olusegun	One Player	150cc	Sound On	102.42
Olusegun	One Player	150cc	Sound Off	98.32
Olusegun	One Player	150cc	Player's Music	104.23
Olusegun	Two Players	150cc	Sound On	102.54
Olusegun	Two Players	100cc	Sound Off	109.29
Olusegun	One Player	100cc	Sound Off	104.45
Austin	Two Players	150cc	Sound On	96.88
Austin	Two Players	150cc	Sound Off	92.32
Austin	One Player	100cc	Sound On	97.9
Austin	One Player	50cc	Player's Music	121.35
Austin	Two Players	150cc	Player's Music	100.89
Austin	One Player	150cc	Sound Off	108.45
Austin	One Player	50cc	Sound On	116.28
Austin	One Player	150cc	Sound On	92.32
Austin	Two Players	100cc	Sound Off	111.67
Austin	One Player	100cc	Player's Music	127.2

Austin	Two Players	50cc	Sound Off	116.13
Austin	Two Players	100cc	Sound On	103.69
Austin	One Player	100cc	Sound Off	99.34
Austin	Two Players	50cc	Player's Music	119.87
Austin	One Player	50cc	Sound Off	117.89
Austin	One Player	150cc	Player's Music	89.82
Austin	Two Players	100cc	Player's Music	108.52
Austin	Two Players	50cc	Sound On	113.79

3.3 Code

3.3.1 Packages used

```
library(tidyverse)
library(pander)
library(MASS)
```

3.3.2 Import the above data

```
dat <- read.csv("data.csv", stringsAsFactors = T,fileEncoding = 'UTF-
8-BOM')
dat$difficulty <- relevel(dat$difficulty, "50cc")
```

3.3.3 ANOVA Table

```
aov_out <- aov(formula = time ~ numPlayers*difficulty*sound +
participant,
               data = dat)
(aovSum <- summary(aov_out)[[1]])
```

3.3.4 Assumption Checking

```
qqnorm(aov_out$residuals)
qqline(aov_out$residuals)
par(mfrow = c(2,2),
    mar = c(2,0,0,0),
    mgp = c(0,.5,0),
    omi = c(0,.25,0,0))
plot(x = dat$numPlayers, y = aov_out$residuals,
     xlab = "", ylab = "Residuals")
plot(x = dat$difficulty, y = aov_out$residuals,
     xlab = "", ylab = "", yaxt = "n")
plot(x = dat$sound, y = aov_out$residuals,
     xlab = "", ylab = "Residuals")
plot(x = dat$participant, y = aov_out$residuals,
     xlab = "", ylab = "Residuals", yaxt = "n")
plot(aov_out$fitted.values,aov_out$residuals,xlab="Fitted Lap
Times",ylab="Residuals",
     main="Residuals vs. Fitted Lap Times")
abline(h=0)
plot(aov_out$residuals,ylab="Residuals",type="l",
     main="Residuals vs. Index", xlim = c(1,72))
res_ind_df <- data.frame(index = 1:length(aov_out$residuals),
                          res = aov_out$residuals)
```

```
res_ind_lm1 <- lm(res~index, res_ind_df[1:18,])
res_ind_lm2 <- lm(res~index, res_ind_df[(18+1):(18*2),])
res_ind_lm3 <- lm(res~index, res_ind_df[(18*2+1):(18*3),])
res_ind_lm4 <- lm(res~index, res_ind_df[(18*3+1):(18*4),])
plot(aov_out$residuals,ylab="Residuals",type="l",
     main="Residuals vs. Index", xlim = c(1,72))
abline(v= 18)
abline(v= 18*2)
abline(v= 18*3)
clip(1, 18, -100, 100)
abline(res_ind_lm1, xlim = 1:18)
clip(18, 18*2, -100, 100)
abline(res_ind_lm2, xlim = 1:18)
clip(18*2, 18*3, -100, 100)
abline(res_ind_lm3, xlim = 1:18)
clip(18*3, 18*4, -100, 100)
abline(res_ind_lm4, xlim = 1:18)
# outlier detection
which(abs(aov_out$residuals/aovSum$`Mean Sq`[9])>3)
boxcox(aov_out)
```

3.3.5 F-test between the pre-trial variance and the experiment variance

```
Fstat <- var(time_obs)/aovSum$`Mean Sq`[9]
1-pf(Fstat, df1 = 19, df2 = 54)
```

3.3.6 Interaction charts

```
Sound <- dat$sound
Difficulty <- dat$difficulty
`Number of Players` <- dat$numPlayers
par(mfrow = c(1,2))
interaction.plot(x.factor = Sound, trace.factor = Difficulty,
                 response = dat$time, xlab = "Sound", ylab = "Mean
Time",
                 legend = F)
legend(x = 2.1, y = 119, legend = c("50cc","100cc","150cc"), lty =
3:1,
       bty = "n", title = "Difficulty")
interaction.plot(x.factor = Sound, trace.factor = `Number of Players`,
                 response = dat$time, xlab = "Sound", ylab = "Mean
Time",
                 legend = F)
legend(x = 1.7, y = 115.5, legend = c("One Player","Two Player"), lty
= 2:1,
       bty = "n", title = "Number of Players")
```

3.3.7 Tukey Comparisons

```
alphaTukey <- 0.03
TukeyHSD(aov_out, "difficulty",ordered = T,
         conf.level = 1-alphaTukey)$difficulty

tuk <- TukeyHSD(x = aov_out,which = "difficulty",
                 ordered = T, conf.level = 1-.03)$difficulty %>%
  apply(2,function(i) as.numeric(i)) %>% as.data.frame()
```

```
difficultyTukey <- cbind(comparison = c("100cc->150cc",
                                        "50cc->150cc",
                                        "50cc->100cc"),
                         tuk)
theme_set(theme_bw())
theme_update(text = element_text(size=12),
panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),
strip.background = element_blank()
)
ggplot(data = difficultyTukey) +
  geom_point(mapping = aes(x =  comparison, y = diff),
             stat="identity", pch = 3, size = 20) +
  geom_errorbar(mapping = aes(x = comparison, ymin = lwr, ymax = upr))
+
  labs(x = "", y = "Difference (Seconds)") +
  scale_x_discrete(labels=c('50cc->150cc'= expression("50cc" %->%
"150cc"),
                            '50cc->100cc'= expression("50cc" %->%
"100cc"),
                            '100cc->150cc'= expression("100cc" %->%
"150cc"))) +
  coord_flip()
```

3.3.8 Polynomial comparisons

```
dat %>%
  group_by(difficulty) %>%
  summarise(mean(time),.groups = "drop") %>%
  select(-difficulty) %>%
  unlist() %>% unname() -> difficulty_means

alpha_poly <- 0.02

MSE <- aovSum$`Mean Sq`[9]

ci_lin <- c(1,0,-1)
ci_quad <- c(1,-2,1)

C_lin <- sum(ci_lin*difficulty_means)
C_quad <- sum(ci_quad*difficulty_means)

mult_poly <- qt(1-alpha_poly/(2*2), df = aovSum$Df[9])

lower_poly_CI_lin <- C_lin - mult_poly*sqrt(MSE)
lower_poly_CI_quad <- C_quad - mult_poly*sqrt(MSE)
upper_poly_CI_lin <- C_lin + mult_poly*sqrt(MSE)
upper_poly_CI_quad <- C_quad + mult_poly*sqrt(MSE)

cbind(`Lower CI` = c(lower_poly_CI_lin, lower_poly_CI_quad),
      `Contrast` = c(C_lin, C_quad),
      `Upper CI` = c(upper_poly_CI_lin, upper_poly_CI_quad)) ->
polyout

rownames(polyout) <- c("Linear", "Quadratic")
```

```
pander::pander(polyout)
```

3.3.9 Relative Efficiency

```
a <- 3
b <- 4
MSblock <-aovSum$`Mean Sq`[4]

sigma2CRD <- ((b-1)*MSblock + b*(a-1)*MSE)/(a*b-1)

dfe_RBD <- (a-1)*(b-1)
dfe_CRD <- a*(b-1)

cf <- ((dfe_RBD+1)+(dfe_CRD+3))/((dfe_RBD+3)+(dfe_CRD+1))

cf*sigma2CRD*MSE
```

3.3.10 Power for future studies

```
alpha=.05
a=3  # number of levels of A
b=3  # number of levels of B
c=2     # number of levels of C
sigsq = 29.56 # The sigma squared from the experiment
n=35 #number of blocks
DA=2 #desired diff (in seconds) in means to detect with prob 1-beta
Fcrit=qf(1-alpha,a-1,(a*b*c-1)*(n-1)) #value at which I reject H0
lam=n*b*c*(DA^2)/(2*sigsq) #non-centrality parameter (ncp)
beta=pf(Fcrit,a-1,(a*b*c-1)*(n-1),ncp=lam)
power=1-beta
nforA=cbind(n,Fcrit,beta,power)  #output for A
nforA
```

4 Citations

Box, G. E., & Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association, 65(332)*, 1509-1526.

R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

RStudio Team. (2021). RStudio: Integrated Development Environment for R. Boston, MA. Retrieved from http://www.rstudio.com/

Tsegelskyi, R., & Daróczi , G. (2021). pander: An R 'Pandoc' Writer. Retrieved from https://CRAN.R-project.org/package=pander

Venables , W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer. Retrieved from https://www.stats.ox.ac.uk/pub/MASS4/

Wickham , H., Averick , M., Bryan , J., Chang, W., D'Agostino McGowan, L., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. doi:10.21105/joss.01686