Data Analysis Project - Understanding Human's Motor system

Lilach Cohen Hadad, Omer Falach, Mor Schenavsky

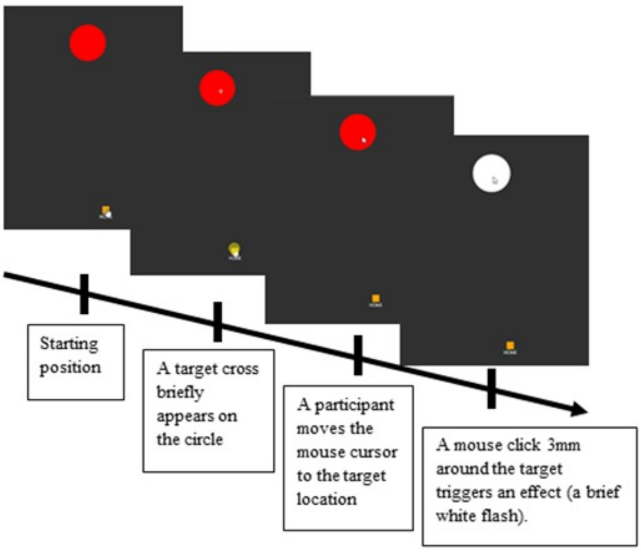
6/27/2021

Background

In our project, we teamed up with two researchers from The Cognitive Development Lab at the Haifa University - Prof. Bat-Sheva Hadad and Dr. Noam Karsh. Their research investigates the impact of low specialization in sensory perception in autism spectrum disorder(ASD) on sensorimotor control.

About the Experiment

During the experiment, TD (typical development) participants were required to perform a reaching movement towards a target. A perceptual effect (a brief white flash) appeared around the target after a correct click on the target using their mouse (Figure 1). In the Short lag condition, the effect could appear between 50-200 ms after reaching the target, and in the Long lag condition, the effect could appear between 450-600ms after reaching the target. To examine the impact of contextual information on motor performance, both conditions included trials in which an ambiguous 300ms lagged effect appeared after the response. Due to the scope of our project, we did not include analysis of these effects. In addition, all participants were required to fill out the aq questionnaire that aims to find behavior characteristics related to ASD.



By sending sensory feedback to participants, the experiment tested whether participants improved their motor functioning (accuracy) after several rounds of performing the experiment, in reaction to the sensory feedback.

About the Data set

The scope of our project will be the data set collected during the experiment, which included 125 TD participants who performed the test 140 rounds each. The data set contains data for each one of the participants:

1. Experiment results (number of trial, lag effect, speed of movement, speed of reaction, x, y coordinates for calculating distance from target, success status etc.)
2. participant data (participant ID, age, gender etc.)
3. Experiment surrounding characteristics.
4. Results and score of aq questionnaire.

Goal

Our goal is to use data visualization and modelling (including T-test for paired samples and linear regression models) to explore two main hypothesis:
1. Action-effect improves accuracy, hence learning proccess occurs.
2. There is a relation between accuracy to the speed of movement, which is another variable that improves due to the learning process.

Importing Raw Data

```
exp1 <- read_csv('exp1.csv')
```

Data Preparation & Tidying

In order to process the data to be able to run our models, we have decided to focus on several main variable. After we cleaned and performed some tidying, the main variables are:
1. subnum - represents the identify number of the participant.
2. lag - long "lag" phases defined as "W" and the short ones as "N".
3. distance - The main data that we have investigated is accuracy. The accuracy rate represents the distance between the mouse click that the participant did to the center of the target. We have calculated the distance from the target by using coordinates X and Y in the raw data and Pythagoras equation. We then calculated the average accuracy rate of all the results for every participant, named it "mean distance".
4. ttime - The speed of movement time for every participant since the moment he started moving until he marked the target.

Other data that we have inserted to the final Excel sheet but we didn't use them in our analysis are -ftime and the aq personal test score. Ftime is our value definition for the average response speed time from the first reaction of the participant until the first mouse movement towards the target. The personal aq test score is the summed score of participant's answers to the aq questionnaire.

Removing outliers:

1. Participant experienced IT and network interruption.

2. We removed the following observations and participants:

- Observations with reaction time of more than 3000ms
- Observations with effect of the 300ms lag as this is not part of our analysis
- Participants who has less than 85% hitting target success rate
- Observations in which participant didn't hit the target

```
###Changing column names
exp1_p = exp1 %>% rename( breaks= "break", subnum = participantid)
exp1_p$block_condition = exp1_p$target

#Removing outliers,Creating new variables,Distance - Calculating sample's accuracy using Pyth
agoras
exp1_p  = exp1_p %>%
  filter(interruptions==2 & device<=2 & controller == 1)%>%
  mutate(x1 = ((x /cm)*10))%>%
  mutate(y1 = (y /cm)*10) %>%
  mutate(xy = (x1*x1)+(y1*y1)) %>%
  mutate(distance = (xy^0.5))

#Creating new variable - p_success - calculating participant's success rate
p_suc<-exp1_p %>%
  group_by(subnum)

p_success<- p_suc%>%summarize(total = n(),correct = sum(success))%>%mutate(pcorrect = correc
t/total)%>%select(subnum,pcorrect)

exp1_p_1 <-exp1_p %>% left_join(p_success)

#calculating flash's impact
exp1_p_ar<-exp1_p_1%>%arrange(subnum,block_condition,trial)%>%select(subnum,block_condition,t
rial,distance,ftime,ttime,success,pcorrect,lag,aq_score)
exp1_p_ar["distance1"]<- 0.05
exp1_p_ar["ftime1"]<- 0.05
exp1_p_ar["ttime1"]<- 0.05
exp1_p_ar["success1"]<- 0.05
exp1_p_ar["pcorrect1"]<- 0.05

for (j in 1:(nrow(exp1_p_ar)-1))
{
  if ((exp1_p_ar[j,1]==exp1_p_ar[j+1,1])&(exp1_p_ar[j,2]==exp1_p_ar[j+1,2]))
  {

    exp1_p_ar[j,11]<-exp1_p_ar[j+1,4]
    exp1_p_ar[j,12]<-exp1_p_ar[j+1,5]
    exp1_p_ar[j,13]<-exp1_p_ar[j+1,6]
    exp1_p_ar[j,14]<-exp1_p_ar[j+1,7]
    exp1_p_ar[j,15]<-exp1_p_ar[j+1,8]
  }
  else
  {

    exp1_p_ar[j,11]<-0.05
    exp1_p_ar[j,12]<-0.05
    exp1_p_ar[j,13]<-0.05
    exp1_p_ar[j,14]<-0.05
    exp1_p_ar[j,15]<-0.05
  }
}

#Cleaning outliers and unnecessary observations for t-test
exp1_p_ar = exp1_p_ar %>% filter(pcorrect1>=0.85) %>% filter(success1 !=0 & success !=0)%>%fi
lter(ttime<=3000 & ttime1<=3000)%>%filter(lag==300)
```

```
#Calculating means calculating for each subnum and block condition pairs
exp1_ready =exp1_p_ar%>% group_by(subnum,block_condition)%>%summarize(mean_distance1 = mean(d
istance1),sd_distance1 = sd(distance1),mean_ttime1 = mean(ttime1),mean_ftime1 = mean(ftime1),
aq_score = mean(aq_score))

head(exp1_ready)
```

```
## # A tibble: 6 x 7
## # Groups:   subnum [3]
##   subnum block_condition mean_distance1 sd_distance1 mean_ttime1 mean_ftime1
##    <dbl> <chr>                    <dbl>        <dbl>       <dbl>       <dbl>
## 1      3 N                         1.48        0.763        543.        96.3
## 2      3 W                         2.11        0.668        522.        65.1
## 3      4 N                         1.14        0.638        708.       129.
## 4      4 W                         1.36        0.699        766.       102.
## 5      9 N                         1.62        0.839       1022.       131.
## 6      9 W                         1.94        1.00        1050.       101.
## # ... with 1 more variable: aq_score <dbl>
```

```
#Arranging data for T-test
exp1_t_test = exp1_ready%>% arrange(block_condition,subnum)


#Preparing data for Linear Regression by calculating differences
ready_for_reg = exp1_ready %>%
  group_by(subnum) %>%
  mutate(Diff_distance = (mean_distance1) - lag((mean_distance1)),Diff_time = (mean_ttime1) -
lag((mean_ttime1)))%>%drop_na()%>%select(subnum,Diff_distance,Diff_time)

p1 = ggplot(exp1_t_test, aes(x=block_condition, y=mean_distance1,fill = block_condition))
```

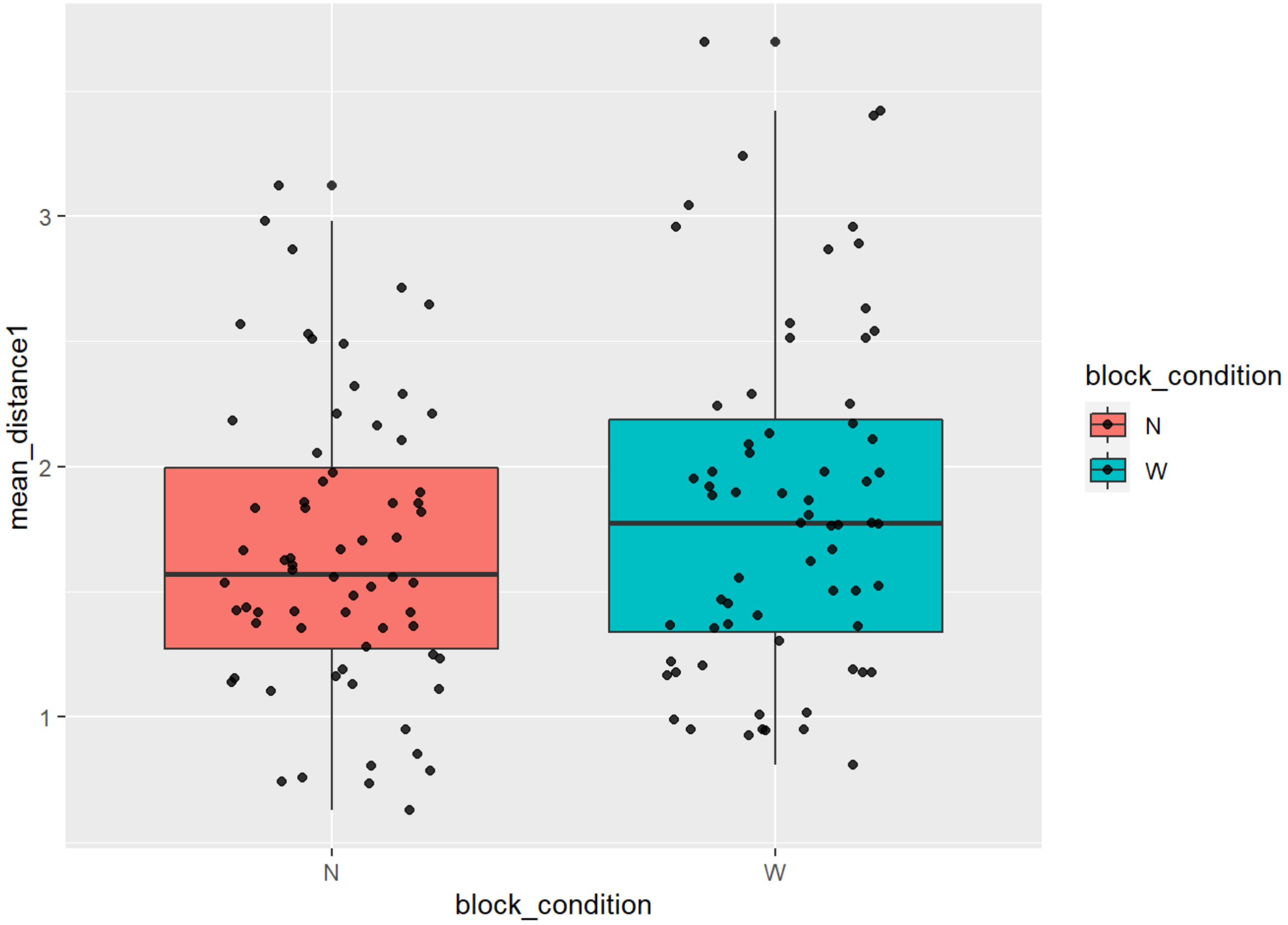
T-Test Model

```
# T-test
head(exp1_t_test)
```

```
## # A tibble: 6 x 7
## # Groups:   subnum [6]
##   subnum block_condition mean_distance1 sd_distance1 mean_ttime1 mean_ftime1
##    <dbl> <chr>                    <dbl>        <dbl>       <dbl>       <dbl>
## 1      3 N                         1.48        0.763        543.        96.3
## 2      4 N                         1.14        0.638        708.       129.
## 3      9 N                         1.62        0.839       1022.       131.
## 4     11 N                         1.37        0.622        536       1040.
## 5     12 N                         2.57        0.931        674.       117.
## 6     15 N                         2.65        1.31         640.       105.
## # ... with 1 more variable: aq_score <dbl>
```

Box Plot

```
p1+geom_boxplot()+
geom_jitter(width=0.25, alpha=0.8)
```



As shown above, participant's accuracy mean in the long lag (W) differs from the short lag (N). We will perform t-test to test if this difference can conclude any improvement in participant's learning process. In addition, it can be seen that the observations are spread over the 4 quartiles with few outliers, which represnet some participants overscoring or underscoring during the experiment.

We use hypothesis testing (t-test for paired samples) to test if accuracy improves with the action-effect. We will define the test as follows:

$$H0 : \mu 1 - \mu 2 = 0$$
$$H1 : \mu 1 - \mu 2 \neq 0$$

μ1 - mean of accuracy for short lag condition (N)
μ2 - mean of accuracy for long lag condition (W)

The null hypotheses suggests that no improvement is made, and thus the means of accuracy for both the short and long lag are equal. Once the level of accuracy differs between the two groups, this can show an improvement in participant's motor reaction. Our Statistical significance is defined at the level of 0.05.

The result of the test are shown below:

```
stat.test = t.test(mean_distance1 ~ block_condition,data = exp1_t_test, paired = TRUE, mu = 0
)
stat.test
```

```
##
##  Paired t-test
##
## data:  mean_distance1 by block_condition
## t = -2.8229, df = 67, p-value = 0.00626
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.31687982 -0.05437281
## sample estimates:
## mean of the differences
##              -0.1856263
```

The p-value result is 0.00626 and therefore we will decline the H0 hypothesis . This means that there is an improvement on participant's motor reaction in response to an immediate effect.

Linear Regression Model

Previous experiments have already found that the speed of movement improves due to the perceptual effect. Our assumption in this experiment is that that the speed of movement is related to accuracy. Thus we would like to test if accuracy improves as well with the time of movement or that there is a negative relation, showing that one comes at the expense of the other.

$\hat{y} = \beta 0 + \beta 1x + \varepsilon$

X is the mean of the difference in distance for each participant
Y is the mean of ttime for each participant

```
# T-test
head(ready_for_reg)
```

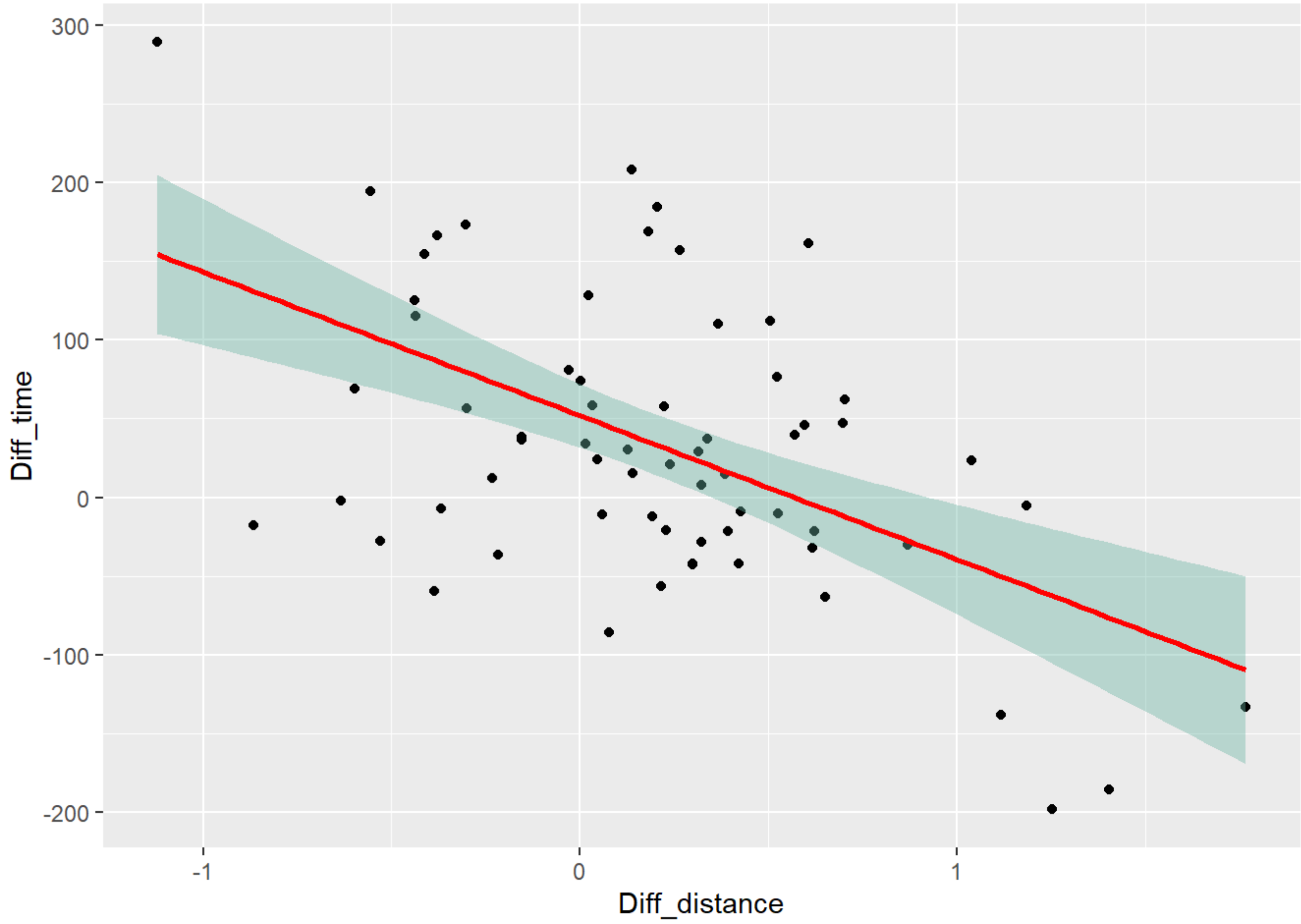
```
## # A tibble: 6 x 3
## # Groups:   subnum [6]
##   subnum Diff_distance Diff_time
##    <dbl>         <dbl>     <dbl>
## 1      3         0.622    -21.3
## 2      4         0.223     57.7
## 3      9         0.315     28.8
## 4     11         0.605    161.
## 5     12         0.323      7.81
## 6     15        -1.12     290.
```

Regression model

```
model <- lm(Diff_time ~ Diff_distance, data = ready_for_reg)
```

Scatter Plot

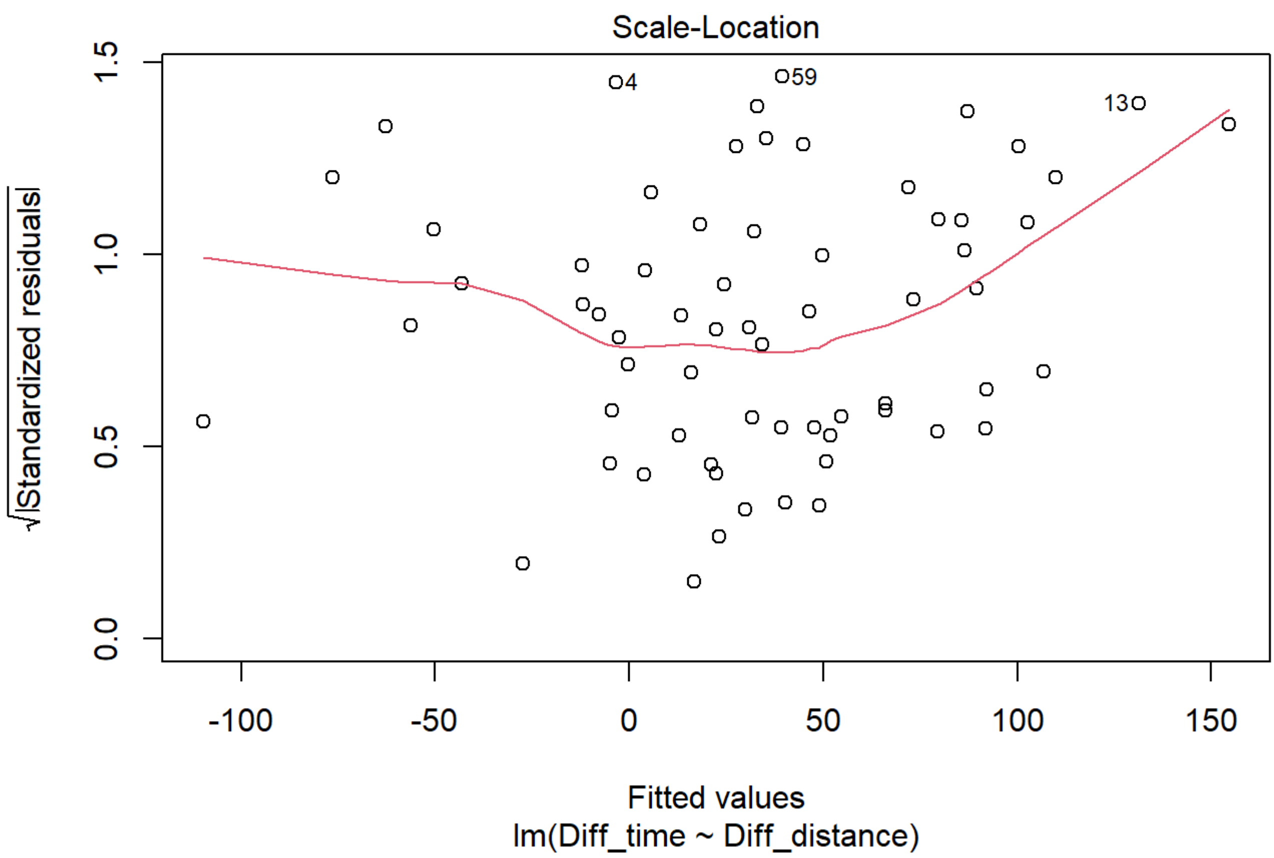
```
ggplot(ready_for_reg, aes(x = Diff_distance,y = Diff_time)) +
geom_point()+
geom_smooth(method=lm , color="red", fill="#69b3a2", se=TRUE)
```



This plot shows the negative relation between the accuracy and ttime, as well as the distribution of the observations.

Scale-Location Plot

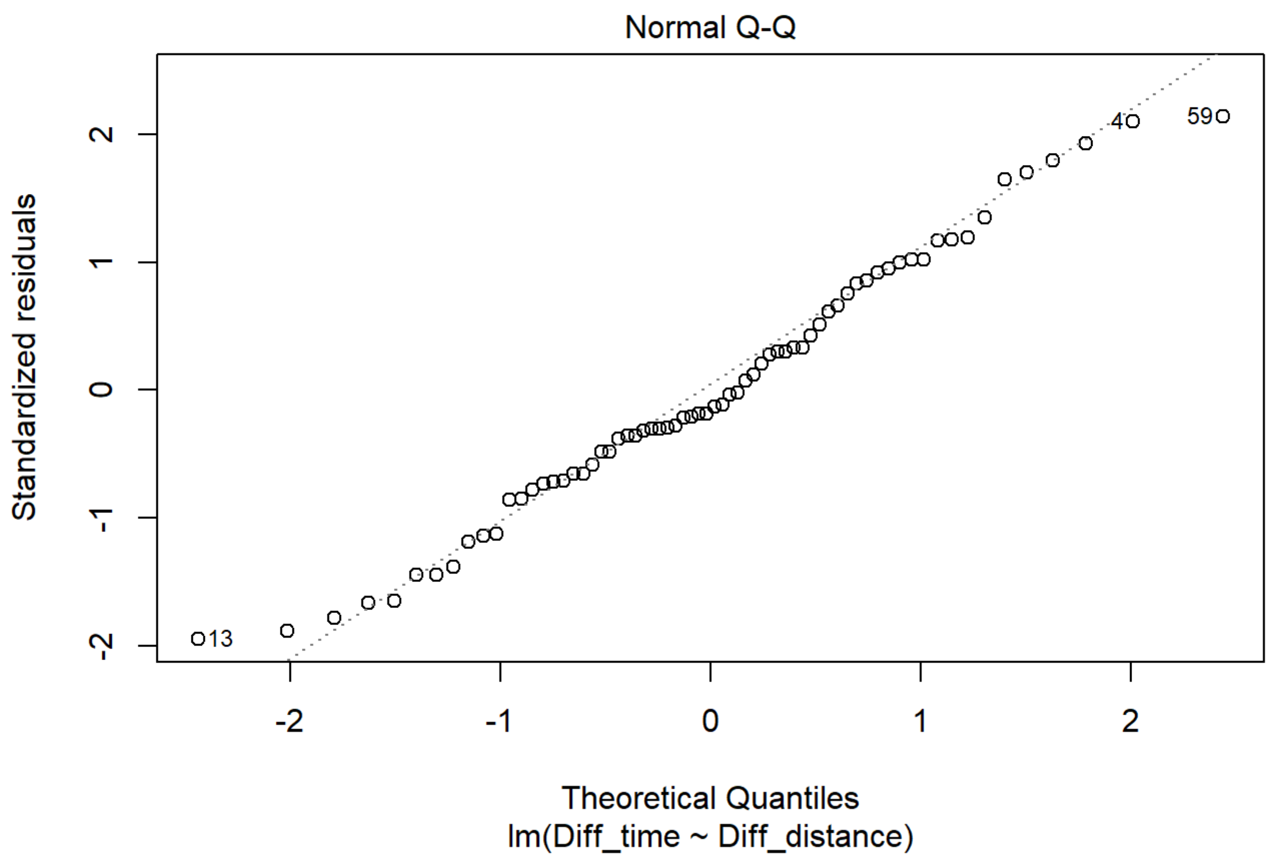
```
plot(model, 3)
```



The curve in the graph seems to be violating the homoscedasticity assumption of constant variance of residuals. We can see that the once the difference in accuracy is significant (both in higher and lower values),There was a wider range of effect on speed. We can see observations of participants who showed improvement in speed of movement in parallel with accuracy improvement as well as participants whose the accuracy improvement came at the expense of the speed of movement at different levels.

QQ Plot

```
plot(model, 2)
```



The Normal Q-Q plot shows that the normality assumption exists, since all points follow the reference line.

Summary regression

```
summary(model)
```

```
##
## Call:
## lm(formula = Diff_time ~ Diff_distance, data = ready_for_reg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -148.92  -52.80  -12.16   60.73  168.84
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      51.97      10.19   5.101 3.07e-06 ***
## Diff_distance   -91.45      17.89  -5.111 2.96e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79.43 on 66 degrees of freedom
## Multiple R-squared:  0.2835, Adjusted R-squared:  0.2727
## F-statistic: 26.12 on 1 and 66 DF,  p-value: 2.957e-06
```

According to the result of the p-value in the linear regression test we can say that there is a significant relation between the accuracy and the speed of movement. This relation is negative, telling that once participant improves his accuracy, his speed of movement will be reduced. The R-squared result shows that the connection is weak, which can be explained by other variables that might affect the trial (such as environment effects).

Conclusions and Summary

We chose to analyze data set of a research that tested improvement of motor learning. From past researchers' conclusions we know that learning process and speed improvements occurred in a response to an immediate affect and decided to find out if there is an improvement in the accuracy rate. We performed two tests in order to analyze the trial results. We assumed that there is an improvement in the accuracy rate in a response to an immediate effect, and chose to do a t-test to examine this assumption and we confirmed it. In addition, we assumed that there is a linear connection between the accuracy and the speed of movement. We did a linear regression test and found out that there is relation between them, but it a weak one due to other factors that affect the experiment. Moreover, the connection shows an inverse relationship- the more accurate the subject, the less fast his speed of movement is. We enjoyed exploring new field during our project and find new discoveries. The research sets a ground for further experiments and hypothesis, such as the effect on ASD population, the relation between the context of the affect and motor learning processes etc.