Quantitative HCI: Comparing Different Input Techniques

Matthias Felix
11-746-625
matthias.felix@uzh.ch

Florian Fischer
11-611-985
florian.fischer2@uzh.ch

Catrin Loch
13-718-143
catrin.loch@uzh.ch

1. DATASET

The dataset consists of a random sample of participants that is drawn from a larger dataset. It contains results of an experiment that was conducted to compare the performance of different input techniques (called conditions) in terms of speed and accuracy of input. As dependent variable, Throughput is used because it is a measure that embeds both the speed and accuracy of responses (throughput is useful as a dependent variable in factorial experiments using pointing techniques or pointing devices as independent variables) (McKenzie 2012: 252).

In the following sections, two research questions are assessed using linear model regressions. The research questions are:

RQ1: To which extent does the choice of input techniques influence the pointing performance?

RQ2: To which extent does the learning effect influence the pointing performance?

1.1 Normality Assumption

Interpretation: The normality assumption does not hold for the subset of the data. The null hypothesis is that the data is normally distributed. Since the p-values are very small (< 0.05), we can reject this null hypothesis and cannot assume that the data is normally distributed. The QQ plots in figure **??** confirm this finding. An exception has to be made for the Phone_Sway condition which is normally distributed (p = 0.17).

1.2 Homogeneity Assumption

Interpretation: The null hypothesis is that the data have a constant variance. Since the p-value yielded by the conducted Levene Test is very small (<0.05), we can reject the hypothesis and cannot assume that the variance is homogeneous.

1.3 Transformation

Due to the above described findings, we perform a logarithmic transformation on the data and test the normality and homogeneity assumption again.

1.4 Normality Assumption after Logarithmic Transformation

Interpretation: The normality assumption does not hold either for the data after logarithmic transformation. The p-values are again very small (< 0.05) and we can reject the null hypothesis and assume that the data is not normally distributed. The QQ plots in figure **??** confirm this finding, this time for all four conditions.

1.5 Homogeneity Assumption after Logarithmic Transformation

Interpretation: The Homogeneity assumption is not confirmed either. The p-value of the Levene Test is again very small (<0.05), and we need to reject the hypothesis and cannot assume that the variance is homogeneous.

2. EFFECT OF INPUT TECHNIQUES

RQ1: To which extent does the choice of input techniques influence the pointing performance?

2.1 Fit the Linear Model

Fitting the linear model reveals that all conditions have a negative interaction effect on throughput. The p-values are below 0.05, with negative interaction effects between -0.34 and -0.61.

2.2 Normality Assumption of Model Residuals (original data)

```
## [1] 928 789
```

The normality assumption does not hold for the residuals of the data either. The QQ plot in 1 shows that the data points are not within the blue lines, indicating clearly that the residuals are not normally distributed. This also holds for the log-transformed data.

2.3 Homogeneity of Model Residuals (original data)

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 1.471843    Df = 1     p = 0.2250552
```

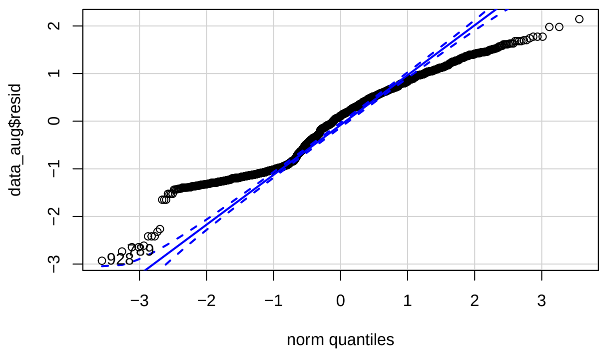

Figure 1: QQ plot of model residuals (original data)

2.4 Homogeneity of Model Residuals (log-transformed data)

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 106.5705    Df = 1     p = 5.528975e-
25
```

The Non-constant Variance Score Test shows a high p-value, meaning that we cannot assume that the variance is not homogeneous. However, this is not true for the log-transformed data, in that test the very low p-value (below 0.05) means that we can reject the null hypothesis and assume non-homogeneity.

2.5 Generalized Linear Hypothesis - Pairwise Comparison

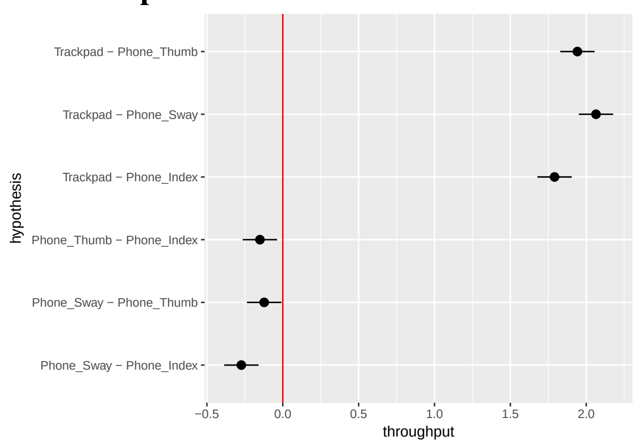

Figure 2: Pairwise Comparison of Conditions.

The pairwise comparison of figure 2 shows that trackpad has a significantly higher throughput than the other conditions, while the other 3 conditions all have a very similar throughput.

3. LEARNING EFFECT

RQ2: To which extent does the learning effect influence the pointing performance?

3.1 Fit the linear model

Fitting the linear model reveals that the blocks have a positive interaction effect on throughput. The p-values are all below 0.05, with positive interaction effects between 0.05 and 0.11, indicating a small learning effect.

3.2 Test assumption on model residuals (original data)

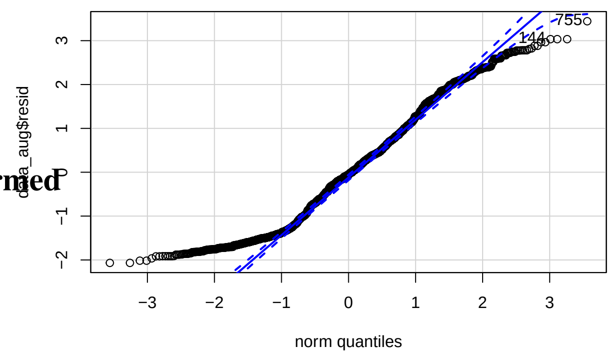

Figure 3: QQ plot of model residuals

```
## [1] 755 144
```

The normality assumption does not hold for the residuals of the data. The QQ plot shows that the data points are not within the blue lines, indicating clearly that the residuals are not normally distributed. This also holds for the log-transformed data.

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.3583796    Df = 1     p = 0.5494076
```

The Non-constant Variance Score Test shows a high p-value, meaning that we cannot assume that the variance is not homogeneous. The same is true for the log-transformed data.

3.3 Test assumption on model residuals (log data)

```
## [1] 755  93
```

```
##    statistic      p.value
## 1 0.07422712 2.111695e-39 Lilliefors (Kolmogorov-
Smirnov) normality test
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.2469969    Df = 1     p = 0.6191976
```

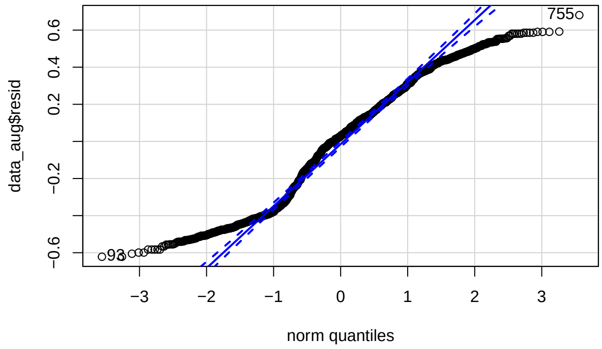

Figure 4: QQ plot of model residuals (log data)

3.4 Model with random intercept

3.5 Analysis of variance (ANOVA)

4. GENERALIZED LINEAR HYPOTHESIS

4.1 Pairwise comparison

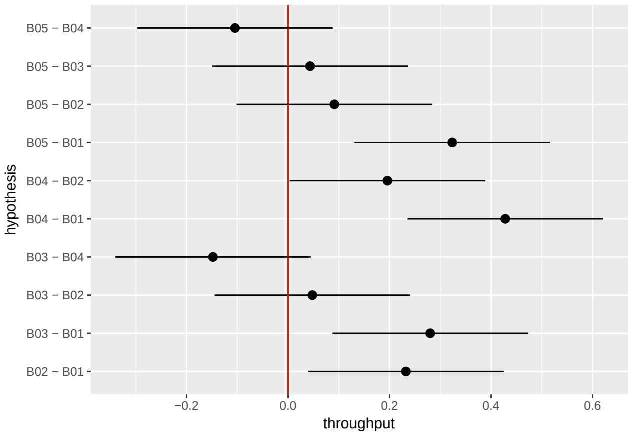

Figure 5: Pairwise comparison of blocks

Finally, the pairwise comparison of figure 5 shows that higher blocks generally have a higher throughput than lower ones, and that the difference is largest for comaprisons with block 1. This confirms that the biggest learning effect is present between block 1 and 2.

5. CONCLUSION

Our analysis with linear models helped us to better understand research questions 1 and 2. It showed that the conditions have a significant interaction effect on the poining performance and that trackpad proved to be the best pointing technique among the four tested techniques, while the other three (thumb, index finger and sway mode) performed similarly well. The analysis for research question two also showed that there is a small learning effect, meaning that participants' pointing performance improved in subsequent blocks, especially in the first few blocks with the effect decreasing over time.

(For grading `group_seed`: 25.66667)

6. REFERENCES

MacKenzie, I.S. 2012. Human-computer interaction: An empirical research perspective. Newnes.