TU Dortmund

Introductory Case Studies

Project 2: Investigating Factors Affecting Distance Perception

Lecturers:

Prof. Dr. Jörg Rahnenführer

JProf. Dr. Antonia Arsova

Dr. Birte Hellwig

M.Sc. Julia Duda

M.Sc. Sven Pappert

Author: Hassan Ali

Group number: 2

Group members: Hassan Ali, André Kafanke, Andrés Arturo Madariaga Ruiz, Sofia Margarita Vergara Puccini

January 5, 2021

Contents

1. Introduction

Distance perception is the process whereby an observer judges the length of an interval between two distinct points in space. This distance does not have to be linear, i.e. in a straight line, but the linear case is most thoroughly studied (Yamamoto 2017). In the scientific literature, multiple studies have been carried out to investigate various factors which could affect the human capability of estimating distances. One such study concluded that there were orientation-dependent biases in judging lengths. Compared to horizontal lines, vertical ones were perceived to be 9.2% ± 2.1% longer on average (Zhu and Ma 2017). On the other hand, another study which compared normal sighted people with low-vision groups found that they were not significantly different in their ability to perceive the dimensions of a room (Legge et al. 2016).

This report is closely related to the two studies cited above (Zhu and Ma 2017; Legge et al. 2016) and considers an additional factor that could potentially affect distance estimation. Here, besides the orientation of the line and the use of the sense of sight, the target length of the line is also taken into consideration. Participants are asked to draw lines of four different target lengths, i.e. 4 cm, 8 cm, 12 cm and 16 cm, in two orientations, i.e. horizontal and vertical, with and without the use of sight.

The statistical analysis of the data comprises of four parts. In the first part, vertical lines are compared with horizontal ones; and in the second part, lines drawn with the eyes closed are compared with those drawn with the eyes open. In both parts, paired samples t-tests are carried out to determine whether there are statistically significant differences in the relative deviations of the measured length of the lines from the target length. In the third part, the relative deviations of the measured length of the lines from the target length between lines of the four different target lengths are compared with each other. This time, a one-way ANOVA test is carried out. In the last part, the ANOVA test is followed up by a Tukey's test to check if there are any pairwise differences in the relative deviations of the measured length of the lines from the target length between the lines of different target lengths.

For each of the tests performed, the assumptions of the tests are first analyzed to judge their reliability. The assumption of normality is examined with Q-Q plots. Box plots, besides being used for data visualization, are also used to check for the presence of outliers.

The goal in carrying out the tests is to determine whether the differences between the lines drawn in the two orientations, between the lines drawn with closed or open eyes, and between the lines drawn with different target lengths in mind, are statistically significant.

The analysis of the data set leads to somewhat counter-intuitive results. The findings suggest that our distance perception is not significantly altered by a change in a line's orientation, its target length or our use of vision.

Besides this Introduction, the report consists of four more sections. The Problem Statement section describes the data set in detail and discusses the data quality. For instance, even though only 21 participants took part in the experiment, there are at least no missing values in this limited available data. The Methods portion of the report provides formulas for the paired samples t-test, the ANOVA test and Tukey's test besides defining some of the terminology used in these statistical tests. The Statistical Analysis section tests the assumptions of these three tests on the data, shows the results of these tests and provides commentary on those results. Finally, the Summary portion concisely rehearses the most important results. It discusses potential improvements to the experiment, such as collecting more data, and suggests ways in which it can be extended, e.g. by also considering non-linear distances instead of just linear ones.

2. Problem Statement

The data is collected from the participants of the course 'Introductory Case Studies' offered at TU Dortmund in the winter semester 2020/2021. Out of the 31 course participants, 21 submitted the data in due time. Each of these 21 students who took the course has done the activity on her or his own. Before data generation, each person was required to draw 4 horizontal lines with a ruler, each exactly equal to the target length of 4 cm, 8 cm, 12 cm and 16 cm respectively. The participants had to look at the lines for one minute to get a feeling for how long they are and then put the sheet of paper away. Next, four lines for each target length were drawn by each person; two horizontal and two vertical lines. Furthermore, for each of these two orientations, the eyes were once closed and once kept open. This sums up to a total of 16 lines per person. Furthermore, the students are split into two groups, A and B; the difference between them being the order in which the 16 lines are drawn by each person from that group. The students in the same group draw the lines in the same order, which is different from the order in which students from the other group draw their lines.

The data set consists of a total to 336 observations for 6 variables, with no missing values. The nominal variable 'ID' takes on the values of the Natural numbers from 1 to 21. It is assigned arbitrarily to each student who submitted the data. The variable 'rand' is nominal and takes on the values 'A' or 'B' indicating the group name. The variables 'horizontal' and 'eye closed' are nominal variables which take on the values '0' or '1'. A value of '1' for the 'horizontal' variable indicates that the line drawn is horizontal, and vice versa. Similarly, a value of '1' for the 'eyes closed' variable indicates that the line is drawn with the eyes closed, and vice versa. Lastly, the target length and the measured length of the line are continuous numeric variables, the latter being accurate to 0.1 cm.

The data quality is consistent. The students are instructed not to draw subsequent lines close to one another or to look at the individual lines after drawing them. This helps reduce the inter-dependency of the readings by preventing the participants from trying to compare the lengths of the lines with each other. Moreover, the lines are to be measured after all 16 of them have been drawn. This is also done to avoid any learning effect while drawing subsequent lines. However, despite all these instructions, it is worth noting that the participants have carried out the experiment on their own, without supervision. Therefore, it is uncertain to what extent the instructions have been followed.

The objectives of the report are to carry out statistical tests on the data to explore whether any significant differences arise in the deviation of the drawn lines from the target lines owing to three factors. These factors are the orientation of the lines, the use of the sense of sight and the target length of the lines. In each case, the assumptions of the tests are also checked to judge the extent to which the results of these tests can be relied on.

3. Statistical Methods

The following statistical methods and mathematical formulas are used. The statistical software R (R Development Core Team 2020), version 4.0.3 has been used for analysis.

3.1. Q-Q Plots

In this report, quantile-quantile or Q-Q plots are scatter plots used to visually check for the assumption of normally distributed samples. The ordered values of the sample, $y_1, y_2, ..., y_n$ are treated as empirical quantiles for the sample. These are plotted on the y-axis against the quantiles of a theoretical normal distribution on the x-axis. The theoretical quantile that corresponds to each ordered empirical quantile y_i is calculated as

$$\Phi^{-1}[p_i]$$

where

$$p_i = \frac{i - (1/2)}{n}$$

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp \frac{-x^2}{2} \, dx$$

(Everitt and A. 2010, p. 339). In R, the *qqline*() function overlays a straight line on the plotted points and connects the 25th and 75th percentiles together. If the points of a Q-Q plot lie roughly on the straight line, this supports the assumption that the sample quantiles originate from a normal distribution.

3.2. Terminology Used in Statistical Tests

An assertion about the distribution of one or more random variables is called a *statistical hypothesis*. A *test* for a certain hypothesis is a procedure used to decide whether the hypothesis should be rejected or not. The *null hypothesis*, denoted by H_0 is the hypothesis that is tested with regards to some parameter θ of a distribution. The *alternative hypothesis* denoted by H_1 is a opposing statement regarding θ. In other words, H_0 is tested against H_1 (Mood 1974, p. 402-403). The parameter space Θ is divided between $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$, where $\Theta_0 \cup \Theta_1 = \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$.

Based on the results of a statistical test, we either reject H_0 or fail to reject it. A *type I error* results when we reject H_0 even though it is actually true. Conversely, accepting H_0 when it is false leads to a *type II error*. (Mood 1974, p. 405)

A hypothesis test is usually specified in terms of a *test statistic*, which is a function of the sample observations (Everitt and A. 2010, p. 427). The numerical value of the test statistic is compared with a *critical value* to decide whether H_0 should be rejected.

The *critical region* is the set of values of the test statistic that cause H_0 to be rejected (Everitt and A. 2010, p. 115). The complement of the critical region or the rejection region is the *acceptance region.* If the value of the test statistic lies in the rejection region, H_0 is rejected. Otherwise, it is accepted.

The concept of *degrees of freedom* is used while calculating some test statistics. This is the number of independent data points in a sample during the statistic's calculation (Everitt and A. 2010, p. 127). The *significance level* $\alpha \in (0, 1)$ is the probability at which H_0 is rejected. This is decided prior to performing the test and is conventionally set to 0.05 (Everitt and A. 2010, p. 393). The *confidence interval* is the range of values which is said to contain the true value of a parameter θ with a certain probability (Everitt and A. 2010, p. 99). This probability is called the *confidence level.* The sum of α and the confidence level equals 1. Consequently, $\alpha = 0.05$, corresponds to a confidence level of 0.95 or 95%.

The value of the test statistic is used to calculate a *p-value.* This is the probability that the test statistic takes on the calculated value, or an even more extreme value, under the assumption that H_0 is true (Everitt and A. 2010, p. 346). H_0 is rejected if, *p-value* $\leq \alpha$.

3.3. Paired Samples t-test

Also known as the dependent samples t-test, this is used when the means of two variables are compared. As this is a paired samples or dependent samples test, both variables are sampled the same group of participants; i.e. if $a_1, a_2, ..., a_n$ and $b_1, b_2, ..., b_n$ are the two samples, then a_i corresponds to b_i. The assumptions of the paired samples t-test are as follows:

- The dependent variable has to be continuous.
- The samples/groups must be related, i.e. both groups have the same subjects.
- The differences between the paired values should be approximately normally distributed.
- The differences between the paired values should not contain any outliers.

The hypotheses are,

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

where μ_1 is the population mean of the first variable and μ_2 is the population mean of the second variable.

The paired samples t-test is based on the T-statistic which, for this particular test, is calculated as follows:

$$t = \frac{\bar{x}_{diff}}{s_{\bar{x}}}$$

where

$$s_{\bar{x}} = \frac{s_{diff}}{\sqrt{n}}$$

where $\bar{x}_{diff}$ is the sample mean of the differences of the paired values; n is the sample size for each group; s_{diff} is the sample standard deviation of the differences; and $s_{\bar{x}}$ is the estimated standard error of the mean.

The calculated t-value is then compared to the critical t-value with $n - 1$ degrees of freedom at the desired significance level α. If $t \geq$ critical value, we reject H_0 (Yeager, Bhattacharya, and Reynolds 2020).

3.4. One-way ANOVA

The one-way analysis of variance test or ANOVA, is used to compare the means of k populations in one variable. Here, $k \geq 3$. The assumptions of the test are as follows:

- The observations are random and the samples observed from the k populations are independent of each other.
- The variances of the distributions in each of the populations are equal to each other.
- The dependent variable should be approximately normally distributed in each group.

The hypotheses are,

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

$$H_1 : \mu_i \neq \mu_j \text{ for at least one } (i, j), \text{ where } i \neq j \text{ and both } i, j \in \{1, ..., k\}.$$

The one-way ANOVA is calculated using the F-statistic. This is a generalization of the T-statistic used to test for the equality of two means. It is given by:

$$F = \frac{\sum n_j(\bar{x}_j - \bar{x})^2/(k-1)}{\sum\sum(x_{ij} - \bar{x}_j)^2/(N-k)}$$

where n_j is the sample size of the jth group; $\bar{x}_j$ is the sample mean of the jth group; $\bar{x}$ is the overall mean for all groups; k is the number of groups; x_{ij} is the ith observation of the jth group; and N is the number of observations in all groups.

The F-statistic has two degrees of freedom, df_1 and df_2, where $df_1 = k - 1$, and $df_2 = N - k$. The calculated value of the F-statistic is compared with the critical F-value with degrees of freedom df_1 and df_2 at the desired significance level α. We reject H_0 if $F \geq$ critical value (Sullivan 2020).

3.5. Tukey's Test

The Tukey's test is usually only carried out when the null hypothesis of the one-way ANOVA is rejected. It is used to identify the pairs of groups whose means are significantly different from each other. As it is used to follow up on the one-way ANOVA, the assumptions of Tukey's test are the same as those for the ANOVA test. If all groups have the same number of observations, the Tukey's criterion is calculated as follows:

$$T = q_{\alpha(k,N-k)} \sqrt{\frac{\sum\sum(x_{ij} - \bar{x}_j)^2/(N-k)}{n}}$$

where α is the significance level; k is the number of groups; N is the number of observations in all groups; $q_{\alpha(k,N-k)}$ is the critical value of the studentized range distribution with significance level α and k and $N - k$ degrees of freedom; x_{ij} is the ith observation of the jth group; and n is the number of observations in each group.

The value of the Tukey's criterion, T is compared with the absolute value of the difference of the means of two groups at a time. If $|\bar{x}_i - \bar{x}_j| \geq T$, where $i \neq j$ and both $i, j \in \{1, ..., k\}$, then we reject H_0 (Abdi and Williams 2010).

In this report, the family-wise confidence level for the Tukey's test is set to 95%. This is the overall confidence level for all pairwise comparisons and corresponds to a type I error rate or α of 0.05. The more the number of pairwise comparisons, the higher the probability that at least one type I error is committed. Therefore, if the overall confidence level has to equal 95%, then each pairwise comparison should have a confidence level higher than 95%. Consequently, the p-value for each pairwise comparison has to be adjusted. The adjusted p-values in a Tukey's test are calculated using the *ptukey*() function in R.

4. Statistical Analysis

4.1. Overall Distance Perception per Participant

Before investigating the factors affecting distance perception in the participants of the experiment, we first consider how distance estimation varies per participant. For this, the relative deviation of the measured length of each line from the target length is calculated. From hence forth in this report, this relative deviation will be used in all analyses.

$$\text{Relative deviation} = \frac{\text{Measured length} - \text{Target length}}{\text{Target length}}$$

There were 16 observations per person. Figure 1 shows the relative deviation of the 16 lines drawn by each of the 21 people as box plots. The graph shows that the participants vary in their ability to perceive lengths. For instance, for ID number 18, the interquartile range of the deviations is very small and there are no readings outside the whiskers of the box plot. This individual is very good at estimating the lengths of the lines. On the other hand, the box plot for ID number 5 has a much larger interquartile range indicating that this person is relatively less proficient at judging line lengths. In the light of these observations, it seems that distance perception greatly varies among individuals.

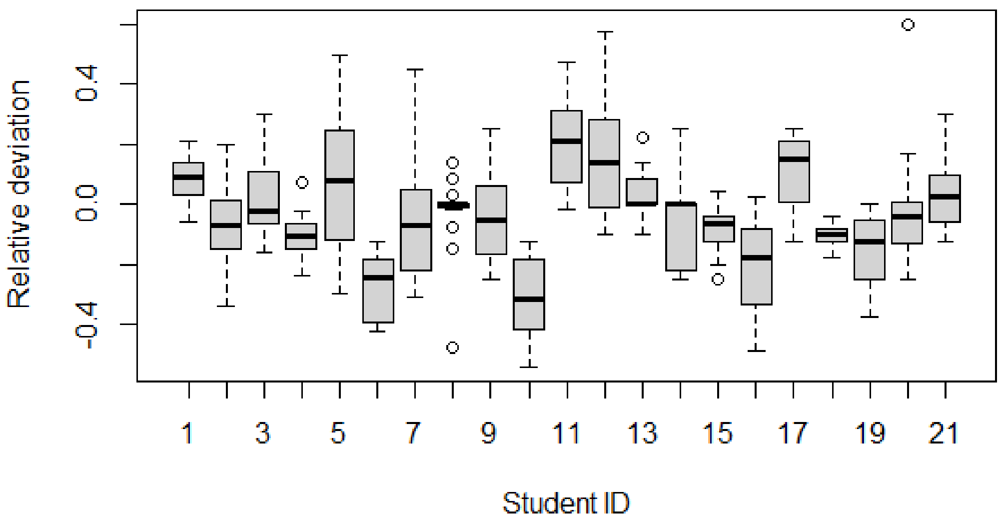


Figure 1: Box plot of the relative deviations for each student ID

The box plot for ID number 8 is different from the others. The interquartile range is extremely small and the median is almost exactly zero. There are 3 outliers above the top whisker and 3 below the lower one. It seems as if this participant has drawn all lines perfectly using, e.g. a ruler, and then intentionally added a few outliers to give the impression of randomness in the readings. Performing the test under supervision would help reduce the skepticism that the readings have been wilfully manipulated.

4.2. The Effect of Orientation on Distance Perception

To study the effect of orientation on distance estimation, the data is divided into two halves: horizontal and vertical lines. The means of the relative deviations of the horizontal lines are compared with those of vertical lines using the paired samples t-test. One of the assumptions of this test is that the observations are independent of each other. As 8 horizontal lines are drawn per person, these lines are dependent on each other as they originate from the same individual. So, for each individual, the mean of the 8 horizontal lines is taken to produce one value per participant. The same is done for the 8 vertical lines per person. This leaves us with 21 readings for each orientation.

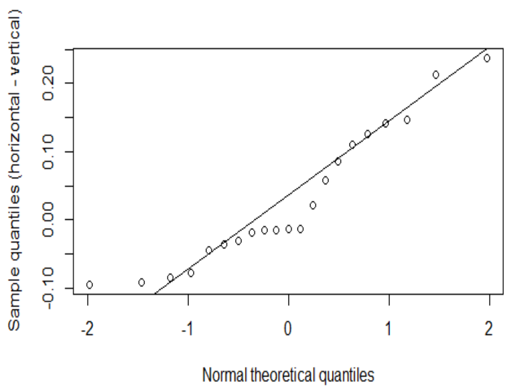


Figure 2: Q-Q plot of the quantiles of the differences of the relative deviations between horizontal and vertical lines, versus normal theoretical quantiles

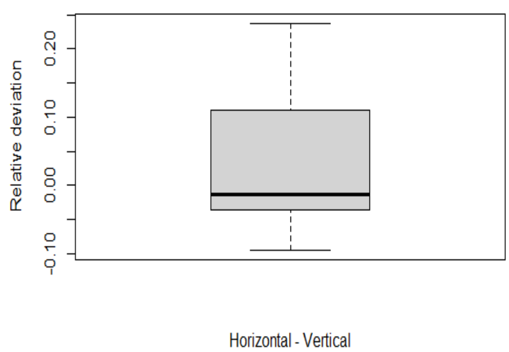


Figure 3: Box plot of the differences of the relative deviations between horizontal and vertical lines

Another assumption of the paired samples t-test is that the differences of the horizontal and vertical relative deviations should be normally distributed. Figure 2 checks for this assumption using a Q-Q plot. The quantiles of the differences of the relative deviations of the horizontal and vertical lines are compared to the quantiles of a theoretical normal distribution. As the plot shows, the points adhere to the straight line connecting the 25th and 75th percentiles very weakly. Points in the center of the plot and on the extreme left diverge from the normal distribution. So the normality assumption is only loosely met, if at all. Nonetheless, as the t-test is relatively insensitive to deviations from the normality assumption (Everitt and A. 2010, p. 420), we carry on with the paired samples t-test.

Figure 8 in the Appendix shows separate box plots for the two orientations. Figure 3 is a box plot of the differences between the relative deviations of the two orientations. The graph shows that there are no points outside the whiskers of the box plot. This supports the assumption of the absence of outliers in the differences of the two groups. Moreover, the graph also shows that the median of the differences is almost zero which suggests that the two means should be similar unless the samples are heavily skewed.

The mean of the relative deviations are -0.0238 and -0.0524 for the horizontal and vertical lines respectively. The mean of the differences in the relative deviations is 0.0285. The results of the paired samples t-test are summarised in table 1. The test has a p-value of 0.2067. As this is larger than the significance level of $\alpha = 0.05$, the difference in the means is not statistically significant. This suggests that the orientation of the line does not affect distance perception.

Table 1: Results of the paired samples t-test comparing the means of the relative deviations of horizontal and vertical lines

t value	degrees of freedom	p-value
1.3051	20	0.2067

4.3. The Effect of Sight on Distance Perception

As in subsection 4.2, for satisfying the independence assumptions, the 8 readings corresponding to closed eyes for each person are averaged to produce one value per participant. The same is done for the 8 readings per student corresponding to open eyes. This leaves us once again with 21 readings for closed and 21 readings for open eyes. The normality assumption is checked using the Q-Q plot in figure 4. The quantiles of the differences of the relative deviations of the two groups are compared to the quantiles of a theoretical

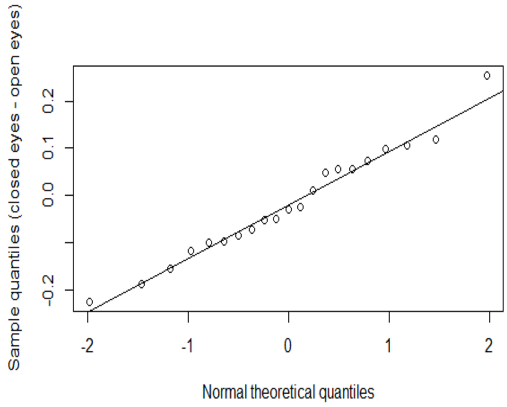


Figure 4: Q-Q plot of the quantiles of the differences of the relative deviations between lines drawn with closed and open eyes, versus normal theoretical quantiles

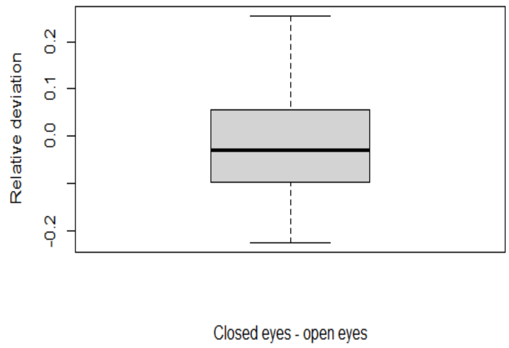


Figure 5: Box plot of the differences of the relative deviations between lines drawn with eyes closed and eyes open

normal distribution. This time around, the points lie close to the straight line joining the 25th and 75th percentiles. The graph supports the assumption that the data is normally distributed.

Figure 9 in the Appendix shows separate box plots for the lines drawn with closed and open eyes. Figure 5 is a box plot of the differences between the relative deviations of the lines, with and without the use of sight. There are no points outside the whiskers of the box plot which supports the assumption of the absence of outliers in the differences of the two groups. This graph also shows that the median of the differences is close to zero. Unless the distributions are very skewed, we would expect the difference in the means to be close to zero.

The mean of the relative deviations are -0.0467 and -0.0295 for the lines drawn with closed and open eyes respectively. The mean of the differences in the relative deviations is -0.0173, which is indeed very close to zero. The results of the paired samples t-test are summarised in table 2. The test has a p-value of 0.5014. As this is larger than the significance level of $\alpha = 0.05$, the difference in the means is once again not statistically significant. This suggests that the perceived distance is not affected by the use or non-use of sight.

Table 2: Results of the paired samples t-test comparing the means of the relative deviations of lines drawn with closed and with open eyes

t value	degrees of freedom	p-value
-0.6847	20	0.5014

4.4. The Effect of Target Length on Distance Perception

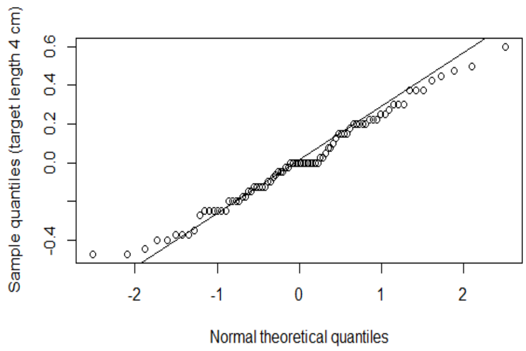
As there are four different target lengths of 4 cm, 8 cm, 12 cm and 16 cm which are being compared in one variable, i.e. the relative deviation, a one-way ANOVA test is used for the analysis. For this analysis, the data is split into four subsets; one for each target length. Each one of the 21 participants drew 4 lines for each target length. These multiple readings per person are not averaged to produce a single value for each participant. Therefore in this analysis, the assumption of independent samples is relatively compromised. There are 84 observations of the relative deviation corresponding to each target length.

The assumption of normality is checked by the Q-Q plots of figure 6. In general, extreme values, i.e. the few largest and smallest quantiles, diverge from the straight line joining the 25th and 75th percentiles. However, the points in the center of the graphs lie close to the line. The exception is the Q-Q plot for the target length of 4 cm, which also has some divergence from the straight line in the center of the plot. The normality assumption is therefore only partially supported by the graphs.

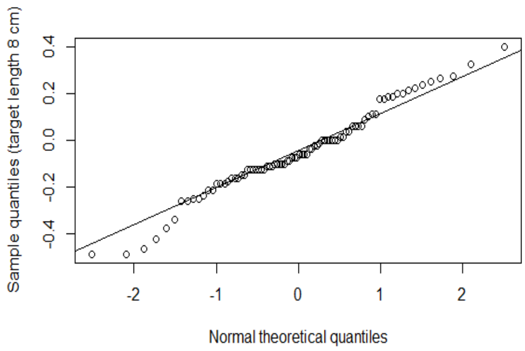
Table 3: The mean and variance of the relative deviations for each target length

Target Length	4 cm	8 cm	12 cm	16 cm
Mean	0.0013	-0.0502	-0.0543	-0.0493
Variance	0.0599	0.0329	0.0261	0.0226

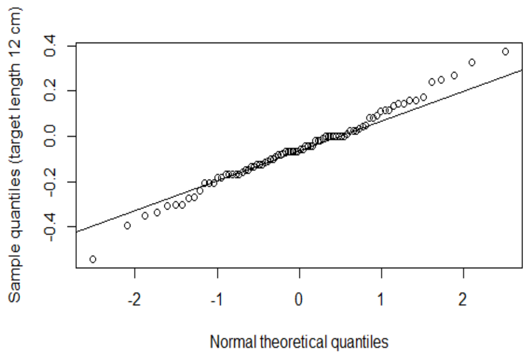
The variance of the relative deviation for each target length is given in table 3. The assumption of equal variances for each target length does not seem to be met. The relative deviation for the target length of 4 cm has roughly thrice the variance when compared to the target length of 16 cm. Nonetheless, the variances for the target lengths of 8 cm, 12 cm and 16 cm are still relatively similar. The same trend is shown by the box plot in figure 10 in the appendix. The interquartile range for the target length of



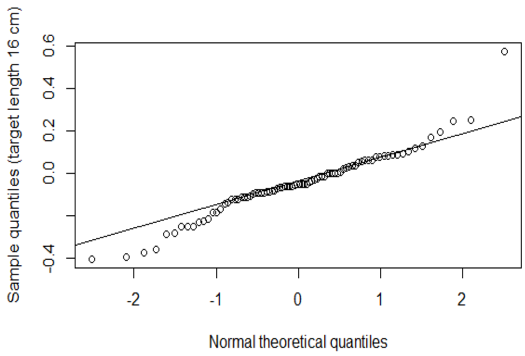
(a) Target Length 4 cm



(b) Target Length 8 cm



(c) Target Length 12 cm



(d) Target Length 16 cm

Figure 6: Q-Q plots of the relative deviations of the lines for each target length

4 cm is larger compared to the other three target lengths, which have comparatively similar interquartile ranges.

The results of the one-way ANOVA are shown in table 4. The p-value of 0.177 is greater than the significance level of $\alpha = 0.05$. This means that the differences among the mean relative deviations of the target lengths are not statistically significant. This supports the claim that the target length does not influence the perceived distance.

Table 4: Results of the one-way ANOVA comparing the mean relative deviation of the four target lengths

F value	degrees of freedom (k)	degrees of freedom (N-k)	p-value
1.652	3	332	0.177

4.5. Checking for Pairwise Differences in Relative Deviation between the Target Lengths

The pairwise comparisons are carried out by Tukey's test which has the same assumptions as those for the one-way ANOVA. The mean relative deviation for each of the four target lengths is given in table 3. The table shows that the means of the lengths 8 cm, 12 cm and 16 cm, all lie roughly around -0.05. The mean of the target length of 4 cm is quite different compared to these three and lies close to 0.001.

The results of Tukey's test are shown in table 5. None of the adjusted p-values is less than the significance level of $\alpha = 0.05$ which is in line with the results of the one-way ANOVA in subsection 4.4. Figure 7 shows that the zero difference line passes through the adjusted confidence interval of each pairwise comparison. Once again, the difference in the mean relative deviations of the three longer target lengths is very close to zero when they are compared to each other. When each of these lengths is compared to the target length of 4 cm, the difference is more pronounced.

Table 5: Results of the Tukey's test comparing the four target lengths pairwise

Target Length Pair (cm)	4-8	4-12	4-16	8-12	8-16	12-16
Difference of Means	-0.515	-0.0556	-0.0507	-0.0041	0.0008	0.0049
Adjusted p-value	0.2877	0.2233	0.3017	0.9990	0.9999	0.9982

Figure 7: A plot of the adjusted confidence levels for each pairwise comparison of the deviations of the target lengths

5. Summary

The report deals with the factors that influence people's perception of distance. Each of the 21 participants of the experiment drew a total of 16 lines for four target lengths combined; the lengths being 4 cm, 8 cm, 12 cm and 16 cm. For each target length, the lines were drawn in a horizontal and vertical orientation, with and without the use of sight. The data set therefore comprises of four main variables of concern for each line drawn: its orientation, the use of sight, the target length and the measured length. For analysis, the relative deviation of the measured length from the target length is used.

The difference in the mean relative deviations of the horizontal and vertical lines is tested using the paired samples t-test. The same test is used to test for the difference of the means in the lines drawn with closed or open eyes. In both cases, the test suggests that there is no significant difference in the means of the two orientations or the means of the lines drawn with the use and non-use of sight. The one-way ANOVA test is used to compare the mean relative deviation among the four target lengths. Once again, the test suggests that there is no significant difference among the means. Likewise, a Tukey's test also suggests that there are no significant pairwise differences between the means of any two target lengths. So none of the three factors affects distance perception.

The assumptions for each test are partially met. The assumption of independent observations is judged by experimental design; Q-Q plots are used to check for normality assumptions; and box plots are used to estimate the variation and central tendency of the observations. Raw values of variances are also used to analyze assumptions.

To improve the reliability of the results, the data should be sourced from more than just 21 participants. Multiple readings from the same individual should not be used in the one-way ANOVA and Tukey's test. The experiments should be conducted under supervision so that compliance with the data collection instructions can be ensured. Additionally, statistical tests can be used to check for assumptions, e.g. the Shapiro-Wilk test can be used to check for normality assumptions. Moreover, non-parametric tests can be used when the assumptions of the parametric tests are not satisfied. Otherwise, these can also be used to support the results of the parametric tests.

As a further extension to the report, non-linear distances, target lengths other than the four used in this report and additional factors affecting distance perception like drunkenness or emotional state, can also be incorporated into the experiment.

References

[1] Hervé Abdi and Lynne J Williams. "Tukey's Honestly Significant Difference (HSD) test". In: *Encyclopedia of Research Design* 3 (2010), pp. 583–585.

[2] Skrondal Everitt B. S. and A. *The Cambridge Dictionary of Statistics, Fourth Edition*. Cambridge University Press, 2010.

[3] Gordon Legge et al. "Indoor Spatial Updating With Impaired Vision". In: *Investigative Opthalmology Visual Science* 57 (Dec. 2016), p. 6757. DOI: 10.1167/iovs.16-20226.

[4] Alexander McFarlane Mood. *Introduction to the Theory of Statistics, 3rd Edition*. McGraw-Hill Inc., 1974.

[5] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2020.

[6] Lisa Sullivan. *Hypothesis Testing - Analysis of Variance (ANOVA)*. Boston University School of Public Health. URL: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_HypothesisTesting-ANOVA/BS704_HypothesisTesting-Anova_print.html (visited on 12/26/2020).

[7] Naohide Yamamoto. *Distance Perception*. Ed. by Caplan B. Kreutzer J. DeLuca J. Encyclopedia of Clinical Neuropsychology, Springer. 2017. URL: https://doi.org/10.1007/978-3-319-56782-2_9103-2 (visited on 12/24/2020).

[8] Kristin Yeager, Preya Bhattacharya, and Victoria Reynolds, eds. *SPSS Tutorials: Paired Samples t Test*. Statistical Consulting at Kent State University Libraries. 2020. URL: https://libguides.library.kent.edu/SPSS/PairedSamplestTest (visited on 12/26/2020).

[9] Jielei Zhu and Wei Ma. "Orientation-dependent biases in length judgments of isolated stimuli". In: *Journal of Vision* 17 (Feb. 2017), p. 20. DOI: 10.1167/17.2.20.

A. Additional figures

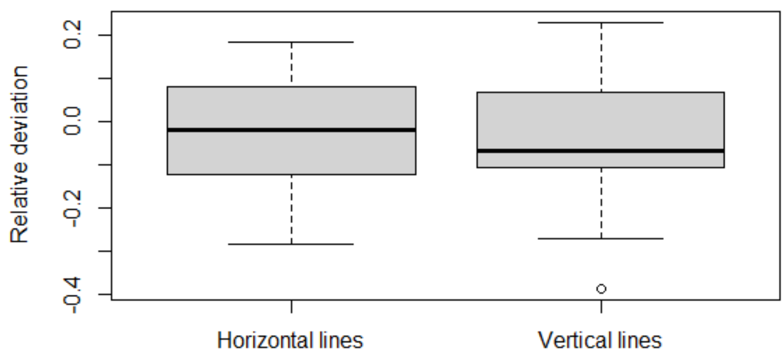


Figure 8: Box plot of the relative deviations of horizontal and vertical lines

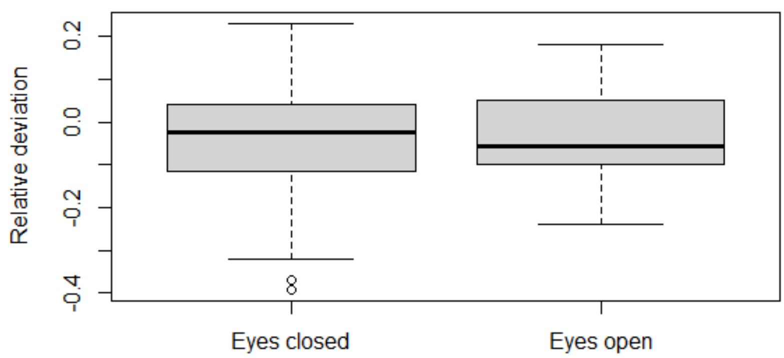


Figure 9: Box plot of the relative deviations of the lines drawn with eyes closed and eyes open

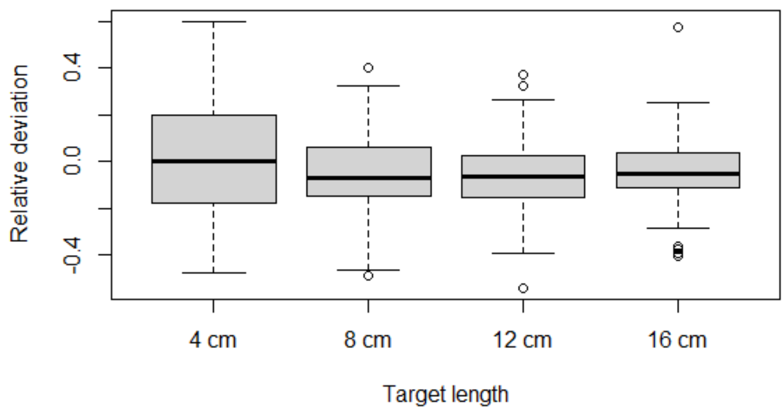


Figure 10: Box plot of the relative deviations for each of the four target lengths