Case Study : Eye-Tracking Analysis

Author: Opeyemi Ayanwale

December 16, 2024

Contents

1 Introduction

Eye-tracking technology enables precise monitoring of where and for how long individuals focus their visual attention within their environment (Duchowski, Andrew T., 2017). Eye-tracking systems use advanced algorithms and techniques to detect eye position and pupil movement: a calibration process maps gaze direction onto the environment, such as screen coordinates, allowing researchers to precisely measure eye movements and determine where participants focus their attention. Fixation points, durations, and saccades are important metrics for studying visual attention patterns (Holmqvist et al., 2011). Eye-tracking is a widely employed tool for behaviour research due to its ability to provide detailed insights into visual attention and cognitive processes (Klaib et al., 2021). However, despite technological advancements, much of the existing research is limited by homogeneity in sample populations, hindering a comprehensive understanding of age-related differences in eye movement behaviours.

The primary goal of this report is to address these gaps by investigating age-related trends in fixation durations and identifying systematic patterns in fixation behaviour over the course of an experiment using the Dutch Science Museum Nemo eye-tracking dataset. The dependent variable, duration, represents the fixation duration of each participant. The dataset also contains several independent variables, including age, gender, order, and other potential metadata or identifiers, such as participant ID, date of birth, coordinates, and timing information (OSF Home, 2024). The original report associated with the data is also available for research purposes (Strauch et al., 2023).

The distribution of fixation durations and participant ages was explored using various visualisation tools such as histograms, scatterplots, and line plots. A linear mixed effects model (LMM) was employed to examine the relationships and systematic patterns between fixation durations and predictors (age, order of each fixation event) while accounting for repeated measures, with participant ID included as a random effect. The assumptions of linearity, normality, homoscedasticity, and residual independence were validated using diagnostic plots: QQ-plot, residual vs. fitted values plots, histogram of residual, and scale-location plots.

The second section provides a more detailed overview of the dataset, description of variables, and information on data quality. The statistical analysis methods are presented and explained in the third section. And in the fourth section, the presentation, analysis, and interpretation of the results are presented. Finally, in the fifth section, the main findings are summarised.

2 Problem Statement

2.1 Dataset and Data Quality

The dataset for this report was collected in 2023 at the Dutch Science Museum Nemo in Amsterdam and provided by the instructors of the case studies course at TU Dortmund during the Winter Semester of 2024. Museum visitors had the option to participate in the study. Participants viewed a feature-rich image for 10 seconds without instructions, and their gaze direction was recorded. Eye tracking data is classified into three types: raw data (participant's original video recording, calibration metadata, and additional information regarding the experimental procedure). This data is not shared due to ethical research requirements (e.g., GDPR). The pre-processed data (tabular format, including timestamps, x/y coordinates, pupil size, events, and quality indicators such as blinks or signal validity), and lastly, the processed data (identifies eye-movement events, fixations, saccades, and gaze shifts between points of interest).

The online repository contains processed data from the original study, including information on the fixations of 2,607 participants. The pre-processed data was run through the processing pipeline to extract fixations, which were combined with the participants' demographic data. The demographic data include participant ID, gender, date of birth (DoB), region of interest (ROI1, ROI2, ROI3), valid and valid free-viewing, while the compiled fixation file includes label, onset, and offset (start and end times of the fixation experiment), and average x and y coordinates of each fixation, order (the order of each fixation event), participant ID, and fixation duration variables. The original data and report associated with the data written by Strauch et al. are available for educational/research purposes on the OSF website (OSF Home, 2024).

To ensure robust analysis and meet our statistical modelling assumptions, the dataset underwent several data preparation and transformation steps, including mapping the participant demographic data to the compiled fixation dataset and calculating participant age from the date of birth. New features (age-scaled and order-squared) are added to the dataset for further use in our analysis and modeling. The dependent variable fixation duration was also transformed to improve reliability by correcting for non-normality in the distribution, making it suitable for model fitting.

2.2 Project Objectives

The primary goal of this analysis is to examine the relationship between participants' age and fixation duration and to investigate potential temporal patterns over the course of the experiment. This will be achieved through two key research hypotheses. To test these hypotheses, the distribution of fixation durations and participant ages will be explored using various visualisation tools, including histograms, scatterplots, and line plots. These exploratory analyses will provide insights into potential skewness, relationships between variables, and temporal trends in the data. A linear mixed effects model (LMM) will be employed to further to examine the relationships and systematic patterns between fixation durations and predictors (age, order of each fixation event) while accounting for repeated measures, with participant ID included as a random effect. The assumptions of linearity, normality, homoscedasticity, and residual independence will be validated using QQ-plot, residual vs. fitted values plots, histogram of residual, and scale-location plots. The statistical method session will go over the hypothesis testing and other methods used to meet the objectives of this report in detail.

3 Statistical Methods

This section outlines the statistical methods used to analyze the dataset based on the objectives of this report. All analyses and visualizations were performed using Python (version 3.11), leveraging key data analysis and statistical libraries, including pandas for data manipulation (McKinney Wes, 2010), numpy for numerical computations (Harris et al., 2020), matplotlib (John D. Hunter, 2007) and seaborn for data visualization (Michael L. Waskom, 2021). Additional analyses utilized scipy (Virtanen et al., 2020) for hypothesis testing, scikit-learn for scaling and preprocessing data (Pedregosa et al., 2011), statsmodels for statistical modeling, including LMM (Skipper and Josef, 2010).

3.1 Hypothesis Testing

Hypotheses in linear mixed models are specified by formulating null (H_0) and alternative (H_A) hypotheses about the parameters of interest. This can also involve comparing a nested model (representing H_0) with a reference model (representing H_A). Hypothesis tests determine whether the inclusion of specific parameters significantly improves the model's explanatory power (West et al., 2022, p. 34).

In this study, we employed hypothesis testing to examine the relationship between participants' age and fixation duration, as well as to investigate systematic patterns in fixation durations over the course of the experiment. Hypotheses were tested using a linear mixed-effects model, with statistical significance assessed through p-values ($\alpha = 0.05$).

Hypothesis 1: Relationship between participants' age and fixation duration:

- H_0: No significant relationship ($\beta_{\text{age}} = 0$).
- H_1: Significant relationship ($\beta_{\text{age}} \neq 0$).

Hypothesis 2: Systematic patterns in fixation durations over the course of the experiment:

- H_0: No systematic patterns ($\beta_{\text{order}} = 0$ and $\beta_{\text{order}^2} = 0$).
- H_1: Systematic patterns ($\beta_{\text{order}} \neq 0$ or $\beta_{\text{order}^2} \neq 0$).

The p-values from the linear mixed-effects model were used to assess the significance of fixed effects, with p-values below 0.05 leading to rejection of H_0.

3.2 Linear Mixed Effects Models (LMM)

Linear Mixed Effects Models (LMM) are an extension of traditional linear models that allow for the inclusion of both fixed and random effects, making them particularly useful for data with multiple levels of variability. These models are commonly used in situations where data are grouped or clustered, such as repeated measures data or hierarchical structures, where observations within the same group may be correlated (West et al., 2022). This analysis employs LMM to investigate the relationship between *fixation duration* and the predictors *age*, *order of event* (and its quadratic term). A quadratic term (order-squared) was generated to capture potential nonlinear relationships. The primary advantage of using this model is its ability to account for *repeated measures* and *within-subject correlation* of observations. This is essential in the context of our data, where each participant (identified by the variable `ID`) provides multiple observations over the course of the experiment. The model allows for the inclusion of both *fixed effects* (such as age and order) and *random effects* (participant-specific variability).

3.2.1 Model Structure

The linear mixed effects model can be expressed as:

$$\text{Fixation Duration}_i = \beta_0 + \beta_1 \cdot \text{Age}_i + \beta_2 \cdot \text{Order}_i + \beta_3 \cdot \text{Order}_i^2 + u_i + \epsilon_i$$

Where:

- $\text{Fixation Duration}_i$ is the fixation duration for participant i.
- $\beta_0, \beta_1, \beta_2, \beta_3$ are the fixed effect coefficients for the intercept, age, age squared, order, and order squared, respectively.
- u_i represents the *random effect* for participant i, capturing the subject-specific variation.
- ϵ_i is the residual error term, representing the variation unexplained by the fixed and random effects.

Fixed Effects

- *Age* and *Order* are treated as fixed effects, assuming that their influence on fixation duration is consistent across all participants.
- The quadratic term $order^2$ is included to account for potential *non-linear relationships* between these variables and the fixation duration.

Random Effects

The random effect for *participant ID* (u_i) accounts for the variation between participants. It assumes that the baseline fixation durations of different participants may vary, but the influence of age and order on fixation duration is assumed to be the same across participants (unless explicitly modelled otherwise).

3.2.2 Model Fitting

The model is fitted using the `mixedlm` function in `statsmodels`, which accounts for both fixed and random effects. The fitting process is done using a maximum likelihood method, with convergence checked to ensure proper model estimation. The model is defined as:

$$\text{duration\_norm} \sim \text{age\_scaled} + \text{Order} + \text{Order\_sq}$$

Age is treated as a continuous variable that has been scaled to have a mean of zero and a standard deviation of one, ensuring easier convergence of the model. On the other hand, Order is treated as a categorical variable and is not scaled, preserving its inherent sequential nature. Since the `mixedlm` package only supports a normal distribution for the dependent variable, the duration variable was transformed using an inverse normal transformation to approximate a normal distribution, making it suitable for model fitting (Wang et al., 2016).

3.2.3 Model Assumptions / Model Diagnostics

Several assumptions must be satisfied for the LMM to produce valid results. It is essential to assess model diagnostics, such as residual plots, to ensure that these assumptions are satisfied and that the model fits the data adequately (West et al., 2022, p. 41):

- **Normality**: The random effects and residual error (ϵ_i) are normally distributed.
- **Linearity**: The relationship between the predictors and the outcome is linear.
- **Homoscedasticity**: The variance of residual errors is constant across all levels of the predictors.
- **Independence**: Observation within each group is independent, and random effects are independent of residual error.
- **Independence of Random Effects**: The random effects bj are assumed to be independent across groups.

These assumptions are validated using QQ-plot, residual vs. fitted values plots, histogram of residual, and scale-location plot (University of Virginia, 2015).

4 Statistical Analysis

This section summarises the distribution of fixation duration and presents the results from the fitted model to evaluate the relationships between the predictors (age and order) and fixation duration.

4.1 Exploratory Analysis

Exploratory analysis of fixation duration and participant ages was conducted using histograms, scatterplots, and line plots to identify potential skewness, variable relationships, and temporal trends. Figure 1 depicts a histogram with KDE (Kernel Density Estimate) and the count of fixation durations. The graph displays a right-skewed distribution, with the majority of fixation durations clustered around 0 to 1, and the majority of fixation durations concentrated around lower values (close to 0), with a rapid decrease in the number of occurrences as the duration increases (see Figure 1).

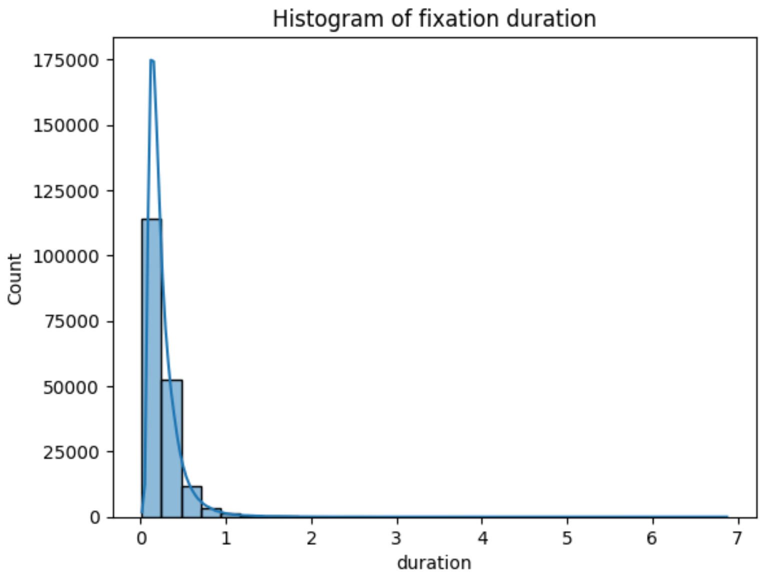


Figure 1: Histogram of Fixation Duration

Figure 2 presents a probability plot (Q-Q plot) that compares the distribution of fixation durations to the theoretical normal distribution. The plot shows significant deviations from the expected straight line, especially in the tail, indicating that the data does not follow a normal distribution (see Figure 2).

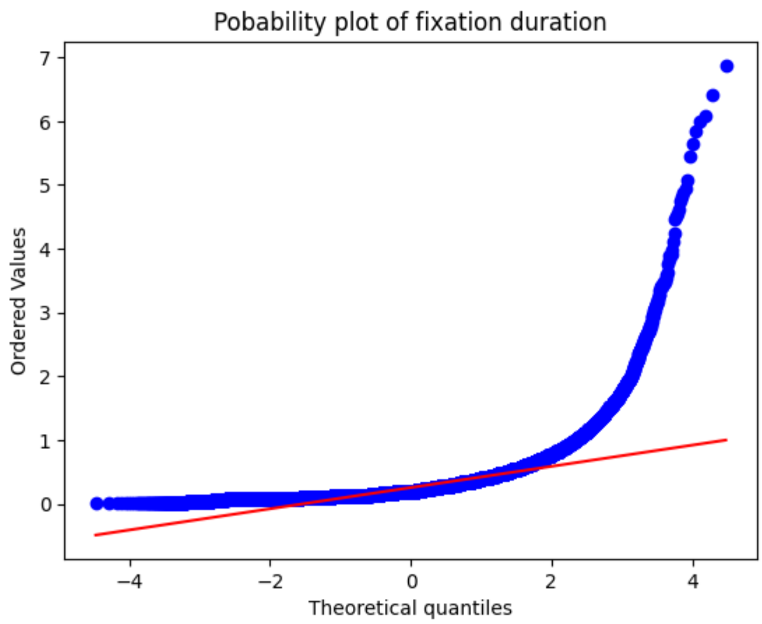


Figure 2: Probability Plot of Fixation Duration

In Figure 3, a scatter plot of age versus fixation duration was plotted to visualize the relationship between age and fixation duration. The plot shows that most data points

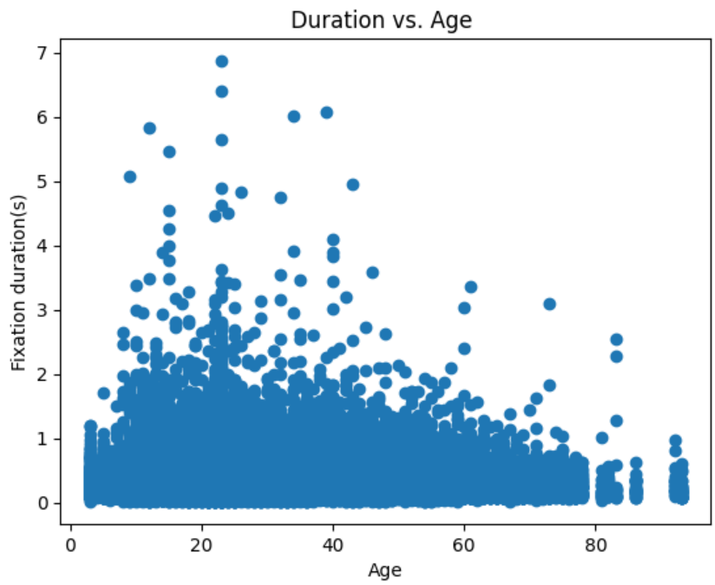


Figure 3: Scatter Plot of Age vs Fixation Duration

cluster around the lower fixation durations (near 0 to 1 on the y-axis). This suggests that people of all ages tend to have shorter fixation durations during the eye-tracking experiment. Younger people tend to have extreme values. The scatter plot does not show a clear linear relationship between age and fixation duration. This could indicate that the two variables do not have an obvious direct relationship (see Figure 3). Figure 4 depicts a line plot of fixation durations over time, with the x-axis representing the order of fixations and the y-axis representing fixation duration.

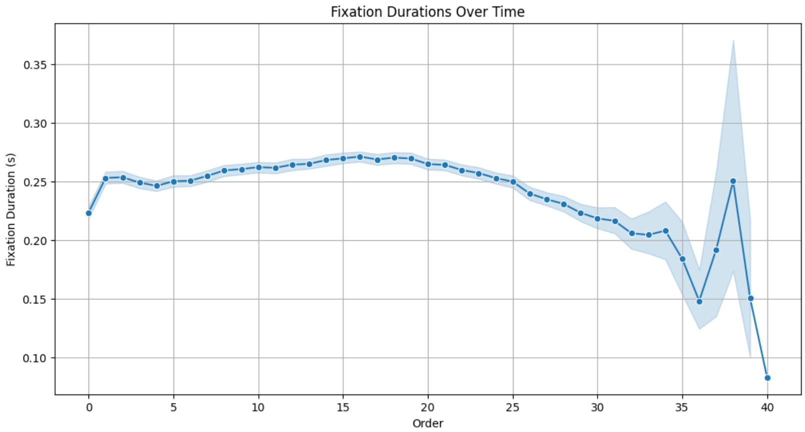


Figure 4: Lineplot of Fixation Duration Over Time

The plot shows a fluctuating trend, indicating that fixation durations change throughout the sequence. This variability suggests that participants may focus their attention more intensely at certain points or intervals, leading to longer fixations, while at other times, the fixations are shorter. These fluctuations could be tied to specific points of interest, distractions, or changes in the task that affect how the participant interacts with the visual stimulus (see Figure 4).

4.2 Linear Mixed Effects Model Analysis

A linear mixed effects model (LMM) was fitted to analyze the relationship between the transformed fixation duration (`duration_norm`) and the predictors. Scaled age (`age_scaled`), fixation order (`Order`), and its quadratic term(`Order_sq`) were included as fixed effects. Participant ID (`ID`) was incorporated as a random effect to account for individual variability. The model included 183,627 observations across 7250 participants. The residual variance of the model was 0.9560, and the log-likelihood was -258814.1027. The model results are summarised in Table 1 and the diagnostic plots are presented in Figure 5 of the Appendix on page 15.

Table 1: Mixed Linear Model Regression Results

Variable	**Coefficient**	**Std. Error**	**z**	**P>|z|**	**[0.025, 0.975]**
Intercept	-0.153	0.006	-23.983	0.000	[-0.166, -0.141]
Age (scaled)	-0.004	0.003	-1.261	0.207	[-0.010, 0.002]
Order	0.025	0.001	24.542	0.000	[0.023, 0.027]
Order_sq	-0.001	0.000	-17.690	0.000	[-0.001, -0.001]
Group Var	0.035	0.001			

The fixed effects show that:

- **Age Effects:** The coefficient for the linear term of age (`age_scaled`, with $p = 0.207$) is not statistically significant at $\alpha = 0.05$ level, suggesting that age does not have a significant linear effect on fixation duration in the model (see Table 1).
- **Order Effects:** Both the linear (`Order`, $p < 0.001$) and quadratic (`Order_sq`, $p < 0.001$) terms for fixation order were statistically significant. The positive coefficient for `Order` indicates that fixation duration tends to increase as the experiment progresses. However, the negative coefficient for `Order_sq` suggests a non-linear relationship, where the increase in fixation duration starts to slow down or may even decline after a certain point in the sequence (see Table 1).

The random effects:

- **Group Variance (Participant ID):** The variance for the random effect (group level) is 0.035, indicating variability in fixation durations across participants. There is significant between-participant variability in fixation behaviours, suggesting that individual differences play a role in fixation duration (see Table 1).

4.3 Model Diagnostic

The QQ-plot of residuals in Figure 5a of the Appendix on page 15 shows that the residuals closely align with the red diagonal line throughout the range. This shows that the residuals have a normal distribution. Figure 5c in the appendix on page 15 shows a symmetrical, bell-shaped histogram, indicating that the residuals are normal. The overlaid smooth density curve provides additional evidence of the residuals' alignment with the normal distribution, confirming that the model's normality assumptions are met (see Figures 5a and 5c of the Appendix on page 15).

Figures 5b and 5d of the Appendix, on page 15, depict the residual vs. fitted plot and the standardised residuals vs. fitted values. These plots evaluate the assumption of homoscedasticity and assist in identifying patterns in the residual spread. The plot shows that the residuals are fairly distributed around the horizontal line at zero, but there is a subtle banding pattern and some minor heteroscedasticity. The standardised residuals have a relatively constant spread across the fitted values, but, like the residual plot, there is a visible lower band where residuals are systematically more negative. However, the deviations are not extreme, indicating that the model's overall performance remains robust (see Figures 5b and 5d of the Appendix on page 15).

4.4 Robustness Analysis

To ensure the stability of our findings, we conducted a robustness analysis by adding random noise to the fixation duration values. This analysis involved several steps, including visual inspection, data transformation, model fitting, and diagnostic evaluations as with the original dataset. Figure 6a of the Appendix on page 16 depicts a boxplot comparing fixation durations between the original and noisy datasets, which revealed that adding random noise did not significantly impact the data's central tendency, variability, or outlier distribution. Both datasets exhibit comparable medians, IQR, and the presence of outliers. Similarly, the histogram and probability plot of fixation durations in Figures 6b and 6c with the noisy dataset have similar shapes as the original data. The scatterplot of age versus fixation duration in Figure 6d for the noisy dataset reflects the same trend as in the original analysis, with a slight deviation observed for the older participant. The relationship between age and fixation duration remains consistent (see Figure 6 of the Appendix on page 16). The linear mixed model with the noisy dataset produced the following results on Table 2 of the Appendix, on page 15. The Order and Order (squared) terms remain highly significant ($p < 0.001$), indicating that the non-linear relationship between the order and fixation duration is robust to the addition of noise. Age (scaled) remains non-significant ($p = 0.692$), consistent with the findings in the original analysis. The intercept remains similar, suggesting no major shifts in the baseline fixation durations (see Table 2 of the Appendix on page 15). The robust diagnostic plots on Figure 7 of the Appendix on page 17 (Q-Q plots, residuals vs fitted, histogram of residuals and scale location plot) for the noisy dataset were virtually identical to those from the original analysis (see Figure 7 of the Appendix on page 17).

5 Summary

The data used in this report is an excerpt from the 2023 Dutch Science Museum Nemo eye-tracking dataset from Amsterdam and was provided by the instructors of the case studies course at TU Dortmund during the Winter Semester of 2024. The objective of this analysis is to examine the relationship between participants' age and fixation duration and to investigate potential temporal patterns over the course of the eye-tracking experiment. A histogram and probability plot was used to visualise the distribution of the fixation duration (dependent variable). The distribution appeared approximately symmetric and right-skewed, prompting normalization to standardize the range. The majority of participants had shorter fixation durations and concentrated in the lower range of fixation duration values. A scatterplot was used to explore the relationship between participant age and fixation duration. The relationship appeared weak, with no strong linear trend observed. The lineplot showed a trend where fixation duration increased with fixation order initially but plateaued and slightly declined later. This suggested a potential non-linear relationship between order and fixation duration.

To account for individual differences (participants) and test the predictors' effects on fixation duration, a linear mixed model (LMM) was fitted. For the first hypothesis, the result shows that the linear effect of age on fixation duration is not statistically significant; therefore, we fail to reject $H0$ as there is no sufficient evidence to support a significant relationship between participants' age and fixation duration. For the second hypothesis, the linear term (Order) is positive and significant, and the quadratic term (Order_sq) is negative and significant therefore we reject $H0$ as there is strong evidence of systematic temporal patterns in fixation duration over the course of the experiment. Also, there is substantial variability in fixation duration at the participant level, suggesting individual differences play an important role. Diagnostic checks were conducted to validate the LMM assumptions. The residuals were symmetrically distributed around zero with no strong patterns. However, slight heteroscedasticity was observed (residual spread increases slightly with fitted values). Standardized residuals were also roughly constant across fitted values, confirming acceptable model assumptions. To ensure the stability of our findings, we conducted a robustness analysis by adding random noise to the fixation duration values. The results from the robustness analysis remain stable after introducing random noise to fixation durations. The significance of key relationships, particularly the non-linear effect of order, was preserved, and model diagnostics showed no degradation. These findings confirm the reliability of the original analysis.

Bibliography

Duchowski, Andrew T. (2017). *Eye Tracking Methodology: Theory and Practice*. Springer, New York, NY.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. (2020). Array programming with numpy. *Nature*, 585(7825):357–362. `https://www.nature.com/articles/s41586-020-2649-2`.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. Oxford University Press, Oxford. url = https://global.oup.com/academic/product/eye-tracking-9780199697083?cc=uslang=en.

John D. Hunter (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95. `https://ieeexplore.ieee.org/document/4160265`.

Klaib, A. F., Alsrehin, N. O., Melhem, W. Y., Bashtawi, H. O., and Magableh, A. A. (2021). Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and internet of things technologies. *Expert Systems with Applications*, 166:114037. `https://www.sciencedirect.com/science/article/pii/S0957417420308071`.

McKinney Wes (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, pages 51–56. SciPy. `https://proceedings.scipy.org/articles/Majora-92bf1922-00a` (visited on 1st December 2024).

Michael L. Waskom (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60):3021. `https://joss.theoj.org/papers/10.21105/joss.03021`.

OSF Home (2011-2024). Center for open science. `https://osf.io/sk4fr/` (visited on 1st December 2024).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830. `https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf`.

Skipper, S. and Josef, P. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, pages 57–61. SciPy. https://proceedings.scipy.org/articles/Majora-92bf1922-011 (visited on 1st December 2024).

Strauch, C., Hoogerbrugge, A. J., and Baer, G. (2023). Saliency models perform best for women's and young adults' fixations. https://osf.io/qwfr6 (visited on 1st December 2024).

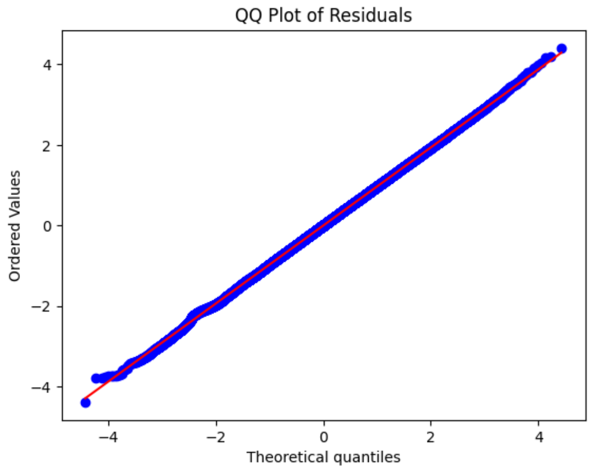
University of Virginia (2015). Understanding diagnostic plots for linear regression analysis. https://library.virginia.edu/data/articles/diagnostic-plots (visited on 14th December 2024).

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272. https://www.nature.com/articles/s41592-019-0686-2.

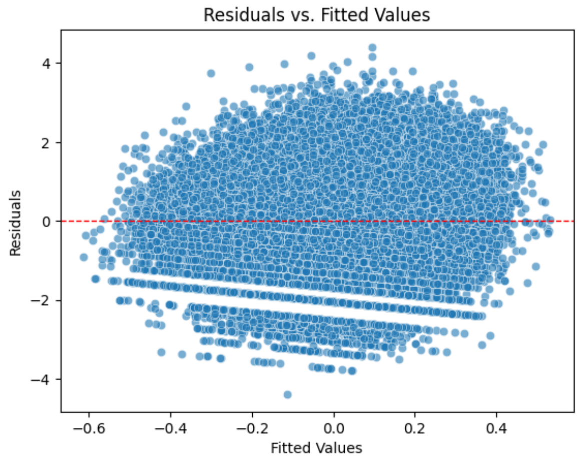
Wang, H., Yang, J., and Hao, S. (2016). Two inverse normalizing transformation methods for the process capability analysis of non-normal process data,. *Computers Industrial Engineering*, 102:88–98. url = https://www.sciencedirect.com/science/article/abs/pii/S0360835216303837.

West, B. T., Welch, K. B., and Galecki, A. T. (2022). *Linear Mixed Models: A Practical Guide Using Statistical Software*. CRC Press, Boca Raton, FL, 3rd edition. url = https://www.taylorfrancis.com/books/mono/10.1201/9781003181064/linear-mixed-models-brady-west-kathleen-welch-andrzej-galecki.

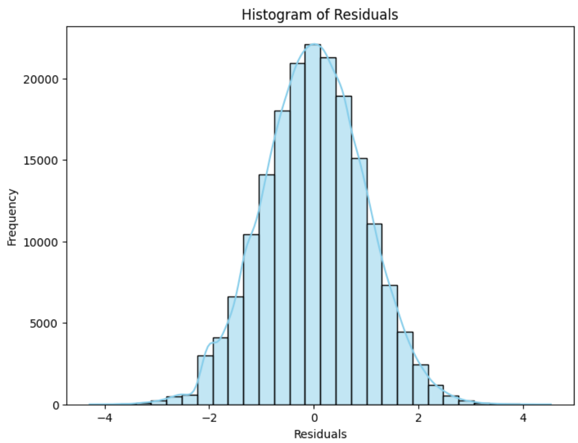
Appendix



(a) QQ-Plot of Residual



(b) Residual vs. Fitted Values Plot



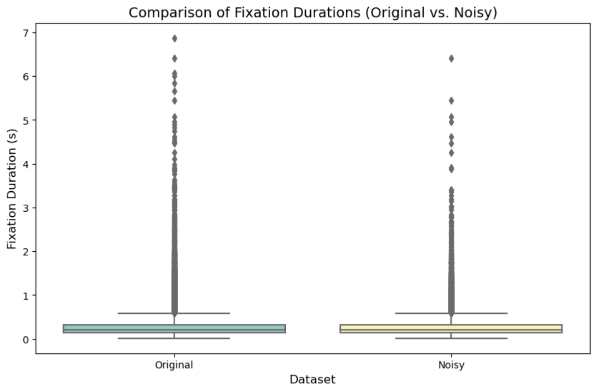
(c) Histogram of Residual

(d) Scale-Location Plot

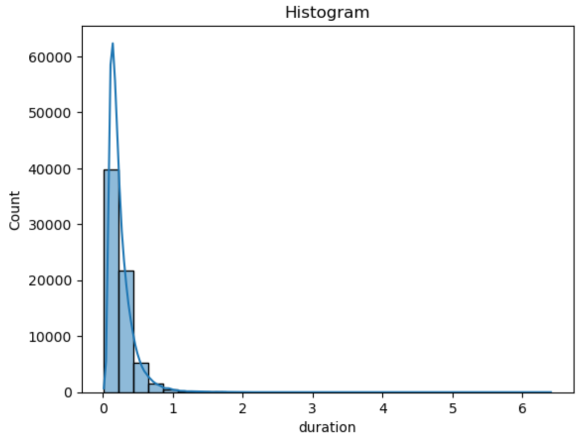
Figure 5: Diagnostic Plots.

Table 2: Linear Mixed Effects Model Results with Noisy Data

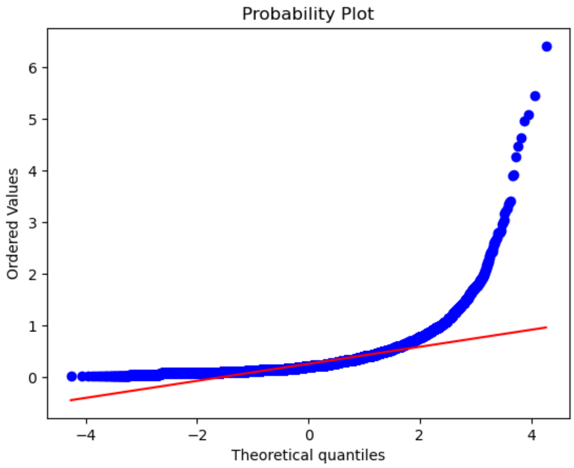
Variable	**Coefficient**	**Std. Error**	**z-value**	**P-value**	**95% CI**
Intercept	-0.153	0.010	-14.724	0.000	[-0.173, -0.133]
`age (scaled)`	-0.002	0.005	-0.396	0.692	[-0.012, 0.008]
`Order`	0.025	0.002	15.213	0.000	[0.022, 0.028]
`Order (sq)`	-0.001	0.000	-11.102	0.000	[-0.001, -0.001]
Group Variance	0.035	0.002			



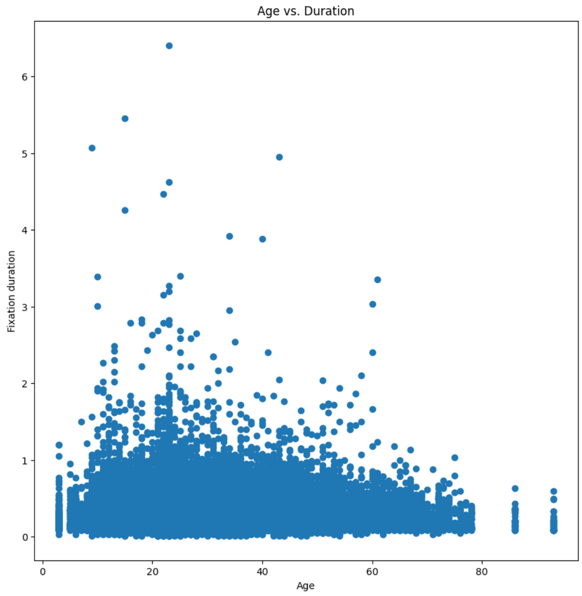
(a) Original vs. Noisy Dataset



(b) Robust Histogram of Fixation Duration

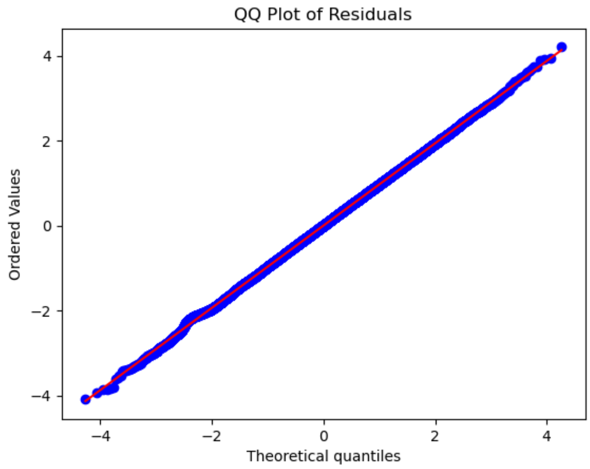


(c) Robust Prob Plot of Fixation Duration

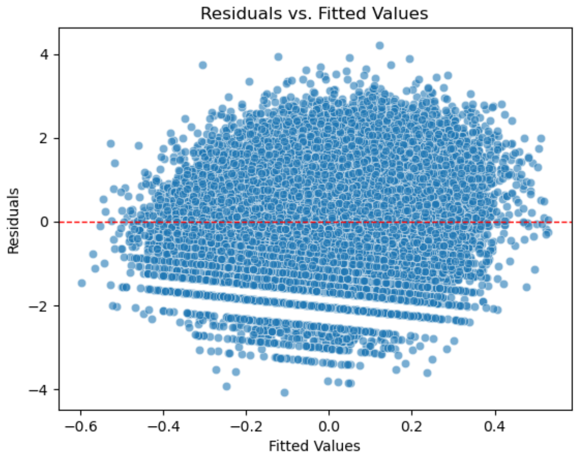


(d) Robust Scatter Plot of Age vs Fixation Duration

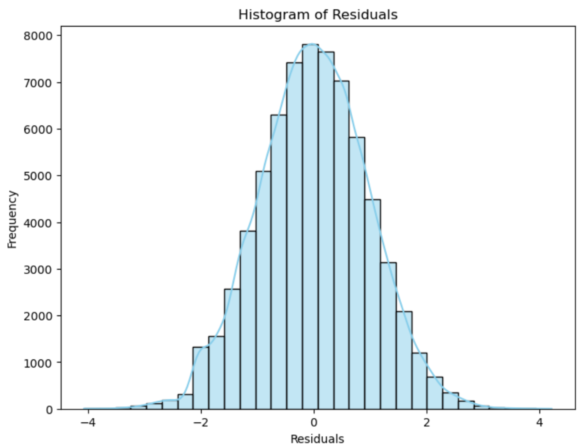
Figure 6: Robust Exploratory Analysis Plots.



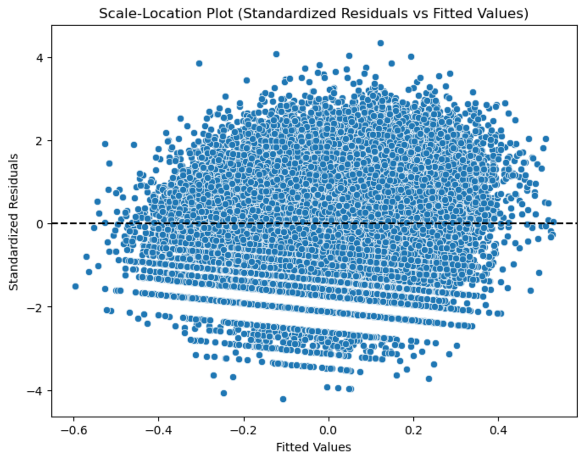
(a) Robust-QQ-Plot of Residual



(b) Robust Residual vs. Fitted Values Plot



(c) Robust Histogram of Residual



(d) Robust Scale-Location Plot

Figure 7: Robust Analysis Diagnostic Plots.