Final Project Report; From Zero to Data Analysis, Eye-Tracking, and LMM

Ali Ibraheem

David Raul Carranza Navarrete

Mitra Gholami

Seyedamirhossein Siadati

Context

Introduction to the study

For this group project, we use the dataset first introduced in the paper ”VREED: Virtual Reality Emotion Recognition Dataset Using Eye Tracking and Physiological Measures”. (Tabbaa et al., 2021)

The Virtual Reality Emotion Recognition Dataset (VREED) combines eye tracking and physiological measures to assess participants' emotional responses, while interacting with different 360° video environments (360-VEs). Specifically, the authors explore the effectiveness of Virtual Reality (VR) in eliciting emotions in immersive environments, drawing on principles of affective computing which highlight the role of technology in measuring emotional responses with greater accuracy (Wang et al., 2022).

Through prior pilot testing, a total of 12 360-VEs were chosen to represent a range of emotions. The authors categorized the 12 VEs into four quadrants based on the Circumplex Model of Affects (CMA), evaluating emotions through arousal (high vs. low) and valence (positive vs. negative). 34 participants were recruited to engage with these selected stimuli for 1-3 minutes in VR, during which their physiological signals (ECG and GSR) and eye tracking data were collected along with self-reported emotional ratings.

The authors performed preliminary machine learning analyses to establish a baseline for emotion recognition. The results validated the authors' hypothesis, suggesting that immersive VR environments can significantly enhance emotional elicitation and recognition.

Dataset and variables

VREED is publicly available on Kaggle. VREED comprises a number of datasets, including the eye tracking data, the Electrocardiogram (ECG) measures, the Galvanic Skin Response (GSR) data, and the data extracted from self-reported questionnaire. We decided to work with eye tracking data, which includes the following eye tracking features: fixations, saccades, microsaccades, and blinks extracted per quadrant.

The dataset contains a total of 50 variables; however, for the scope of this project, we decided to focus on the following two categories:

1. Fixation Metrics to capture the duration of participants' fixations within each quadrant. 2. Blink Metrics to gain insight into participants' blink rate, which may reflect the emotional states across quadrants.

The table below shows the variables we considered.

Variable	Description
subjectID	A unique ID assigned to each participant
Quad_Cat	Quadrant category
Mean_Fixation_Duration	Mean fixation duration
Num_of_Blink	Number of blinks

Table 1: Description of variables

Quadrant category is a nominal categorical variable where 0= High Arousal and High Valence, 1= Low Arousal and High Valence, 2 = Low Arousal and Low Valence, and 3=High Arousal and Low Valence.

Hypothesis

Eye movements, including the number of blinks and the duration of fixations, reflect underlying cognitive and attentional processes (Skaramagkas et al., 2021). We hypothesize that fixation duration and blink numbers can predict the quadrant categories corresponding to the emotional response while accounting for individual differences using bayesian regression models.

H1

We predict that longer fixation duration and fewer blink numbers are associated with quadrant category 1 (Low Arousal and High Valence), while shorter fixation durations and higher blink numbers are associated with quadrant category 3 (High Arousal and Low Valence).

H0

Fixation duration and blink numbers do not significantly predict the Quad category, indicating no relationship between eye movement patterns and quadrant categories.

Methods and Justification

For this project, our primary goal is to investigate how blink frequency and fixation duration predict the likelihood of participants's emotional states across different quadrants. Specifically, the analysis aims to evaluate whether these eye tracking metrics vary across each quadrant and identify potential interactions between blink and fixation behaviors.

The Analysis of choice and why it is appropriate

Initially, we planned to use linear mixed-effects models (LMM) via the **lmer** function. However, due to the multinomial and categorical nature of our dependent variable (Quad_Cat), this method was unsuitable.

Instead, we use the **brm** function from the **brms** package, which supports *Bayesian hierarchical modeling* and accommodates multinomial outcomes.

This approach allows us to model the probability of each quadrant category while accounting for both fixed effects (blink and fixation metrics) and random effects (subject-level variation). Bayesian methods also provide meaningful inferences through posterior distributions and credible intervals, improving interpretability. This method is appropriate for answering our research question, because (1) our dependent variable (Quad_Cat) is categorical rather than continuous, which violates the assumptions of LMM, (2) the categorical nature of the outcome, which requires a model that can handle multinomial data rather than linear relationships, and (3) the discrete outcomes, meaning that each individual observation results in one distinct categorical outcome.

The brm function fits *Bayesian regression models* using the brms package and Stan probabilistic programming language, to perform full Bayesian inference using Markov Chain Monte Carlo (MCMC) methods or approximate algorithms like variational inference. For this project, the outcome has more than two categories; therefore, multinominal logistic regression would be the appropriate model to fit the data. The multinomial logistic regression model is specified as:

$$\log\left(\frac{P(y=j\mid x)}{P(y=k\mid x)}\right)=\beta_{0j}+\beta_{1j}x_1+\beta_{2j}x_2+\cdots+\beta_{pj}x_p$$

$$\text{for}\quad j=1,2,\ldots,k-1$$

Where:

- y is the categorical outcome variable with k categories.
- $P(y=j\mid x)$ is the probability of y being in category j (compared to the reference category k).
- β_{0j} are the intercepts for each category j (compared to the reference category).
- β_{ij} are the coefficients for predictor variables $x_1,x_2,\ldots,x_p$ for category j.
- $x_1,x_2,\ldots,x_p$ are the predictor variables.

We fit four models to investigate how fixation and blink metrics predict quadrant category:
Model 1: Number of blinks as a fixed effect.
Model 2: Mean fixation duration as a fixed effect.
Model 3: Both metrics as fixed effects.
Model 4: Interaction between blink and fixation metrics.

This approach enables us to capture complex relationships between eye-tracking measures and emotional responses while properly accounting for the categorical nature of the outcome variable.

Results and Interpretation

Results and findings

The summaries of the four Bayesian multinomial logistic mixed models provide parameter estimates (posterior means), standard errors, and 95% credible intervals (CIs) for each predictor. Model summaries are available in the script added to the GitHub repository. These outputs quantify the effects of blink and fixation metrics on the likelihood of quadrant category assignment while accounting for subject-level variability.

All models show strong Markov Chain Monte Carlo (MCMC) convergence, with Rhat values near 1.00 and large effective sample sizes, which confirm reliable parameter estimation and accurate posterior distributions. Overall, these models indicate that blink frequency and its interaction with fixation duration significantly influence quadrant category probabilities.

Model comparison

We performed model comparison using the Leave-One-Out Cross-Validation Information Criterion (LOOIC) to evaluate predictive accuracy across the four models. A lower LOOIC value indicates better predictive performance.

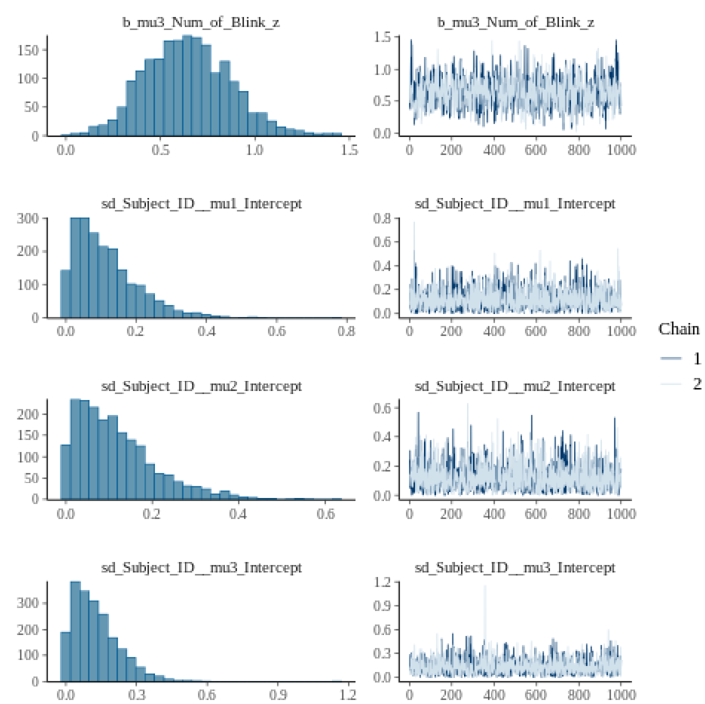

Figure 1: Summary for Model 1.

This comparison revealed that Model 1 (using the number of blinks as a predictor) has the lowest LOOIC value (873.9977), indicating the best fit. Additionally, model 4 (which includes the interaction of the two fixed effects) follows closely with a LOOIC of 875.7951. Based on LOOIC analysis, Model 1 has better predictive accuracy and provides the most reliable fit for the data, while Model 4 suggests potential interactive effects worth further exploration.

Model 1 diagnostic

Model 1 shows a significant positive association between the number of blinks and the probability of quadrant assignment, indicating that increased blink frequency is associated with specific quadrant categories. Additionally, model 4 reveals an interaction effect between blink frequency and mean fixation duration, suggesting their combined influence on quadrant assignment is interdependent rather than additive.

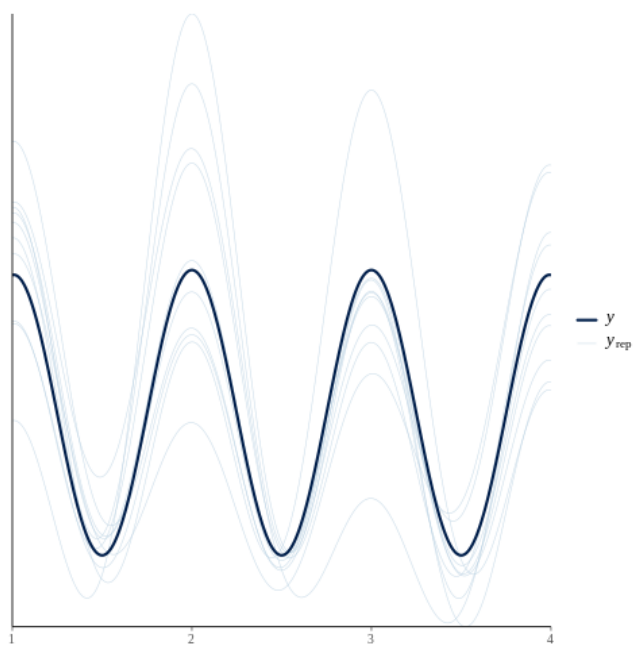

Figure 2: Model Evaluation showing recovered parameters vs original parameters.

Additionally, the diagnostic plots for Model 1 (Figure 1) validate these findings. The posterior density plots show well-defined peaks and appropriate spread, especially for the effect of blink frequency, indicating informative parameter estimates. The trace plots for the two MCMC chains show consistent mixing and overlap across iterations, which reflects stable and reliable parameter estimation. These diagnostics confirm the robust fit and convergence of the model.

Model performance evaluation

To assess the model's predictive performance, we performed a posterior predictive check (PPC) using 10 posterior draws with the 'dens_overlay' method. The plot compares the observed data (y) with model predictions (y_rep) from the model. The close alignment between the dark line (y) and the lighter lines (y_rep) shows that the model predictions closely match the observed data. As shown in Figure 2, the model fit is strong, supporting the model reliability.

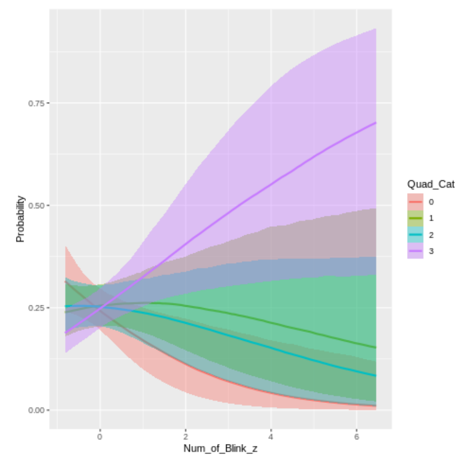

Figure 3: Conditional effects for model 1

Visualization of predicted effects

Furthermore, we visualized the predicted effects of the number of blinks on the probability of quadrant assignment for each model:

- Model 1

Figure 3 illustrates how increases in the number of blinks influence the likelihood of belonging to each quadrant category. Particularly, as blink frequency increases, the probability of being in category 3 rises, while the probabilities for other categories decrease. This visualization highlights the strong positive association between increased blink frequency and the likelihood of quadrant category 3, consistent with the model's estimates and diagnostics.

- Model 2

Figure 4 shows how the average fixation duration impacts the probability of each quadrant category. As fixation duration increases, category 2 becomes more likely, while the probabilities for categories 0, 1, and 3 decrease. However, the highly overlapping shaded areas convey the uncertainty in these estimates. This indicates that longer fixation durations are slightly associated with a higher probability of category 2.

- Model 3

Figure 5 demonstrates the influence of the number of blinks and fixation duration jointly on the likelihood of each category. As the number of blinks and the duration of

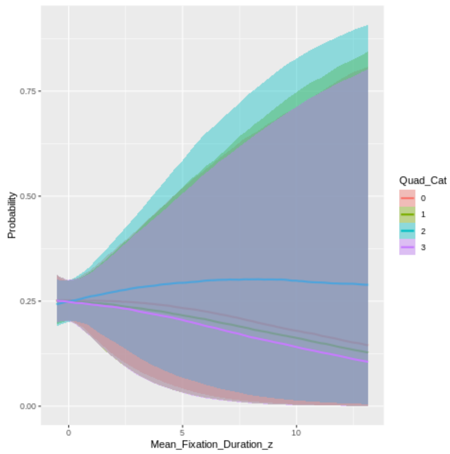

Figure 4: Conditional effects for model 2

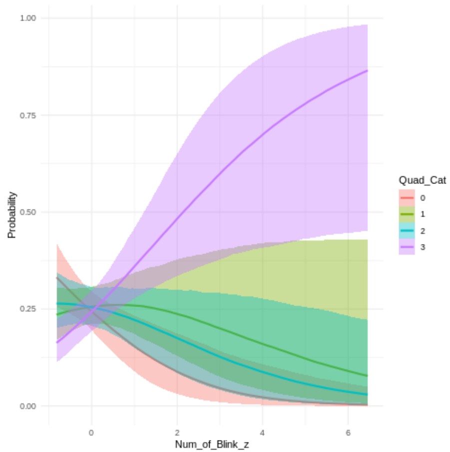

Figure 5: Conditional effects for model 3

fixations increase, category 3 becomes more probable, while categories 0, 1, and 2 decrease. This reinforces the finding that increased blink frequency and fixation duration drives up the chances for category 3.

- Model 4

Figure 6 reveals how the number of blinks interacts with mean fixation duration to affect category probabilities. As the interaction effect increases, the likelihood of category 3 rises significantly, while categories 0 and 1, and 2 decrease.

Implications

Overall, the analysis indicated that the model using blink frequency as a predictor provided the best fit with the lowest LOOIC, reflecting its superior ability to predict the categorical outcome in terms of quadrant category. This suggests a significant relationship between blink frequency and category assignment, implying that eye movement metrics like blinks play a crucial role in differentiating emotional states. These findings contribute

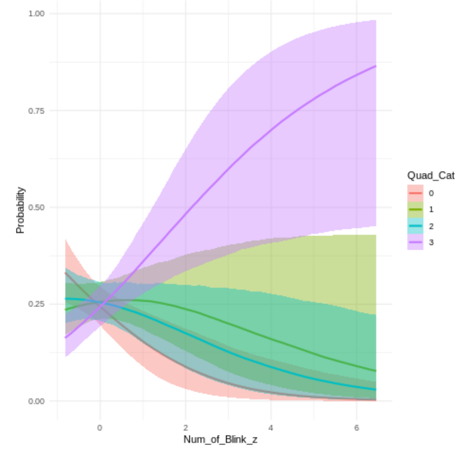

Figure 6: Conditional effects for model 4

to our understanding of eye-tracking metrics as informative indicators of underlying psychological processes.

Conclusion

The analysis supports the hypothesis that eye-tracking features, such as blink frequency, significantly influence the probability of categorical outcomes. The model capturing the effect of blink frequency aligned well with observed data, validating its predictive relevance and supporting the concept that variations in eye movement can reflect distinct emotional states.

Authors' contributions

All members of the group have contributed equally with the conception of the hypothesis, implementation of the models, analysis of the results, and writing of the final report.

References

Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., … Tsiknakis, M. (2021). Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering*, *16*, 260–277.

Tabbaa, L., Searle, R., Bafti, S. M., Hossain, M. M., Intarasisrisawat, J., Glancy, M., & Ang, C. S. (2021). Vreed: Virtual reality emotion recognition dataset using eye tracking & physiological measures. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, *5*(4), 1–20.

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., … others (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, *83*, 19–52.