Group 21: Cognitive Modeling - Lab 1

Questions 1-4

Olusanmi Hundogan - 6883273
Evangelia Giannikou - 6988229

November 20, 2020

Question 1

1.a Mean, SD and SE for total dialing time of 2 conditions

For the two experimental conditions (steering focus or "dualSteerFocus"; dialing focus or "dual-DialFocus") report the mean (M) of the total dialing time, standard deviation (SD), and standard error of the mean (SE) of total dialing time.

Condition	Mean	SD	SE
DualDialing	3.43 s	0.90	0.26
DualSteering	5.07 s	1.32	0.38

1.b Mean, SD and SE for lateral deviation of 2 conditions

Make a similar table for average absolute lateral deviation (in m), when you only look at the data points where a key or "." sign was pressed.

Condition	Mean	SD	SE
DualDialing	0.59 m	0.22	0.06
DualSteering	0.38 m	0.19	0.05

1.c Visualization of lateral deviation over time

Make a plot of how lane deviation changes over time with each keypress, similar to the plot in Figure 1 of Janssen Brumby (2010).

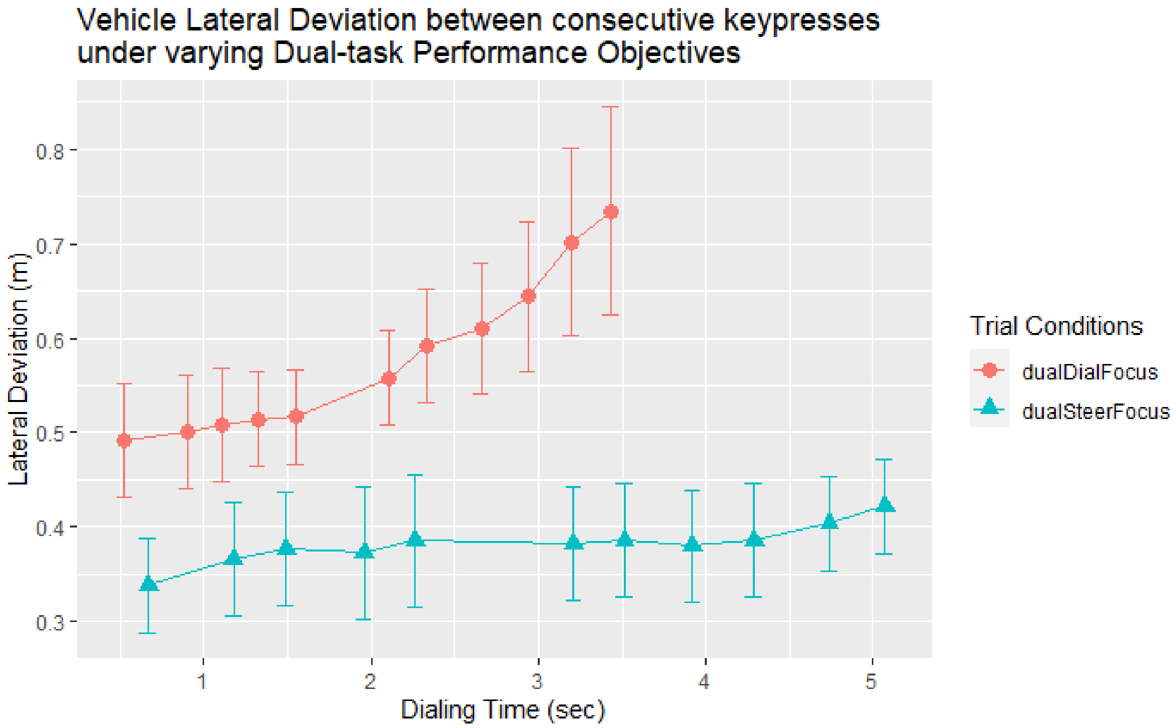


Figure 1: Visualization of lateral deviation over time

1.d Natural breakpoint

For each condition argue whether you think the average participant waited until the natural breakpoint (between 5th and 6th digit) with interleaving dialing for driving.

As observed in the above figure, in the dialing condition, there is a gradual increase in lateral deviation. This is even more visible between the 5th and 6th digit, where the natural breakpoint occurs. Similarly, on the steering condition, there is a steady gradual increase of lateral deviation, but during the natural breakpoint, this remains almost constant. In addition, when the participants have dialing as a priority (main focus), they seem to deviate from steering in a shorter time from the beginning of the trial, and eventually, the natural breakpoint occurs faster. In this case, the participants didn't wait before interleaving dialing for driving, which explains the rapid increase of lateral deviation after the 6th digit. In contrast, when participants give priority to steering, the natural breakpoint occurs later in time, which means that they waited to dial the digits and tried to keep their focus on driving, explaining the constant lateral deviation between 5-6 digits.

Question 2

2.a Visualization of lateral position over time - Human Data

Create a plot in which you show how the lateral position of the car changes over time for the subset of data between 15000 and 18000 milliseconds.

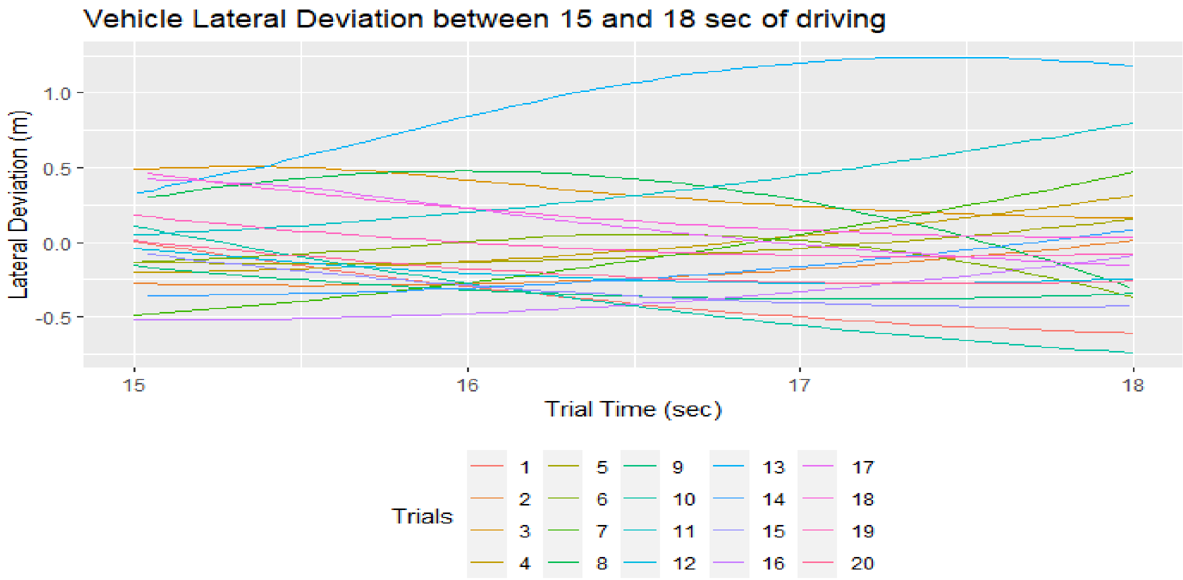


Figure 2: Lateral position over time of human data

2.b Visualization of lateral position over time - Simulated Data of 50 trials

Generate simulated lateral position data for (at least) 20 simulated trials in which you assume that the car starts drifting from a point 0, and samples values from the modeling distribution (M = 0, SD = 0.13) every 50 milliseconds for a period of 3000 milliseconds.

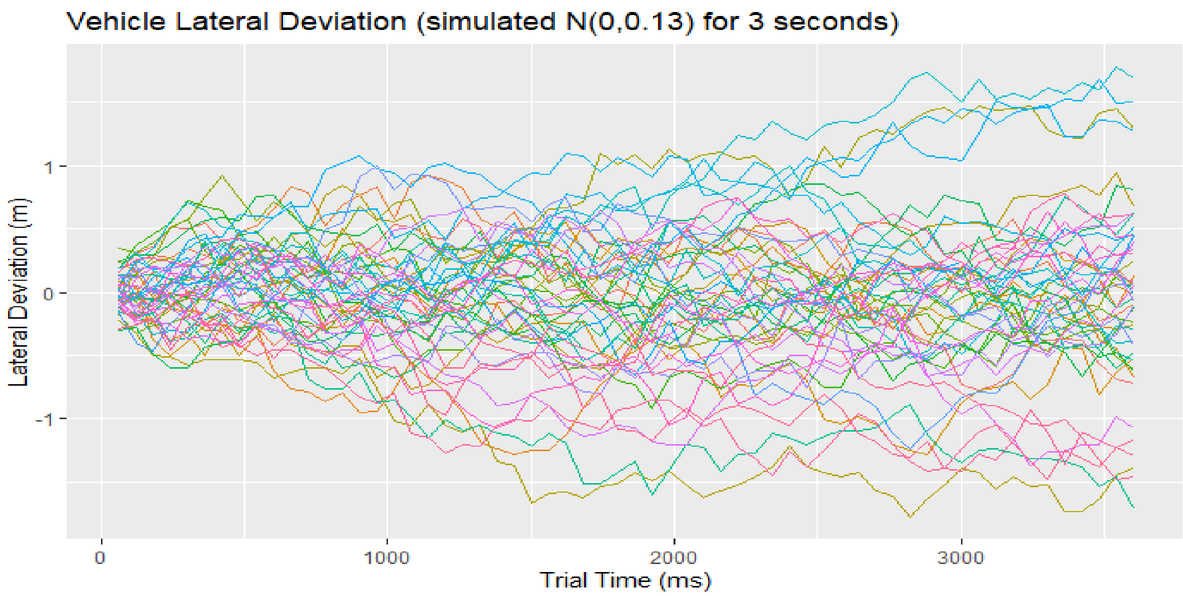


Figure 3: Lateral position over time of simulated data

2.c Histograms

Create two histograms for your report: (i) one histogram that shows the distribution of car positions as measured in the human trial (i.e., a histogram of the data from Question 2A) (ii) one histogram that plots the simulated data (i.e., the data from question 2B).

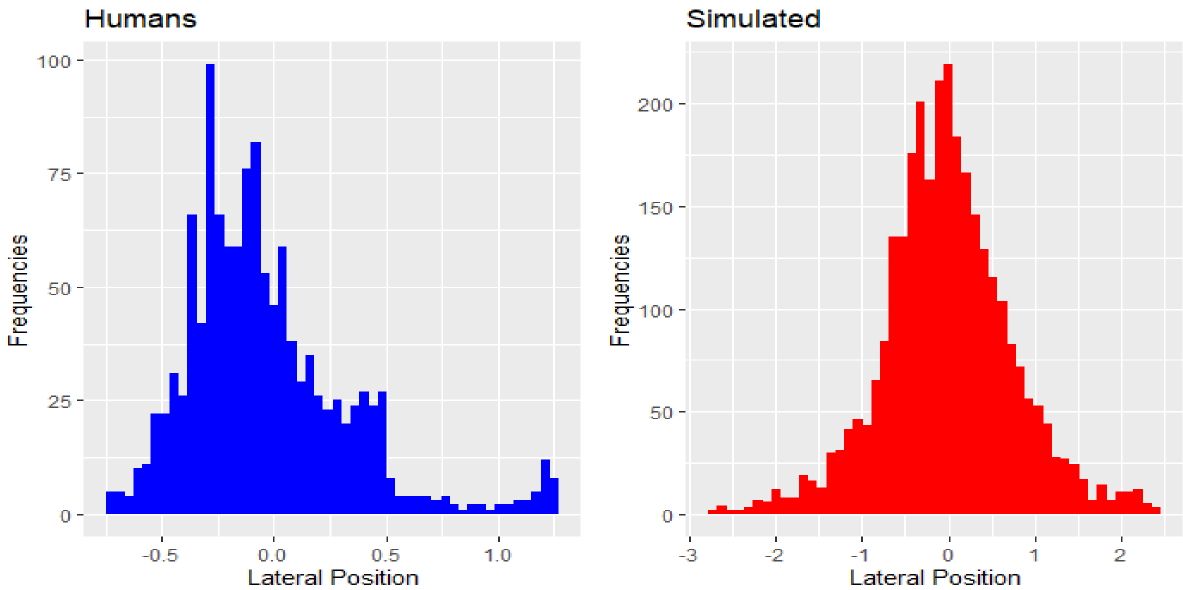


Figure 4: Histogram of human and simulated data

2.d Standard deviations

For the two histograms you plotted in question C you can also calculate the standard deviation of each of the two datasets.

Human data: **SD = 0.36**
Simulated Data: **SD = 0.73**

2.e Simulated data close to human data

2.e.1 Adjusted SD

Try different values in this model (instead of SD = 0.13), and decide which value results in a distribution that is most close to the human data.

We set the standard deviation to **SD = 0.06** as data generated by this Gaussian shows closest resemblence to the human data.

2.e.2 Visualization of the lane position over time of the simulated data with adjusted SD for 50 trials

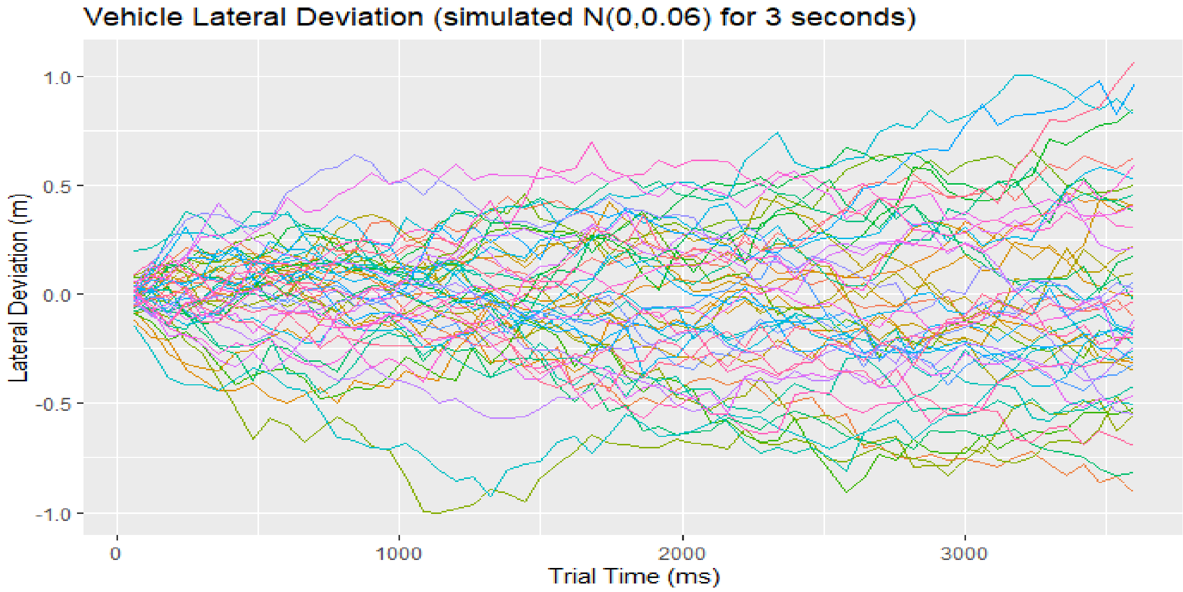


Figure 5: Lateral position over time of simulated data

2.e.3 Histogram of adjusted SD

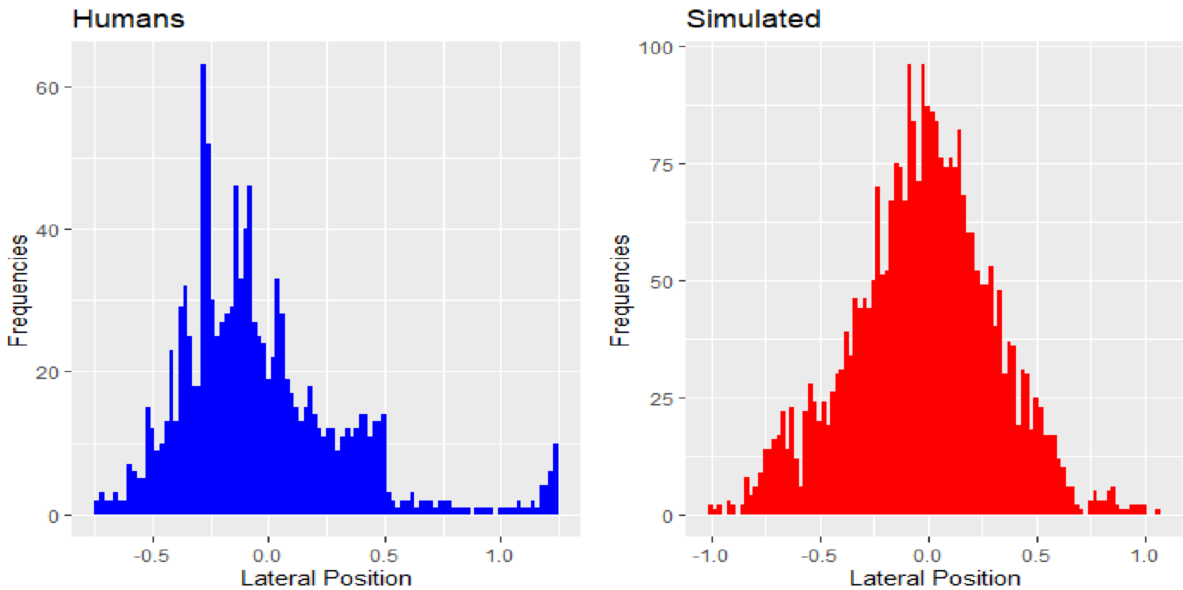


Figure 6: Histogram of new simulated data compared to human data

2.e.4 New SD of resulting distribution

The new SD of the histogram in section 2.e.3 has a **SD = 0.38** which is closer to the human data **SD = 0.36**.

Question 3

3.a Calculated interkeypress intervals

What was the average value of the interkeypress intervals?

The average value of the interkeypress intervals we found is 263.65

3.b Selected interkeypress intervals

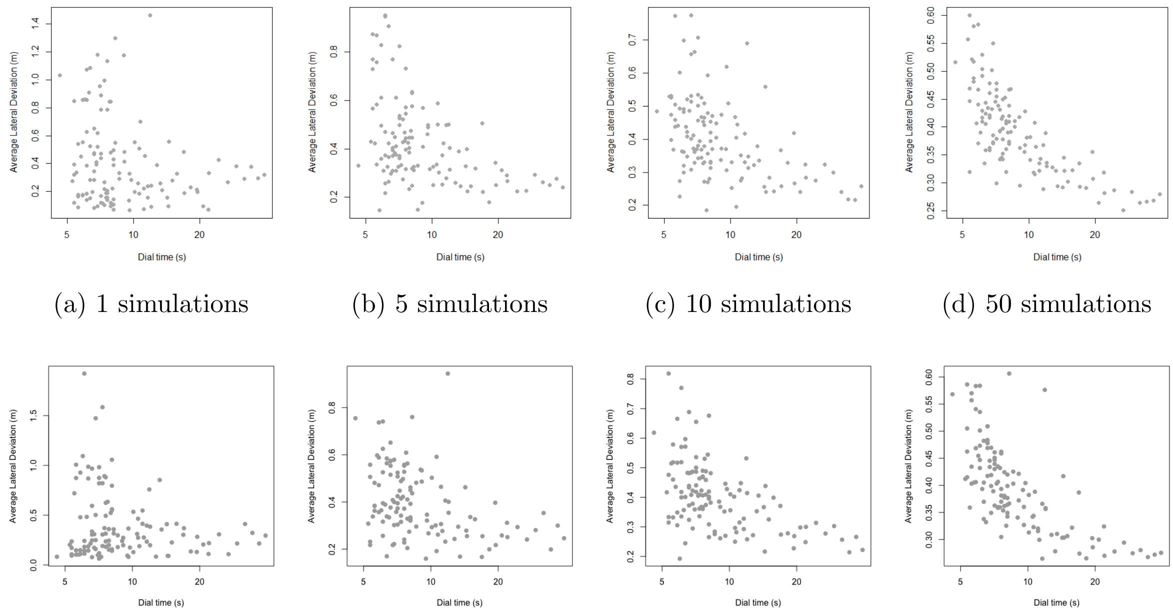
What value did you pick for your model and why?

We pick 264 by rounding up the calculated interkeypress. This is the average across all participants and therefore, a valid estimation of the parameter.

Question 4

4.a Multiple simulation plots

Plot the graphs of the various simulations in your report, for at least 4 different number of simulations.



(a) 1 simulations (b) 5 simulations (c) 10 simulations (d) 50 simulations

(e) 1 simulations (f) 5 simulations (g) 10 simulations (h) 50 simulations

Figure 7: First row was done by Olusanmi Hundogan (Intel Core i7-9750H 2.60 GHz) while the second was computed by Evangelia Giannikou (Intel Core i5-8259U 2.30 GHz).

4.b The right number of simulations

Explain the following: what would be a good number of simulations for your model and why?

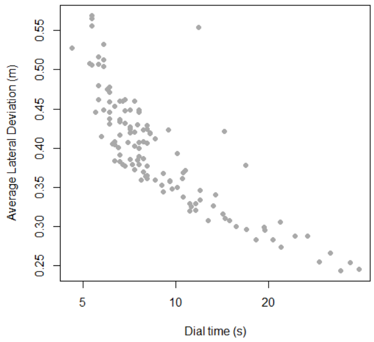


Figure 8: Plot showing the average lateral deviations after 200 simulations.

As shown in Figure 7, the higher the number of simulations, the lower the variation in the distribution of *average lateral deviations* over *dial time*. Meaning, that the standard deviation decreases, which makes sense as the *law of large numbers* in probability theory states that larger numbers of trials tend to approach the expected value of the distribution. We emphasize *expected value*, because this value does not necessarily need to align with the *true value* as the expected value is constrained by modeling assumptions. As a result, increasing the number of trials indefinitely does not necessarily yield values closer to the true distribution, either. In fact, the opposite might be true as the model might generalizability. Another issue with too many simulation runs is the amount of ressources required to run the simulations. Having a higher number might not lead to similarly more knowledge or better results. Hence, it is also always an efficiency trade-off.

We chose 200 simulations as the number of counts as it shows in Figure 8 clearly the general shape of the distribution. This heuristical approach is not optimal. Better approaches would require appropriate measures for the quality of fit.