Updated Nov 18, 2013

Questions and Answers for 2014 TRB contest

This document will be updated on a regular basis given e-mails we have received. We will do our best to answer your question, but may not have a response to all your questions immediately, so please have patience. Also, when you send us a question, PLEASE include "TRB Data Contest" in the subject header so your e-mail can be quickly located.

Q1. Based on my understanding, the contest requires us to develop a model based on the given data. But it does not quite clear to me what is the objective of the question, i.e., what is the response variable of the model (or the output of the model).
A1. The objective of the data contest to "**identify differences in driver behavior in a dilemma zone while distracted."** This information is now included in point #1 in the TRB Data Contest file.

Q2. Do you want us to model all the simulator's outputs (distance, velocity, etc.) or only a specific one?
A2. It is up to you. We will not be telling you what to model.

Q3. Can you please confirm that the experiment was conducted in a simulator which means that the drivers were NOT driving in real roads?
A3. Yes, the experiment was done in a simulator. A picture and description of the simulator can be found at http://www.nads-sc.uiowa.edu/sim_nads1.php

Q4. For Column B: what does presentation order mean?
A4. This person also had questions regarding Columns C, E, F, and L. Please see the data definition file for a more detailed explanation of each field (including the coding). If the explanation is still not sufficient, please e-mail me again.

Q5. Does HH-T1BOI mean– hand held phone through all the driving, then baseline(?), outgoing call, then incoming call? When do these actions occur and how frequent? Do they occur right before the lights? Did they hold the phone all the time?
A5. Please see the data definition file for an explanation of each field (including the coding), and the data contest file, which also includes a description of the study.

Q6. What does baseline mean? Does it mean no phone?
A6. Yes, it means no phone.

Q7. Posting a video in the time frame the data was collected will be helpful too.
A7. We will not be posting a video. However, we have provided a more detailed description of the events in the TRB data contest file.

Q8. I understand how to interpret the "Run Name" variable for all but 26 cases. These are as follows: Case 150-155; "R21", Case 200-203; "R22", Case 389-390; "R13", Case 908-913; "R12", Case 942-943; "R23", Case 1070-1073; "R12"

In all other instances, these characters in the case sequence refer to one of three "drives" (D1, D2, or D3) or a practice run (FAMILIAR). Could you explain what the "R" code stands for, as well as the subsequent two digits?
A8. Please see the latest version of the data definition file; there is a clearer explanation for the difference between "D" and "R".

Q9. "Dist From Stop Line (feet)" entries: Namely is there a typo in the instructions which at present says "-9999" if driver didn't stop, but in the data set when the driver does not stop (e.g. First Stop Frame = -1), the distance from stop line is +9999.
A9. You are correct, this should be +9999 and it has been corrected in the Data Dictionary as well (updated on Aug 23, 2013).

Q10. The three interface alternatives are: handheld, hands-free, and handset. Should the third option be headset? It is unclear what the difference would be between a handheld phone a handset option.
A10. You are correct, "HS" should be HEADSET and this is now revised in the Data Dictionary as well (updated on Aug 23, 2013).

Q11. Could you please explain the meaning of accel pedal changed 10%? 10% of what? From what time to what time or during what time inverval? Is my assumption that this change occured during the 4 second interval between the two events GY and YR correct?
A11. The accelerator pedal value changes 10% from the time the light changes to yellow. The initial value is obtained by averaging the value over the preceding 4 seconds and is meant to indicate the driver's nominal pedal usage during normal driving. This has now been updated in the Data Dictionary (updated on Sep 10, 2013).

Q12. I assume that the "N" means that the subject's Acceleration Pedal did not change more than 10%. Is that correct?
A12. Yes, that is correct.

Q13. If the data in Column F is "N", how are we to interpret the Min and Max Accel? Are they the minimum and maximum acceleration for that entire run?
A13. The minimum acceleration (col J) and maximum acceleration (col K) are defined in the data dictionary. If there was no large accelerator pedal change, then these values are the minimum and maximum acceleration for that specific run.

Q14. What age brackets are represented by the categories young, middle-aged, and older?
A14. This information was included in the document titled, "TRB Data Contest" (Item #3) and still is. But we have now included it in the "Data Definition for TRB Contest" as well.

Q15. What is the difference between a handheld and a handset phone?
A15. This is now better clarified in the "Data Definition for TRB Contest."

Q16a. In many instances (the 4th run of Run_001YM_D1HH_T1BOI, for instance) the green to yellow, yellow to red, and red to green (columns C, D, E) values all occur within

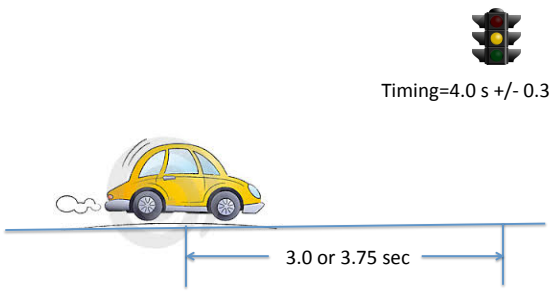
a few frames of each other, essentially indicating that they happened simultaneously. Since this seems to violate the study description of a 4-second yellow followed by a 5-second green, can these observations be considered erroneous?
Q16b. For more than half of the observations, it appears that the yellow and/or interval is 0.3 seconds or less. Is this correct? The first such example is the fourth run for the first subject, which indicates only 8 frames to have elapsed from green to yellow and from yellow to red.
A16. We have received many versions of this e-mail, and we apologize, but there was an error in the coding. It has been corrected in an updated version of the datafile (now dated Sept 10, 2013). The three variables that were impacted by the change was:

Ov Vel at Green To Yellow (Col L)
Ov Dist at Green to Yellow (Col M)
Vel at Yellow to Red (Col P)

[updated Q&A as of Oct 13, 2013]
Q17. The "Data contest guideline and Data description (updated on Aug 20, 2013)" document states that ?traffic signal changes to yellow at one of two pre-determined timings (3.00 or 3.75 seconds)?, what is the reference for these timing changes, are they the time headway from the stop line or something else?
A17. I apologize beforehand as I may have misinterpreted your question, but the timing is computed when the virtual FRONT of the vehicle reaches a specific point before the intersection. So, I will tend to agree that it is 3.0 (or 3.75) seconds from the front of the vehicle to the intersection, given the vehicle speed.

Q18. Could you please explain what does R12 stand for?
A18. This is described in the Data Definition as has been there since Aug 24, 2013 (see the definition of Run Name [col B].

Q19a. The same document also states that "The light remains yellow for 4.00 seconds and then the light transitions to red for another 5.00 seconds", but according to calculation results of the Sep 10th updated data sets, there are various yellow lengths (ranges between 0 and 4.38) and red lengths (ranges between 0 and 5), how should we perceive these different yellow and red times?
Q19b. The duration of yellow light is varying from 0 to more than 4 sec, why would that happen? Isn't it constant, say fixed 3 or 3.75 seconds, for each situation?
A19. In response to Q19a, the description posted came from the original documentation associated with this study (conducted in 2004). In response to Q19a and Q19b, the yellow lengths (based on the Sept 10th data file) should actually not vary from 0 to 4.38, but is theoretically suppose to be held constant at 4 seconds after it is triggered at 3 or 3.75 seconds away. That said, the yellow phase is not constant given variations in the simulator algorithm. Based on the update from Sept 10, 2013, the values do range from vary somewhat from 2.78 to 4.38 seconds. The next figure [***updated on Nov 11, 2013***] is my attempt to explain what is going on.



[**updated on Nov 11, 2013**]. There are some wide ranges in the red phase. Once the driver (participant) passes an intersection, the simulation no longer records the signal state of any previous scenes.

Q20a. As indicated in the data definition file "There should be 6 rows of data for each run condition as there were two signal changes per condition (2 in the baseline, 2 in outgoing, and 2 in incoming)". I just want to make sure if I understand it right. For example, "Run_001YM_D1HH_T1BOI" had six rows. Are the first two rows for baseline condition, following by two rows for outgoing call and the last two rows for incoming call?
Q20b. Each drive number for each subject ideally contains 6 observations: 2 for baseline, 2 for outgoing call, and 2 for incoming call. The treatment number variable indicates the order that these conditions are presented, but it does not say how the pairs of identical conditions are ordered. For example, for the treatment "BOI", it is unclear whether the conditions are presented as "B O I B O I" or "B B O O I I" for the 6 observations per drive. Which one of these orderings is the correct interpretation?
Q20c. We have a question related to the order of drives. For example, there are 6 rows for each run condition/visit as there were two signal changes per condition (2 in the baseline, 2 in outgoing, and 2 in incoming). Let's say the order is 'BOI'. Does that mean first 2 rows (of 6 rows) are baseline, the next two are outgoing, and the last two are incoming? Or, does that mean the first three rows are BOI and then the next three rows are BOI?
A20. We received many versions of this question. Using the example from Q20b, the correct ordering is B B O O I I in your example scenario. So the answer to Q20a is, yes – that is correct.

That said, we can see why this is confusing, particularly if the participant did not complete all six runs. There is a new column now called "Event ID" that is a number between 300 to 317. This can be used to indicate run (or row number) is B, O, and I.

Event IDs from 300 to 305 would be events in the 1st segment; 306 to 311 would be events in the 2nd segment, and 312 to 317 would be events in the 3rd segment. So if the order of runs were BOI, then events between 300 and 305 would include the baseline case, between 306 to 311 would include the outgoing phone case, and between 312 to 317 would be the incoming case. For simplicity, this code is added as a last column and is now updated in the datafile and data description.

Q21. The study description states that the length of red for each drive is supposed to be 5

seconds. However, there are a number of observations with negative values of red, and it appears that all of these correspond to drivers who did not stop at the light. However, some of these "non-stopping" drivers also have positive values for the red duration, so I'm not sure if this indicates some sort of error with the "yellow to red" timeframe coding or not. Could you clarify why some of these red lengths are negative?
A21. These represent incomplete data. That is, the rows with negative values should not be included in your analysis.

Q22a. Distance from stop line (Col I), as you said "—" is beyond the stop line, does it mean the vehicle stopped in the intersection?
A22a. No, if the value is negative, it means the vehicle stopped beyond (or after passing the stop line).

Q22b. If the value is not large, I may guess the vehicle stopped a little far beyond the line (Maybe over the crosswalk), which would be possible. However, there are values as large as -40 ft, which would be fully entered the intersection or more. If it is, why they need or should stop? I don't believe it is a good operation for safety consideration.
A22b. I will admit that I am not sure what you are asking. However, it is correct to say that some drivers are not as safe as others.

Q23. You mention that the ambient traffic remain similar for all the cases. Is it safe to assume that there is no other car in front of the driver when light changes to yellow?
A23. There could be a lead vehicle present but during the yellow light events, the lead vehicle is programmed to be so far in front that there is no potential for conflict.

Q24. There appears to be row duplication for Subject 104. That is, there are entire rows for this Subject that is exactly the same.
A24. We apologize for that. This has now been corrected in the datafile dated Oct 13, 2013.

Q25. For the variable "First stop frame", does "first stops" indicate when the velocity first equals 0, or when the driver first applies the brake?
A25. The first stop frame indicates that frame number when the vehicle's velocity is at zero. The data dictionary has been revised to include this clarification.

Q26. I am writing to clarify if the conditions for the 6 rows with 'FAMILIAR' in the 'Run Name' column are under baseline conditions.
A26. This is described in the data definition file in the description of Column B (Run Name).