This table is a result of some psycholinguistic experiment. The experimental settings are as follows. Researchers select random person (respondent) and ask them a question (for example, "describe what do you think about technologies"). Then the answer is recorded, splitted into sentences and duration of each sentence is measured and recorded in the column sentence duration (in seconds). If there are several sentences in the answer, several rows are created. Then second random respondent is selected independently of the choice of the first respondent, the same question is asked and answer is recorded in the same way. Then process continues with the next repondent, and so on.

Now consider column sentence duration, denote its values (in the same order as they are presented in the table) by $x1,\ldots,xn$.

There are several random factors here. First of all, we select respondents randomly, and each respondent has their unique speech preferences, i.e. some people speak faster than the other, etc. Then, the choice of answer to a particular questions is unpredictable to some extent and can has some random component: even we ask the same question to the same person, we should get different answers. So values $x1,\ldots,xn$ can be considered as realizations of some random variables.

The question is: can we say that $(x1,\ldots,xn)$ can be considered as an i.i.d. sample from some random variable? I.e. can we assume that there exists some random variable X and we can treat values $(x1,\ldots,xn)$ as independent realizations of this random variable?

No, $(x_1, \ldots, x_n)$ is not an i.i.d sample. To prove that, let us address the setup of the experiment: the breakdown of the human speech in sentences violates the i.i.d. condition of independence due to the different speech habits. There exists a dependence between spech preferences and consequent number of sentences. Essentially, the number of seconds is directly influenced by number of sentences to answer the question. Given that the sample size is big enough, there will be an inverse relation between the number of sentences and seconds spoken. So, the feature " `sentence duration` " is expected to be negatively correlated to the number of sentences per respondent. ■