

How people perceive expressions of probability

Matej Miočič¹

Abstract

In this report, our goal was to provide insights on how people perceive expressions of probability based on a survey. We try to understand how and if people differ in their perception based on their demographics. Survey was conducted on Slovenian and English speaking participants separately.

Keywords

Expressions of probability, Survey, Perceived probability

¹mm9520@fri.uni-lj.si, 63180206

Introduction

In Project 2 DEMO, we were tasked to explore and provide insights on conducted survey. Survey was conducted on Slovenian and English speaking participants separately. In the report we are working with English expressions, but it is worth mentioning Slovenian participants were prompted with the Slovenian expression. Each respondent had to assign a probability (on a scale from 0% to 100%) to each of the 15 expressions. Each response has also included population weight and demographics such as sex, year of birth and whether or not the respondent has completed a higher education program. Dataset contains **1046** responses with **20 attributes** mentioned before. All gathered insights are provided graphically and are commented on.

Figures

Perception of expressions of probability

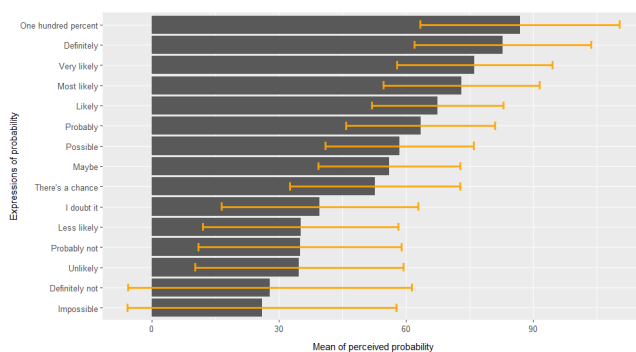


Figure 1. Comparison of mean of perceived probability and their standard deviation show us how people rank expressions of probability and how much they differ from one another.

In Figure 1 we can notice high standard deviation which indicates that the values are spread out over a wider range. That means people in fact perceive these expressions differently. Standard deviations of each expression are quite similar which means every word is perceived differently by a similar amount. This could indicate that dataset contains non-serious responses.

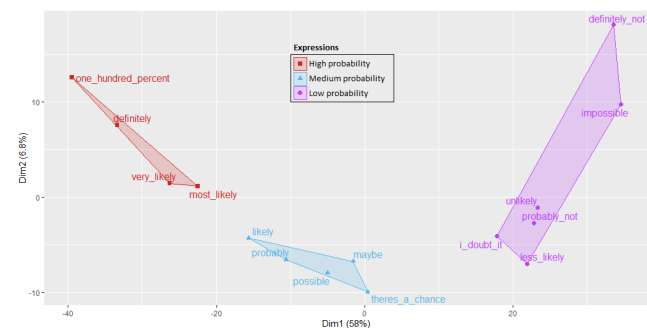
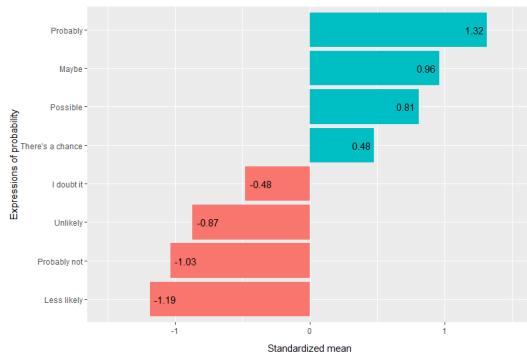


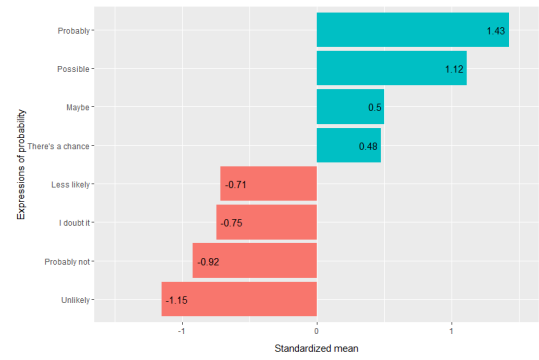
Figure 2. Cluster of expressions show us how people group 15 expressions into 3 groups – **high, medium** and **low** probability expressions.

We decided to group expressions into 3 clusters shown in Figure 2 for easier further analysis. We have also shown that expressions group into objectively "most sensible" clusters. For grouping we needed to remove 6 participants from the dataset, since variance of their ranking are 0.

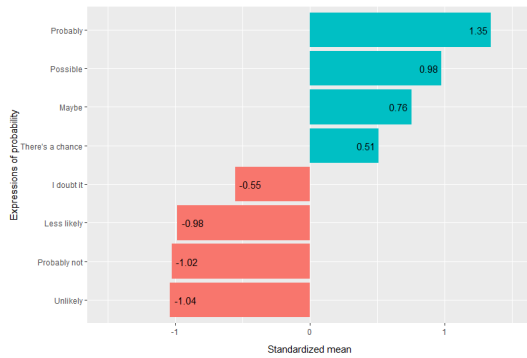
Next we show similarities/differences between responses based on demographics in Figure 3. We decided to show only 8 middle ranked expressions based on their standardized means. Almost all graphs seem indistinguishable, which means participants' rankings of expressions are close. Main difference can be seen between Slovene and English speaking samples. Most probable cause is translation of words into Slovene for Slovene participants.



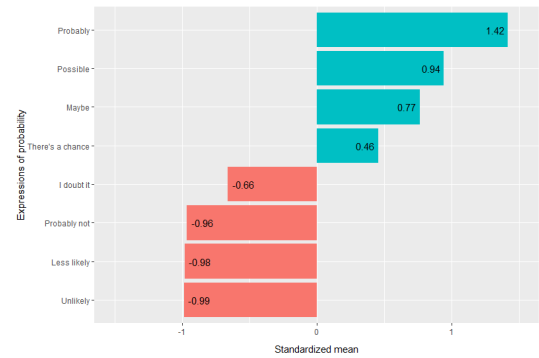
(a) Means of responses of **Slovene** speaking participants.



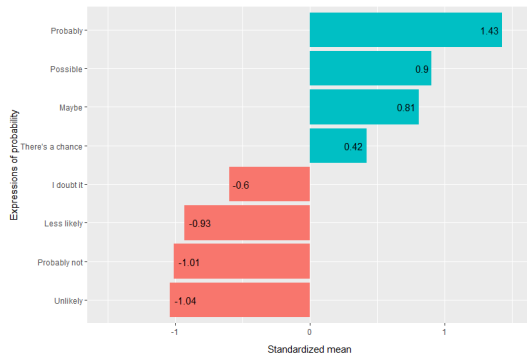
(b) Means of responses of **English** speaking participants.



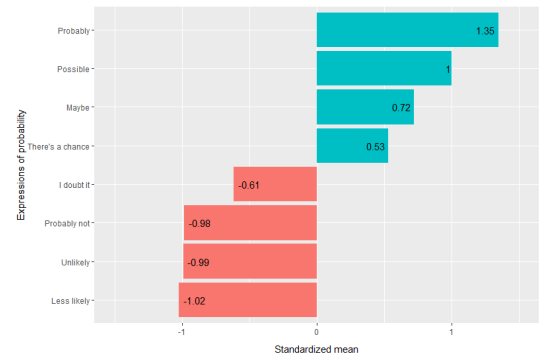
(c) Means of responses of **female** participants.



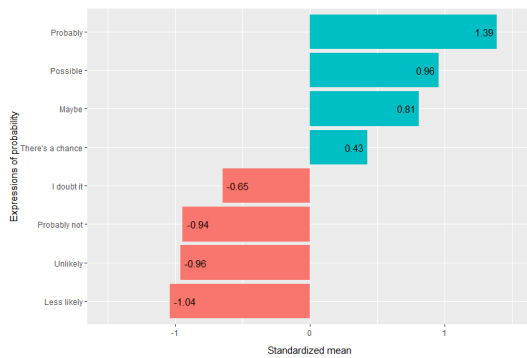
(d) Means of responses of **male** participants.



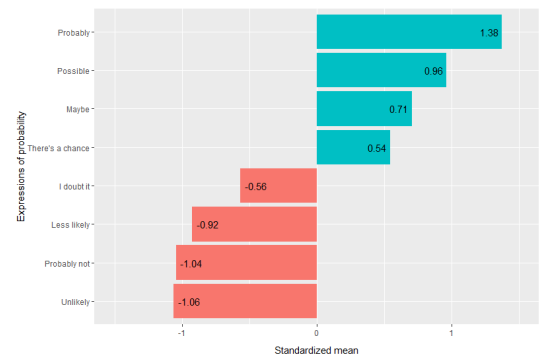
(e) Means of responses of participants with **completed** higher education program.



(f) Means of responses of participants with **uncompleted** higher education program.



(g) Means of responses of participants born **before** the year 1982.



(h) Means of responses of participants born **after** the year 1982.

Figure 3. Comparison of standardized means of 8 middle expressions based on different demographics.

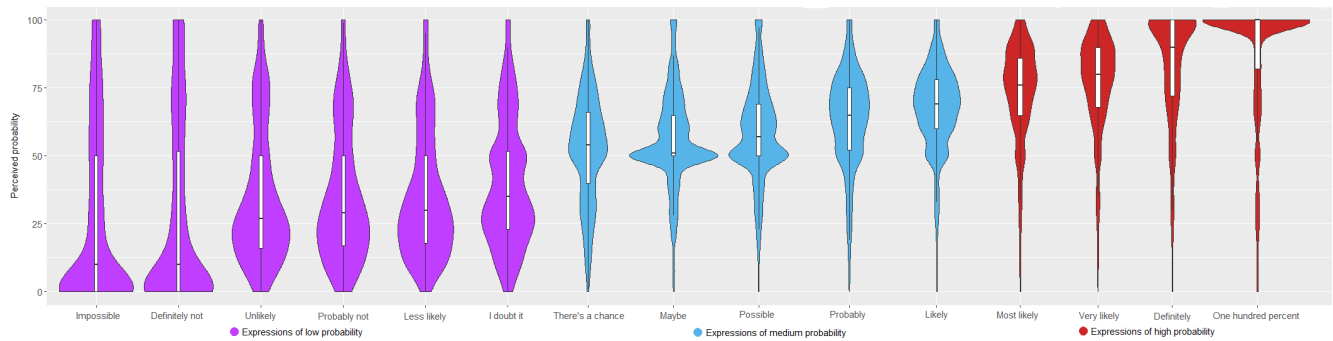


Figure 4. Violin plots of probabilities of expressions divided into obtained clusters.

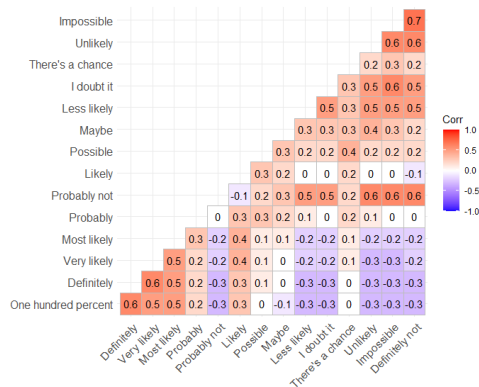


Figure 5. Correlation between each expression of probability based on participants' rankings.

We decided to show violin plots based on groups obtained in Figure 2. In Figure 4 we show distribution of each expression. We can see that low probability expressions have widest

tails, which means they are the most distributed out of three groups.

Finally we show correlation between each expression of probability in Figure 5. We can see that expressions are most correlated in the extremes. For instance 2 expressions ("Definitely not" and "Impossible") with the lowest mean of participants' ranking are the most correlated. We can observe the same with 2 expressions ("One hundred percent" and "Definitely") with the highest mean of participants' ranking. This means most participants ranked correlated expressions alike.

Conclusion

In this report, we analysed how people perceive expressions of probability. We found out that people's perception differs from one another, but when comparing samples based on demographics, we found out that perceptions are more or less the same.