IIT GANDHINAGAR
MA202

Project on Choice Of Prediction

Supervised by: Krishna Miyapuram

Project By:

Sr. No.	Roll Number	Name
1	21110210	Vinit Singh
2	21110165	Pratham Sagar
3	21110112	Lakshya Mehta
4	21110005	Abhinab Mondal
5	21110018	Akshay Mishra

ACKNOWLEDGEMENT

We would like to convey our sincere gratitude to Professor Krishna Miyapuram for his guidance, insight and mentorship during MA202 project on 4th choice prediction competition. His insightful feedback and constructive criticism have been instrumental in shaping the direction and quality of our project. Moreover, his guidance has not only helped us to achieve our project objectives but also enhanced our overall understanding of the subject matter.

What was the 4th Choice Prediction Competition?

The 4th Choice Prediction Competition, held in 2015 by The Hebrew University of Jerusalem, was a contest that challenged participants to create accurate predictive models of human decision-making in situations involving choices between two options. The competition focused on a particular type of decision-making in which participants had to predict which of two options (A or B) would be chosen by a person in a given situation.
List pattern of the experiment
A trial is defined as the set of the 30 questions which are answered by the participants randomly. Each participant did 25 trials in which 5 were done without feedback and 20 with feedback. So all in all the participant made a total of 25*30 = 750 choices.
In the context of this competition, option B was considered "risky" because it offered a higher potential payoff, but also a higher probability of receiving nothing or losing more. The higher probability of receiving nothing with option B was meant to be posed as a "risk" by the experimenters. However, whether or not option B is actually riskier depends on the individual's perception of risk and their decision-making preferences.

Let us explore the selection of the lottery option (option B) on the basis of different demographics:

We will take a normalizing factor which is equal to the number of trials (i.e 5 for without feedback and 20 for with feedback)

1. **Gender:**
    a. **Without Feedback:**

$$\text{Proportion of men taking option B in all trials (in } = \frac{100 * \text{total option B taken}}{\text{normalizing factor * no of men}}$$

The proportion of men taking option B in all trials was found to be 1.99259%, while the proportion of women taking option B was found to be 2.06098%.

    b. **With Feedback:**

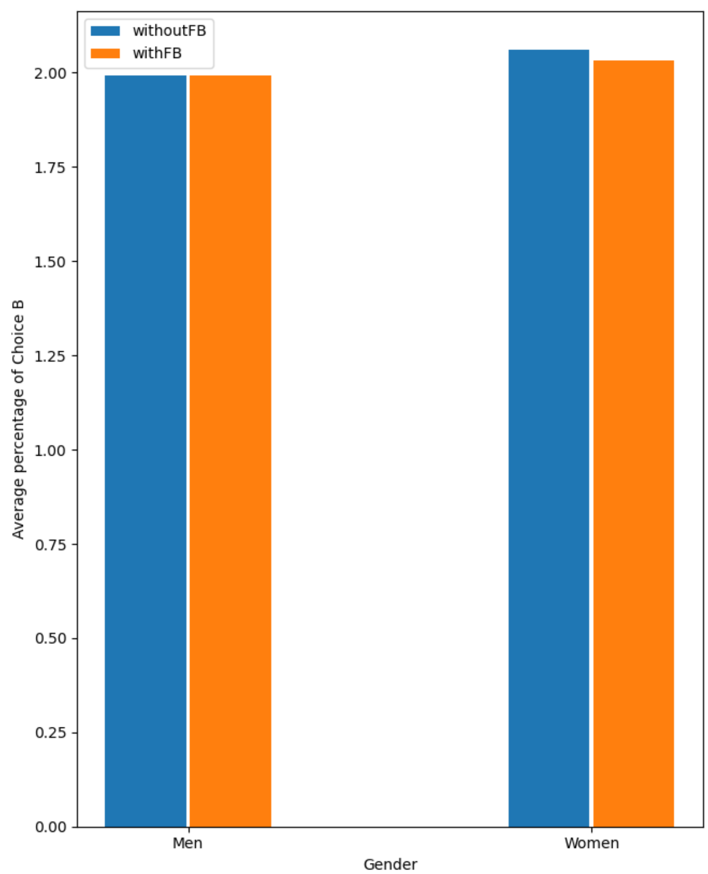
$$\text{Proportion of men taking option B in all trials (in} = \frac{100 * \text{total option B taken by men}}{\text{number of men * normalizing factor}}$$

$$\text{Proportion of men taking option B in all trials with feedback} = 1.99185$$

$$\text{Proportion of women taking option B} = 2.0320$$

$$\delta \text{ in proportion of men} = 0.4979629 - 0.4981482 = -0.0001853$$

$$\delta \text{ in proportion of women} = 0.508 - 0.5152463 = -0.0072463$$



2. **Age:**

We will divide the participants into 3 age groups -

i. Group 1 - 23 and below

ii. Group 2 - 24-28

iii. Group 3 - 29 and above

a. **Without Feedback:**

$$\text{Proportion of choosing option B in Group 1} = 2.03379$$

$$\text{Proportion of choosing option B in Group 2} = 2.01821$$

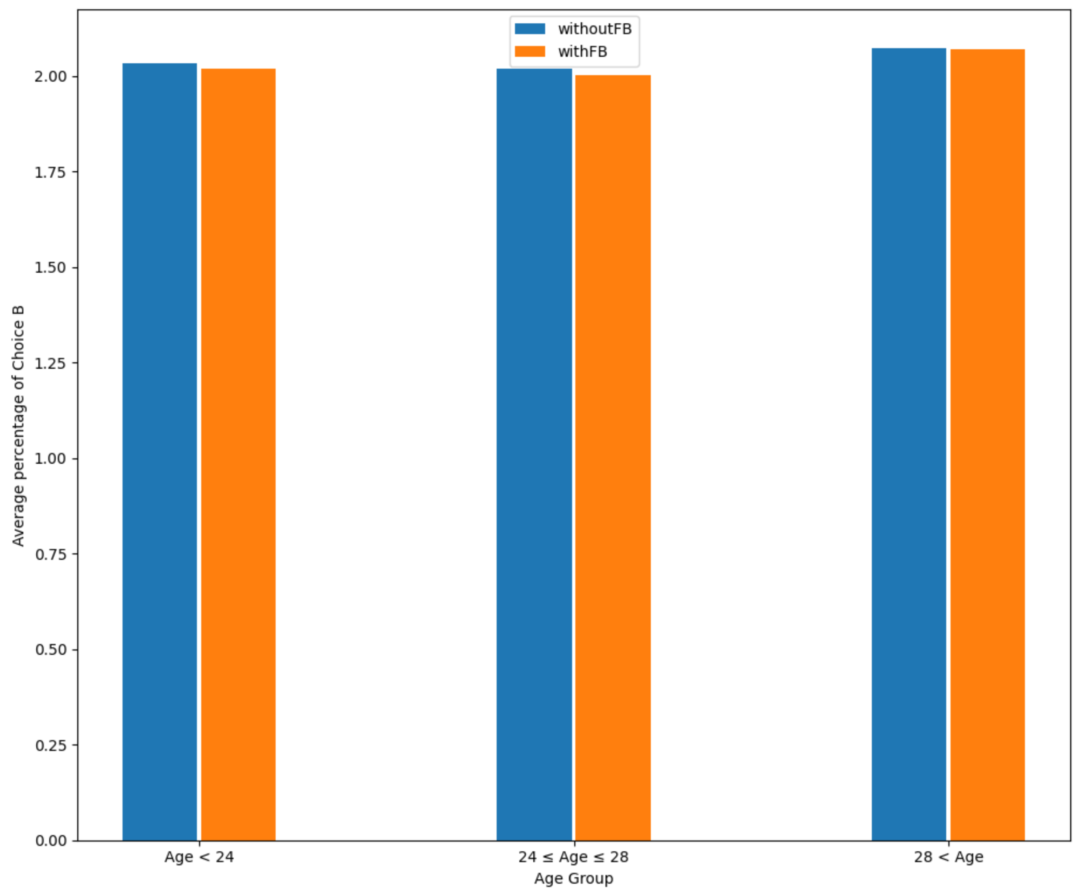
$$\text{Proportion of choosing option B in Group 3} = 2.07246$$

b. **With Feedback:**

$$\text{Proportion of choosing option B in Group 1} = 2.01856$$

$$\text{Proportion of choosing option B in Group 2} = 2.00093$$

$$\text{Proportion of choosing option B in Group 3} = 2.06913$$



3. **Location:**

The experiment to generate the dataset was performed in two universities -

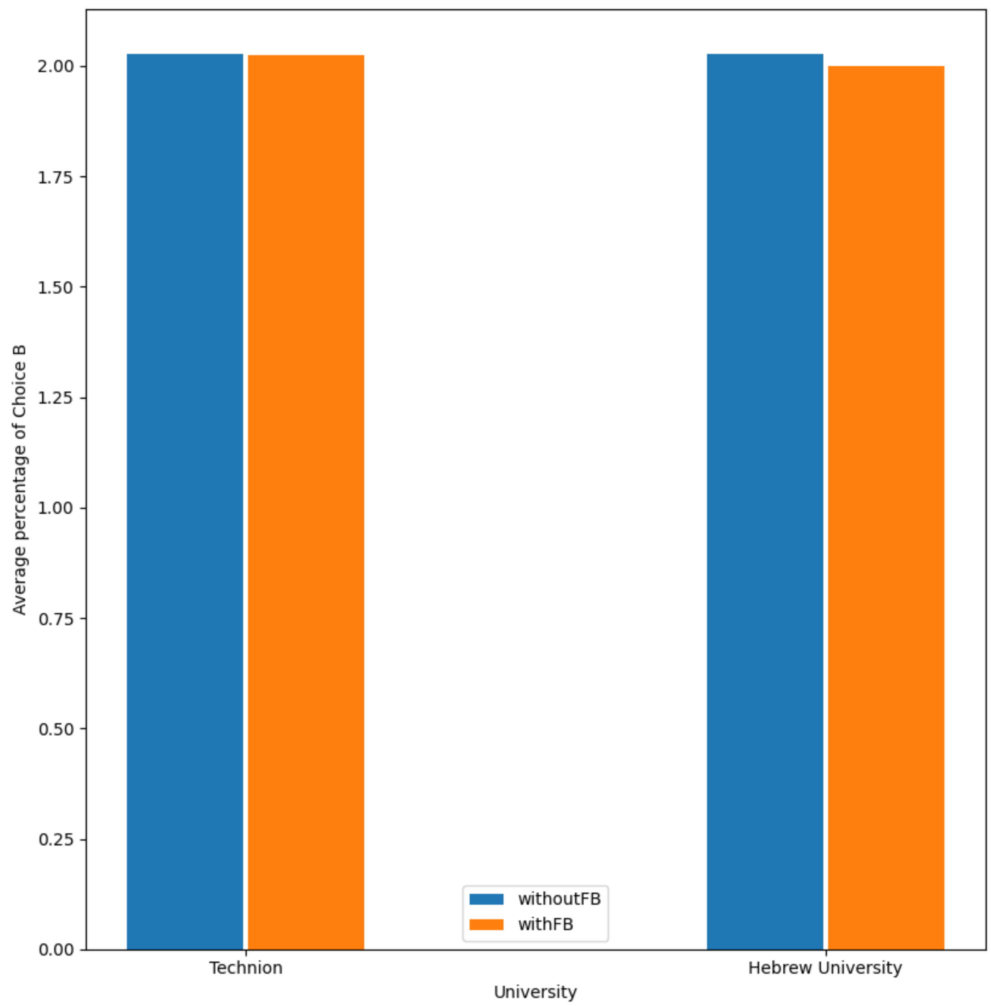
Institute 1 - Technion - Israel Institute of Technology
Institute 2 - Hebrew University of Jerusalem, Rehovot

a. **Without Feedback:**

Proportion of choosing option B in institute 1 = 2.02737%
Proportion of choosing option B in institute 2 = 2.02836%

b. **With Feedback:**

Proportion of choosing option B in institute 1 = 2.02492%
Proportion of choosing option B in institute 2 = 1.99984%



Although, we can see that there is not much dependence between parameters like age and gender on the choice between A and B, but let us more formally perform a chi-squared test to establish this hypothesis.
First, let us form our null and alternate hypothesis. The null hypothesis in this case would be that there is no significant difference between the observed and expected frequencies of choices (A or B) across different age groups and gender. The alternate hypothesis would be that there is a significant difference between the observed and expected frequencies of choices (A or B) across different age groups and gender.

Next, we calculate the expected frequencies for each category based on the marginal distributions of age group and sex:

$$E_{i,j} = \frac{R_i \cdot C_j}{n}$$

where E_{ij} is the expected frequency for category ij, r_{ik} is the frequency count for age group i and sex k, c_{kj} is the frequency count for choice j and sex k, and N is the total sample size.

Using python, we calculated the expected frequencies as to be found in the following contingency table:

		Choice	
Gender	**Age Group**	**A**	**B**
Female	23 and below	13394.5829	13605.4170
Female	24 to 28	65856.6995	66893.3004
Female	29 and above	6325.2197	6424.7802
Male	23 and below	9673.8654	9826.1345
Male	24 to 28	49485.5426	50264.4573
Male	29 and above	21208.0896	21541.9103

Also, the observed frequency table is as follows:

		Choice	
Gender	**Age Group**	**A**	**B**
Female	23 and below	13602	13398

Female	24 to 28	65098	67652
Female	29 and above	5920	6830
Male	23 and below	9588	9912
Male	24 to 28	50914	48836
Male	29 and above	20822	21928

Note that these values are of all trials that every participant has done. We did not normalize it by any factor because as we will see ahead that the p-value will come out to be much less than the tolerance level anyways.

We then calculate the test statistic:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

In this case, we have 6 categories (2 sexes x 3 age groups) and 1 independent variable (choice of A or B), so the number of degrees of freedom is (6-1) = 5.

Chi-squared value: 172.5276

$$p = P(\chi^2 \geq \chi^2_{obs})$$

$$p = 1 - F_{\chi^2}(\chi^2_{obs}|df)$$

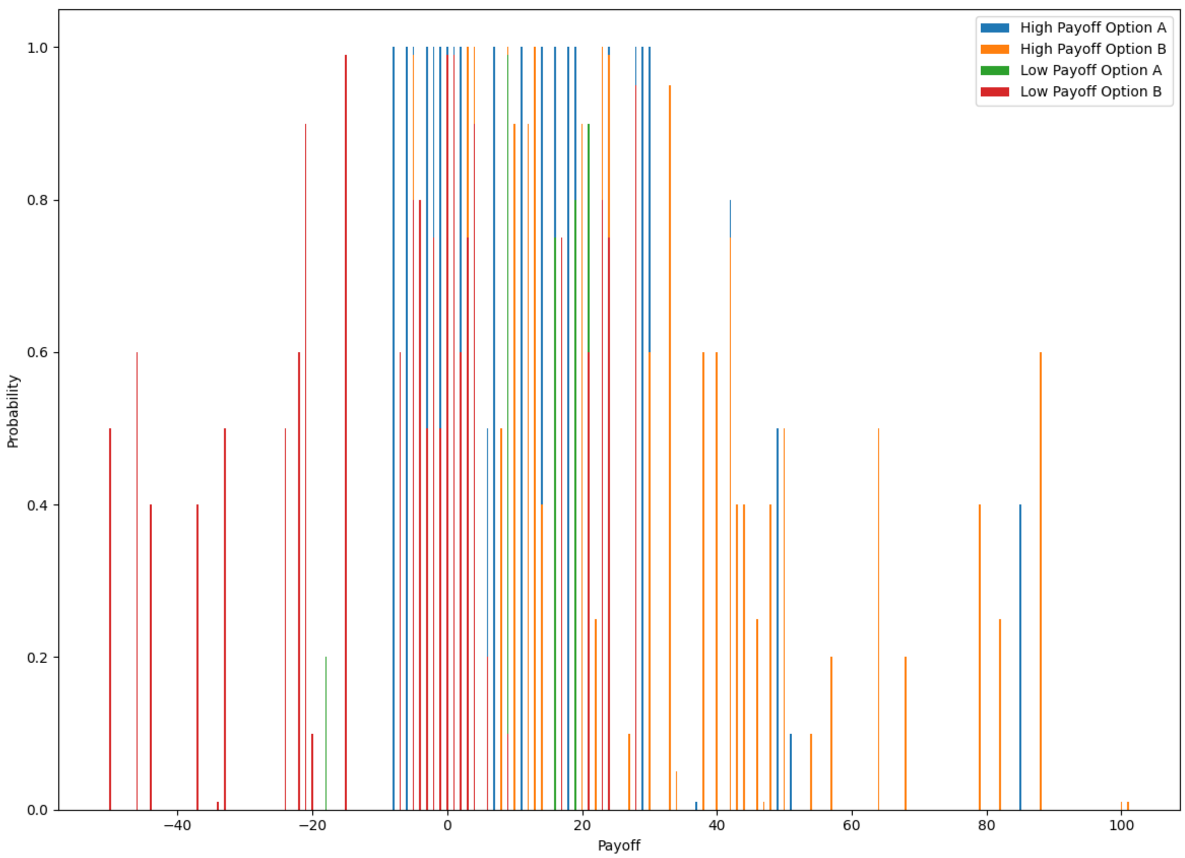
We can calculate the p value using software and tools to get:

P-value: 2.1073e-35
e-35 is scientific notation for 10^-35

Which is an extremely-extremely trivial quantity as compared to the standard tolerance level of 95%. So we can say that we fail to reject the null hypothesis.

Probability v/s Payoff Distribution:

Just as an exercise, we can plot the probability v/s payoff distribution to get an idea of the probability of getting a higher payoff in the experiment and what the participants are risking in order to do so.



CONCLUSION :

We can observe that there is a somewhat normal distribution between the payoff and probability.

We can conclude that the choices do not depend on anything but the individual and their perception of risk and their risk taking abilities. We can also observe from the data that not much of the choices change across trials but most of them change with the framing and manipulation of the statement which was given to the participants which was the root of emergence of some of the paradoxes that were captured in this experimental paradigm.