# Introduction:

I wanted to see whether student standing had a statistically significant effect on participants’ answers to any of the questions.

# Data Cleaning:

The age information in the 2024 survey had the data stored as floats while the age information in the Max and Fardina survey was stored as an object. All of these columns also had Nan values. When plotting the age distribution, I first checked the percentage of participants that actually reported their age, and have visualized the data as a pie chart below. Since the vast majority of participants reported their age, I chose to exclude the NaN values when displaying data for the age distribution bar plot.

When performing my hypothesis test for independence between student standing and participant answers, I found that including NaN values skewed the test to make more questions appear statistically significant. This was because, since such a small number of people did not report their student standing, any variation in their answers were statistically significant. As a result, I chose to exclude entries where students either didn’t report their student standing or didn’t answer the question.

# Background:

There were 3 CSV fails that recorded the survey responses of students from 3 different classes. The wording of the questions in each survey was slightly altered and an extra question was added to one of the surveys, but the surveys were largely similar. They consisted of 21 questions (22 for the survey with the extra question) and each had a varying number of responses ranging from 80 to 132 responses. The first 7 questions (8 for the survey with the extra question) asked about the background of the surveyee (things like age, religious alignment, gender, etc.). The other 14 questions presented a scenario involving a person performing some controversial action and the participants were asked to share their opinions on whether the person in the scenario was justified in taking the actions they did. From hereon, the 2 surveys consisting of 21 questions will be referred to as “2024 survey” and “Max survey” respectively. The survey consisting of 22 questions will be referred to here on as “Fardina survey.”

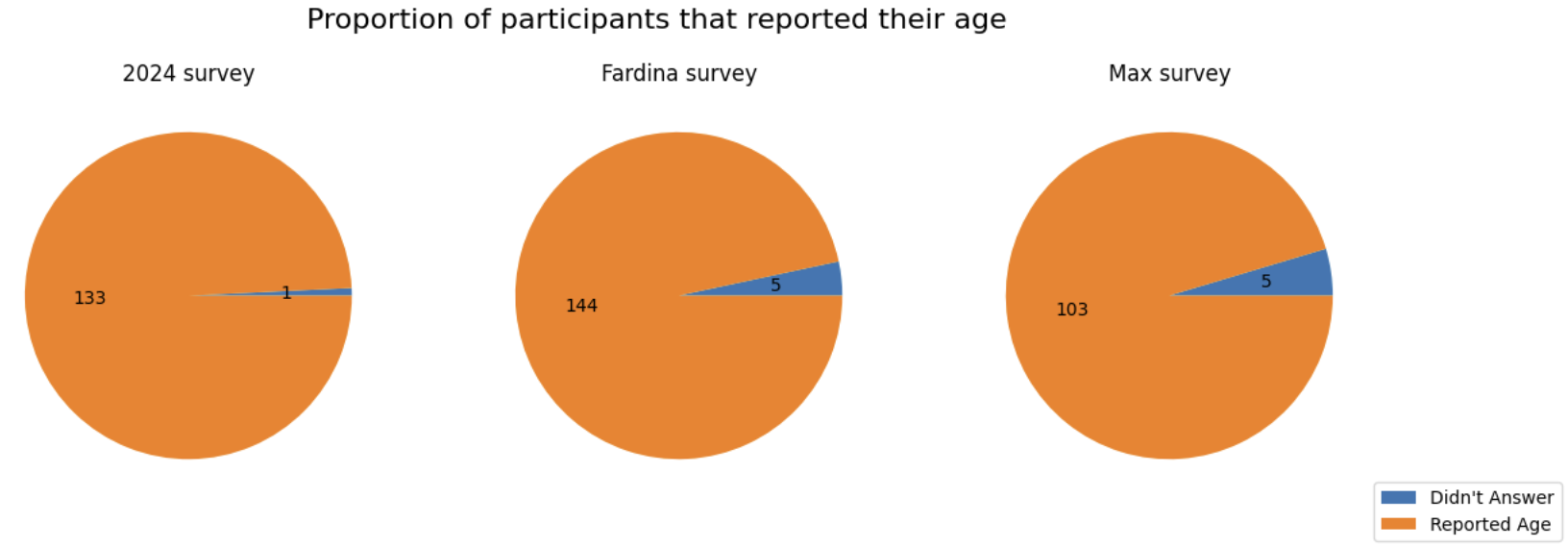


Fig 1. Proportion of participants that reported their age.

As can be seen in the bar plot below, the majority of participants were college aged, between 18 and 22, the vast majority of participants being 20. This is to be expected because most people in CMSC 320 are college-aged and students usually take CMSC 320 during sophomore and junior year, which is usually also when most students turn 20.

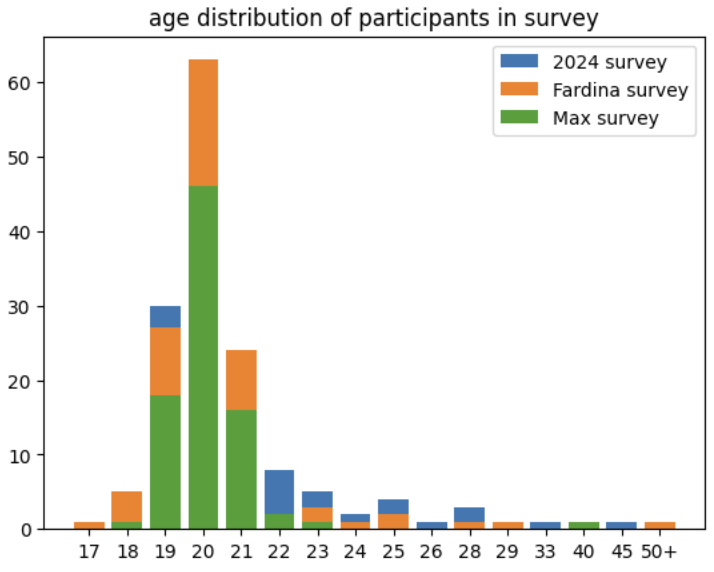


Fig 2. The distribution of ages for each survey respectively. The bulk of students clustered around 20 years of age.

The gender distribution is described in the bottom table. The majority of survey participants identified as men. Overall, the genders of participants across age groups were similarly spread. That is, there was no large imbalance where an age group was predominantly one gender.

|  | 2024 survey | Fardina Survey | Max Survey |
| --- | --- | --- | --- |
| Male | 96 | 98 | 59 |
| Female | 27 | 32 | 19 |
| Other/Prefer not to say | 2 | 1 | 7 |

Table 1. The gender breakdown of each survey respectively

Next, I wanted to examine the political leaning of each class. As can be seen in the pie chart below, the political split of each class was around the same, with most people being one of either mildly liberal, neutral, strongly liberal or mildly conservative.

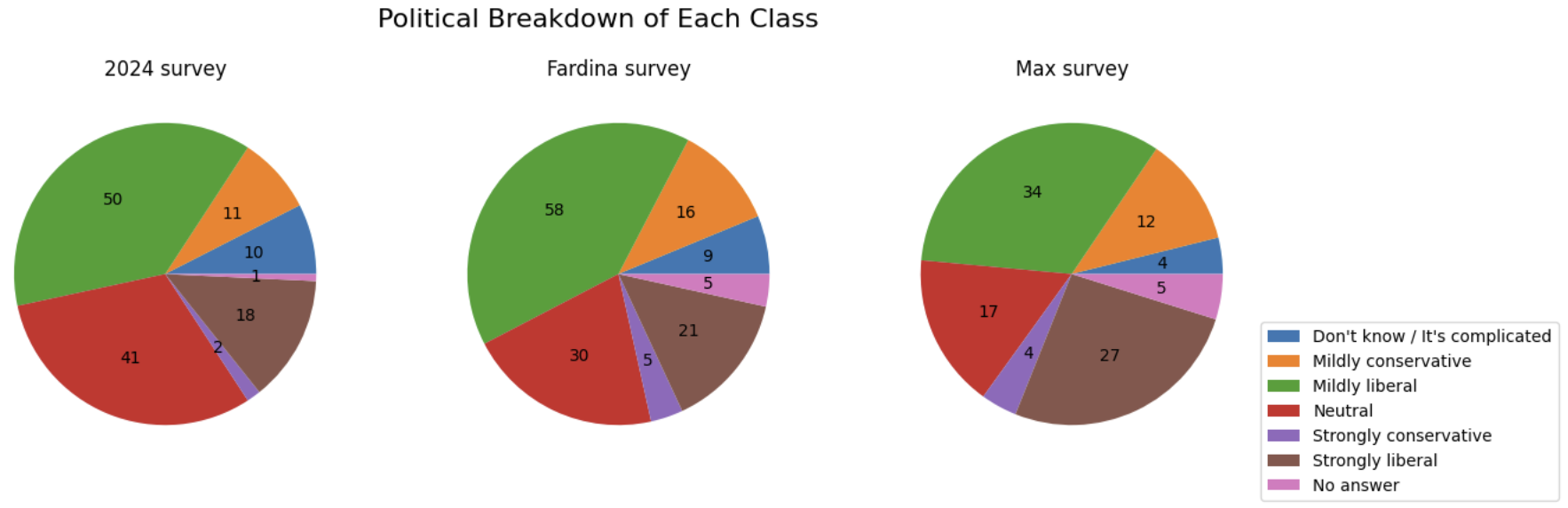


Fig 2. Political breakdown of each class

# Findings

I used a Chi-Squared test to see, for each question, whether a students’ standing and their answers were independent. I binned students as being either a graduate student or an undergraduate student, and looked at their answer to each of the questions. I excluded table entries where the student did not state their standing or the student didn’t answer the question. I designed the test as follows: The null hypothesis was that students’ standing and their answers were independent. The alternative hypothesis was that students’ standing and their answers were not independent, and I would reject the null hypothesis if the chi-squared test yielded a p-value less than 0.06. I ran this test on each of the questions, generating the following table:

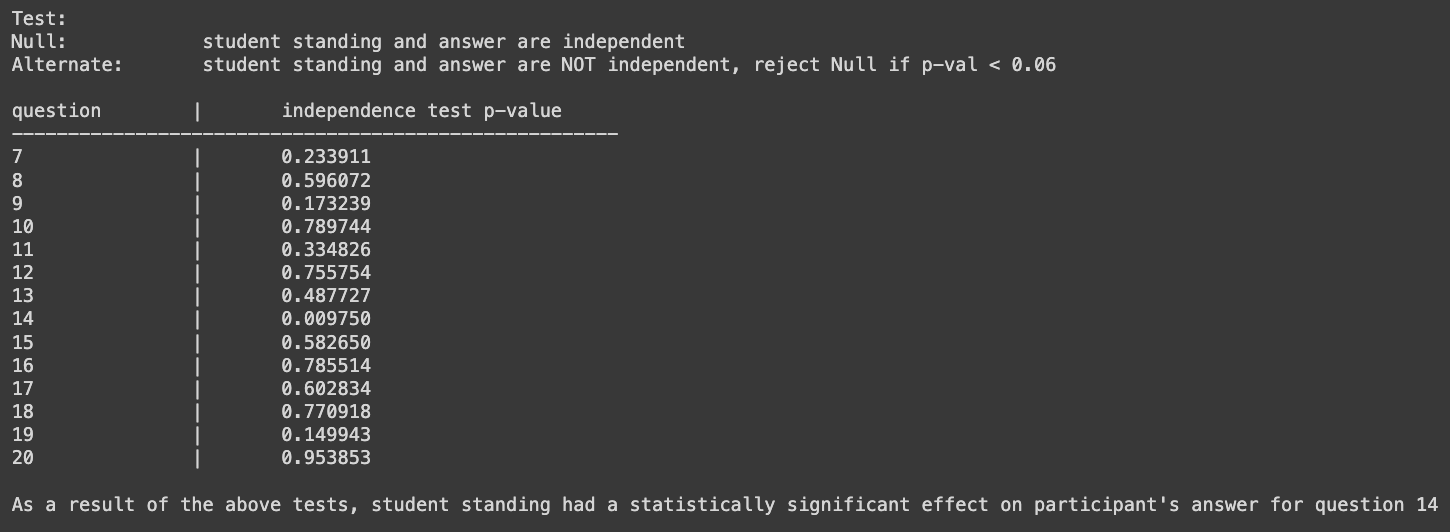


Fig 3. p-values for each of the different questions. As can be seen, question 14 had a p-value of 0.0098

I found that for question 14, which was the scenario about the single mother using ⅔ of her child support on the kid with a different father, student standing had a statistically significant effect.