NAME:PARIPOORNA D

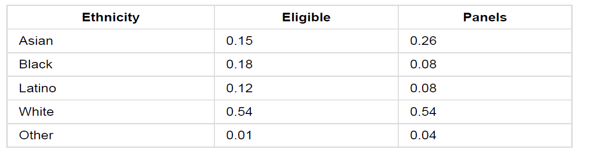
ROLL\_NO:CB.EN.U4CSE20245

ASSIGNMENT -3

The focus of the study by the ACLU of Northern California was the ethnic composition of jury panels in Alameda Country. People who applied for jury service was 1453

The ACLU gathered demographic data on all of these prospective jurors, and compared those data with the composition of all eligible jurors in the country.

The data is shown below



Column details

Eligible--- >proportion of all eligible juror candidates for each ethinicity

Panel---- >proportion of people of that ethinicity appeared for the process of selection into the jury

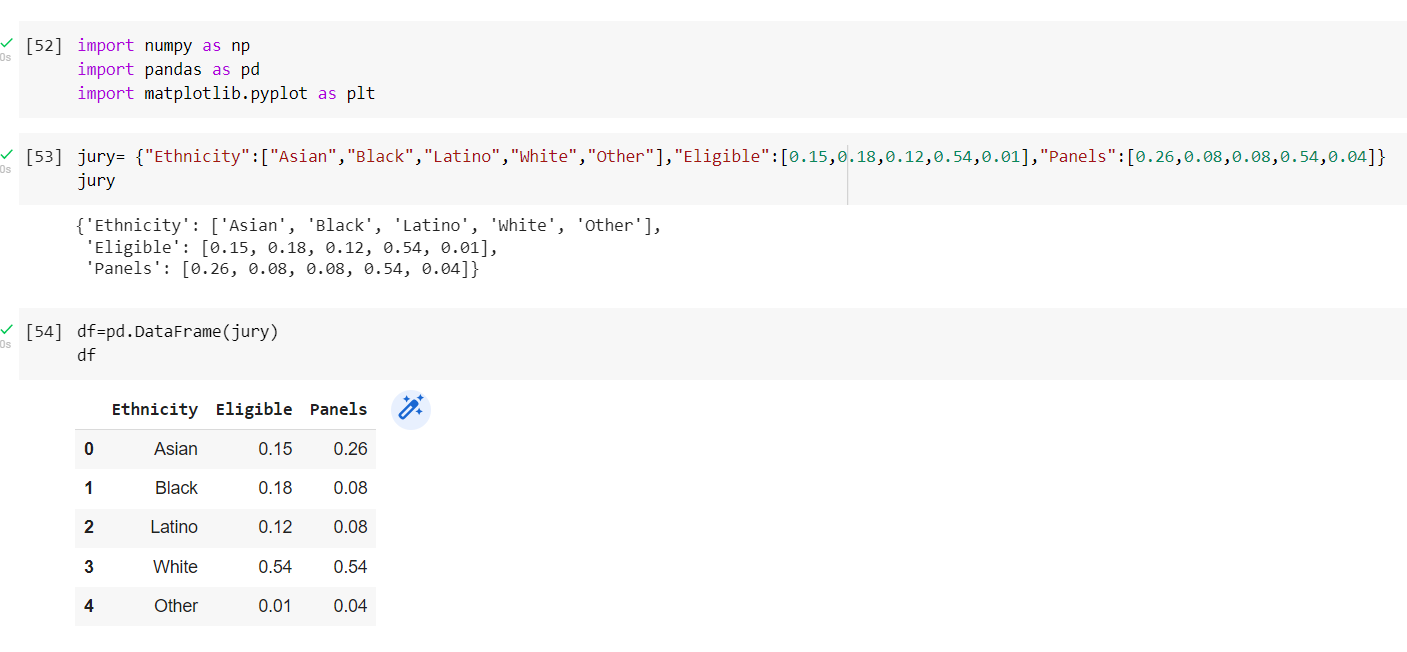
As you see from the table some ethinicities are underrepresented and overrepresented on the panels. But the initial panel should come from the random sample of the population of eligible jurors

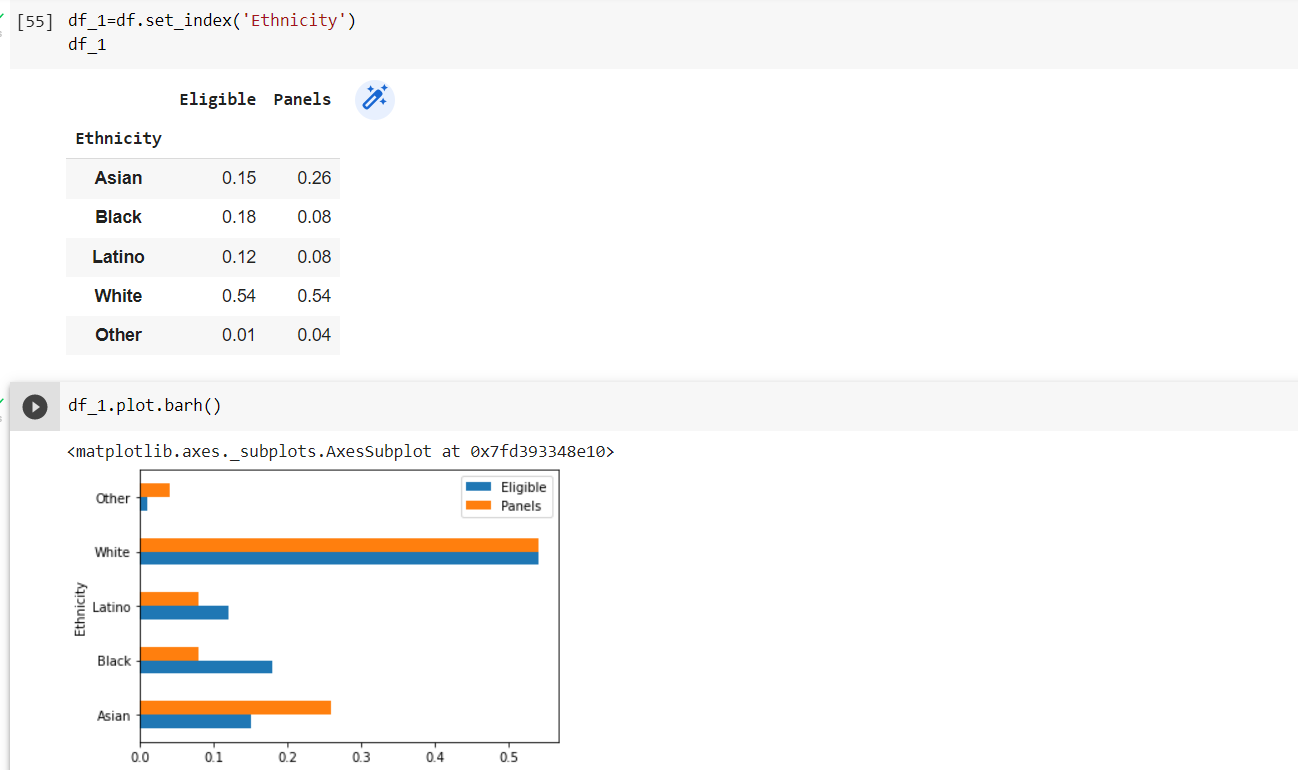
So here our goal is to check whether the panel is actually came from the selection at random

So

**Null Hypothesis**:Panels were the representative of the distribution provided for the eligible jurors

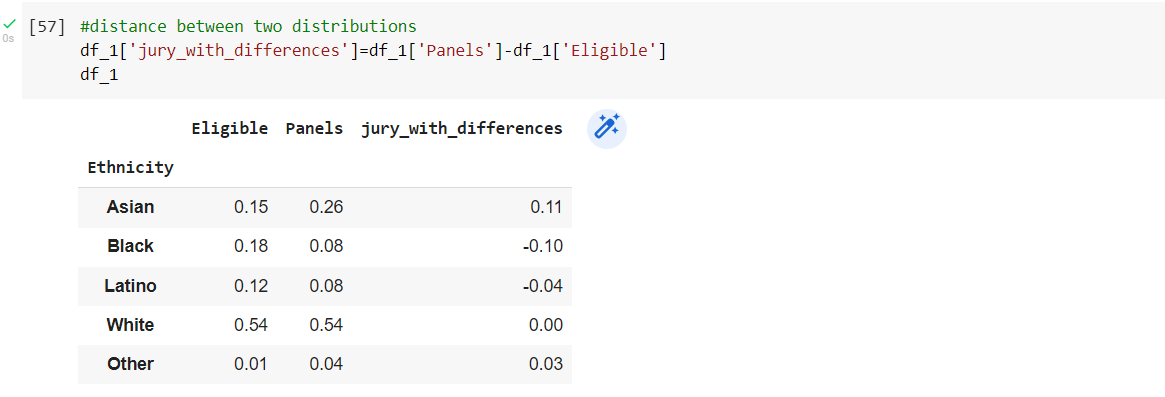
**Alternative Hypothesis**: Panels were not selected at random



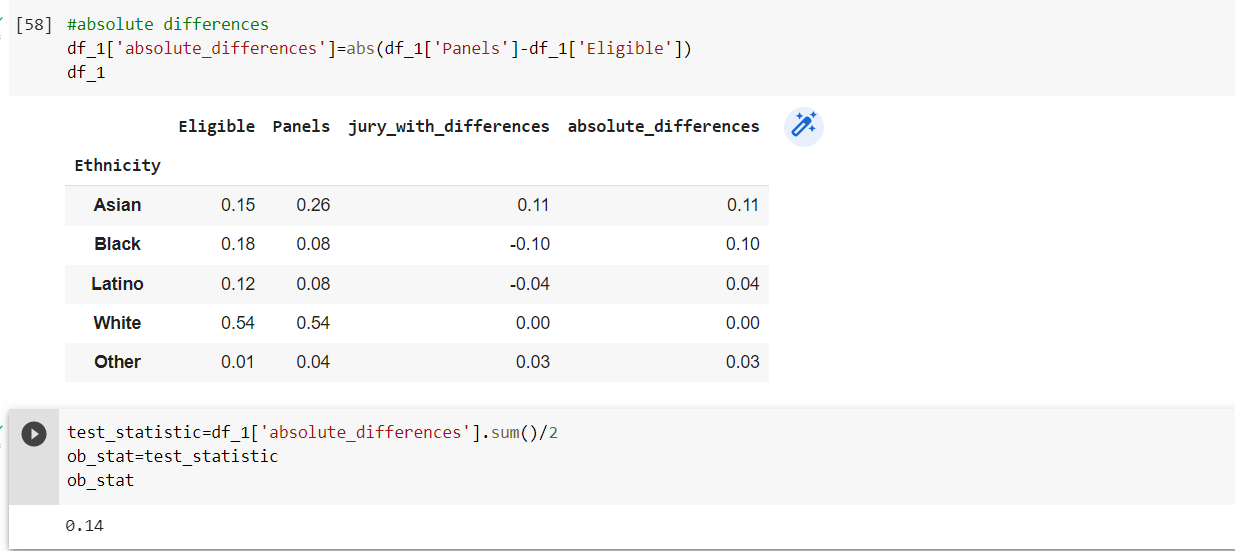


As you see from the above bar graph, for most ethnicities eligible and panel bars are not same

Now we will calculate how much the distance is varied, so will add a new column to findout the differences

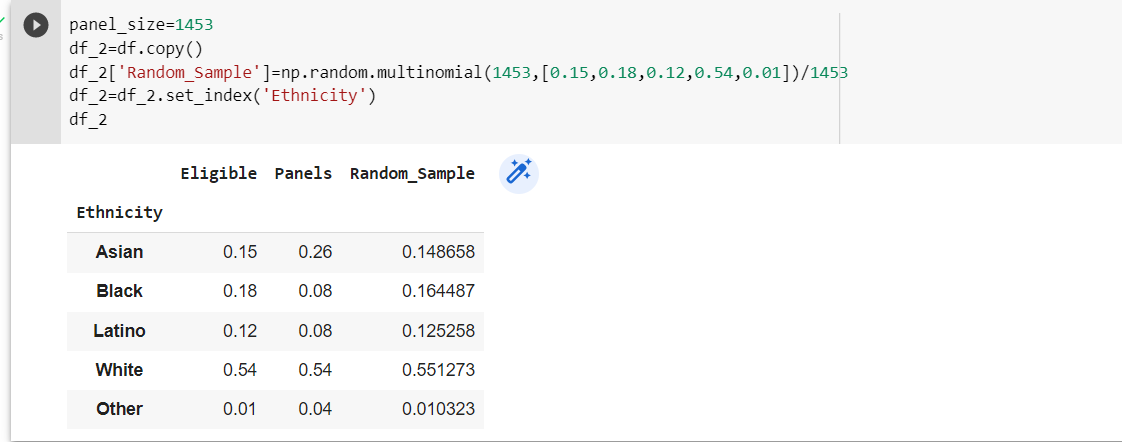


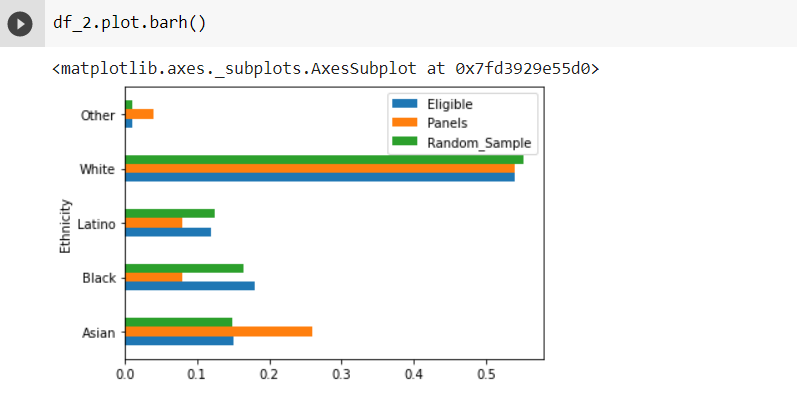
To calculate the total distance varied, we have to sumup the differences but here the sum will be equal to zero(0.11+(-0.10)+(-0.04)+0.00+0.03). To avoid this cancellation we will take the absolute value of each difference and sum it and divide it by 2.



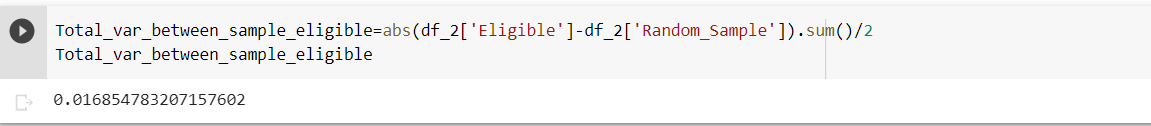
As you see the observed total variation distance is 0.14

Now checking whether the panels are the representative of the eligible population



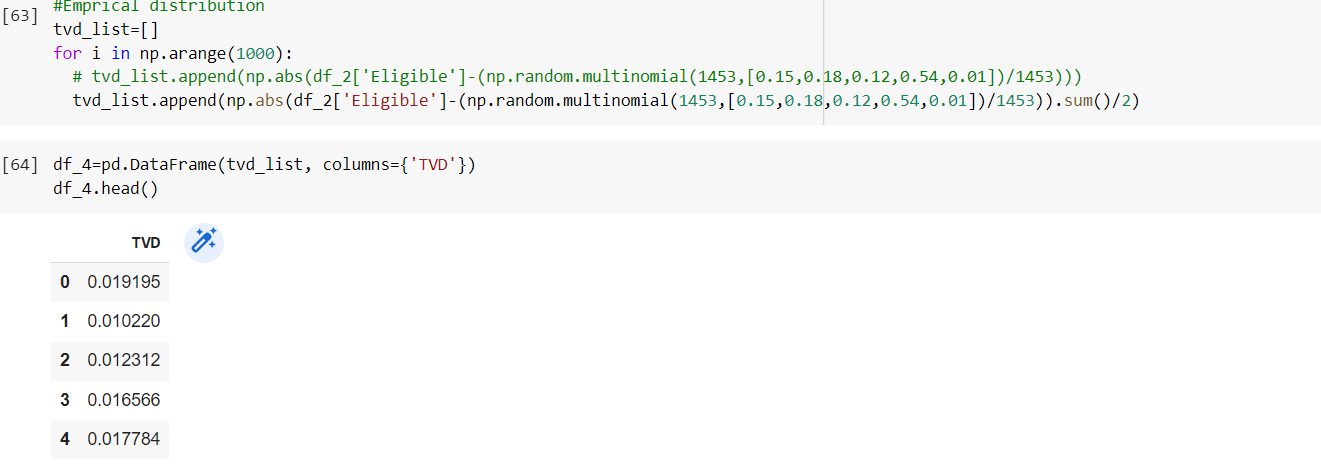


Here the distribution of the random sample is close to the eligible population but not likely to the panel distribution(as the green bars are closer to blue bars than the orange bars)

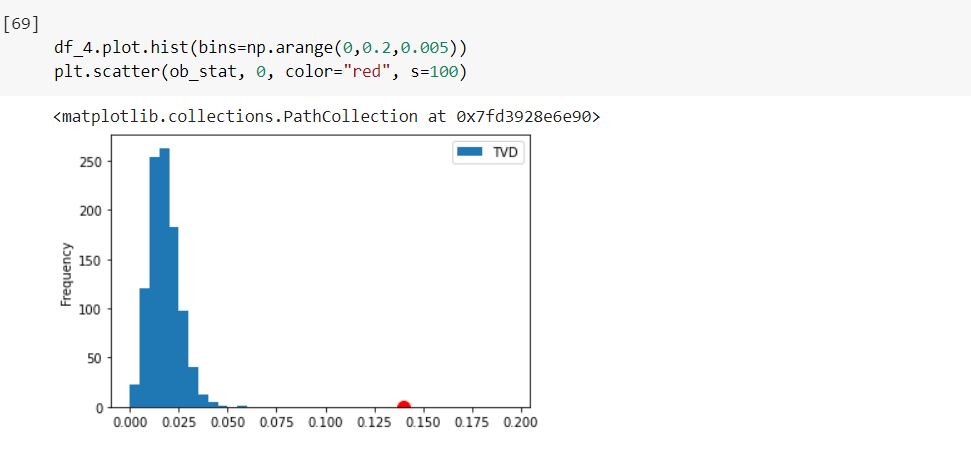


The total distance varied between the eligible population and random sample is 0.0168 but the observed test static is having the value of 0.14 which is greater

Emprical Distribution Of The total distance varied



Each row of df\_4 has the total distance varied between a random sample size of 1453 and the population of eligible jurors



The emprical histogram of the simulated distances (rawing 1453 jurors from the pool of eligible candidates at random)rarely deviates from the eligible jurors just more than 0.05 but the observed stat shows the variation of 0.14 which is far from the distribution(red dot in the figure).

So the datas in the panels are inconsistent with the predicted values of the statistic based on the model of the random selection. Hence the panels were not representative of the distribution provided for the eligible jurors.

Null Hypotheis is rejected