Case Study:

**Jury Selection in Alameda County**

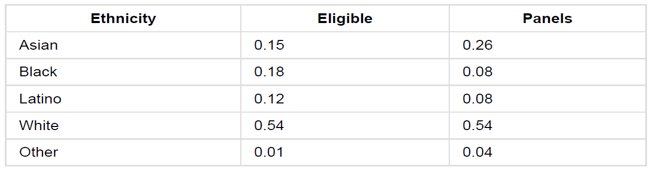
**Problem Statement:**

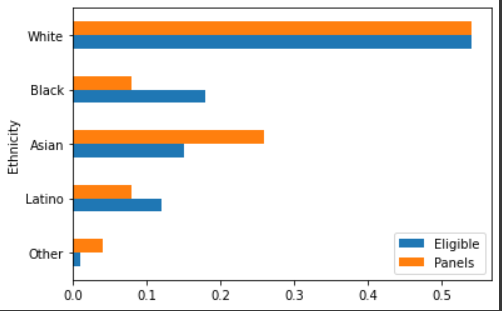
In 2010, the ACLU presented a report on jury selection in Alameda County, California. The report concluded that certain ethnic groups are underrepresented among jury panellists in Alameda County, and suggested some reforms of the process by which eligible jurors are assigned to panels. In those panels, the total number of people who reported for jury service was 1,453.

**The Goal:**

**The distribution of ethnicity among the jury panellists in Alameda County should resemble to some degree – The distribution of eligible jurors based on ethnicity in the total population.**

**Initial jury panel distribution:**



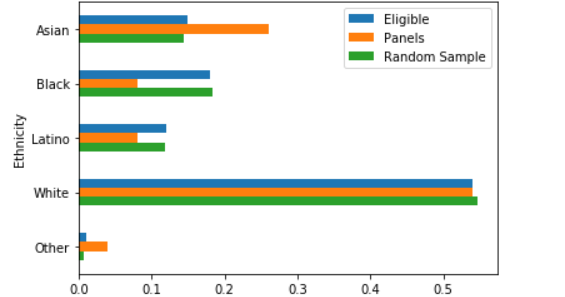


**Hypothesis:**

**Null hypothesis (H0):** The initial panel was randomly selected from the population of eligible jurors.

**Alternative hypothesis (H1):** The panel was not randomly selected.

Let’s pick 1,453 jurors from the eligible list randomly:



* **Here, we see that the random sample is more representative of the Eligible Juror distribution**

**Total Variation Distance Derivation:**  
 Here, we take the “absolute value “of the differences and add them (if we don’t do this, they add to zero.). But, since

**Sum of positive differences = abs(Sum of negative differences) -- (1)**

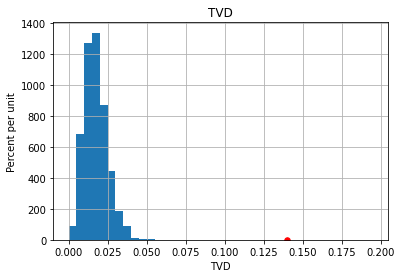
The sum of absolute differences is redundant because:

Sum of absolute difference = Sum of positive difference + abs(Sum of negative differences)

* **Sum of absolute difference = Sum of positive difference + Sum of positive difference** -- derived from (1)
* **Sum of absolute difference = 2 \* Sum of positive difference**

**Total Variation Distance (TVD) =** **Sum of absolute difference / 2**

* **TVD between the random sample and the eligible jurors: 0.0086097.**
* **This is much smaller than the TVD for the panel and eligible population which is 0.14.**

Now we repeat this random sampling process for **5000 times**: 

* The histogram above shows the distribution of the TVDs of 5000 random samples from the eligible juror population.
* In this plot, the maximum TVD doesn’t exceed 0.059. Keep in mind that getting this value (0.059) is already a very rare case.
* But, the initial panel distribution’s TVD (0.14) is shown far out in the right marked by the red dot. **0.14 >> 0.059**

From this, we can deduce that this **makes it highly improbable for the initial panel members’ distribution to have been selected at random** from the eligible population.

Conclusion:

We have significant evidence that the initial panel members’ distribution to have been selected at random from the eligible population. Therefore, we **reject the null hypothesis.**