lec17and18and19

July 18, 2022

[1]:
```
from datascience import *
import numpy as np

%matplotlib inline
import matplotlib.pyplot as plots
plots.style.use('fivethirtyeight')
```

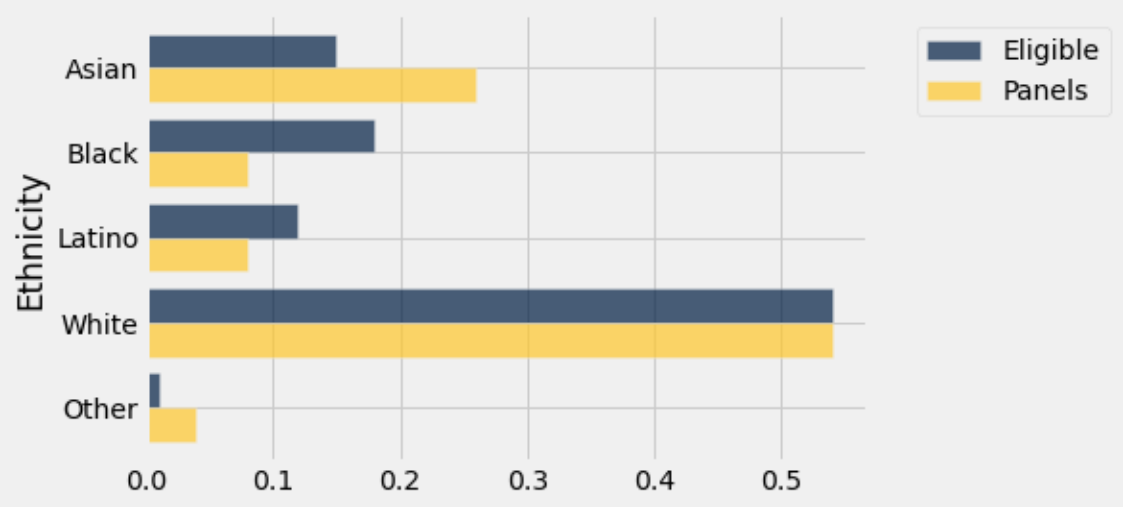
0.1 Alameda County Jury Panels

[2]:
```
jury = Table.read_table('alameda.csv')

jury
```

[2]:
```
Ethnicity | Eligible | Panels
Asian     | 0.15     | 0.26
Black     | 0.18     | 0.08
Latino    | 0.12     | 0.08
White     | 0.54     | 0.54
Other     | 0.01     | 0.04
```

[3]:
```
jury.barh('Ethnicity')
```



[4]:
```
jury_with_diffs = jury.with_column(
    'Difference', jury.column('Panels') - jury.column('Eligible')
)
```

[5]:
```
jury_with_diffs
```

[5]:
```
Ethnicity | Eligible | Panels | Difference
Asian     | 0.15     | 0.26   | 0.11
Black     | 0.18     | 0.08   | -0.1
Latino    | 0.12     | 0.08   | -0.04
White     | 0.54     | 0.54   | 0
Other     | 0.01     | 0.04   | 0.03
```

[6]:
```
jury_with_diffs = jury_with_diffs.with_column(
    'Absolute Difference', np.abs(jury_with_diffs.column('Difference'))
)
```

[7]:
```
jury_with_diffs
```

[7]:
```
Ethnicity | Eligible | Panels | Difference | Absolute Difference
Asian     | 0.15     | 0.26   | 0.11       | 0.11
Black     | 0.18     | 0.08   | -0.1       | 0.1
Latino    | 0.12     | 0.08   | -0.04      | 0.04
White     | 0.54     | 0.54   | 0          | 0
Other     | 0.01     | 0.04   | 0.03       | 0.03
```

[8]:
```
sum(jury_with_diffs.column('Absolute Difference'))
```

[8]:
```
0.28
```

[9]:
```
sum(jury_with_diffs.column('Absolute Difference')) / 2
```

[9]:
```
0.14
```

[10]:
```
def total_variation_distance(distribution_1, distribution_2):
    return sum(np.abs(distribution_1 - distribution_2)) / 2
```

[11]:
```
total_variation_distance(jury.column('Eligible'), jury.column('Panels'))
```

[11]:
```
0.14
```

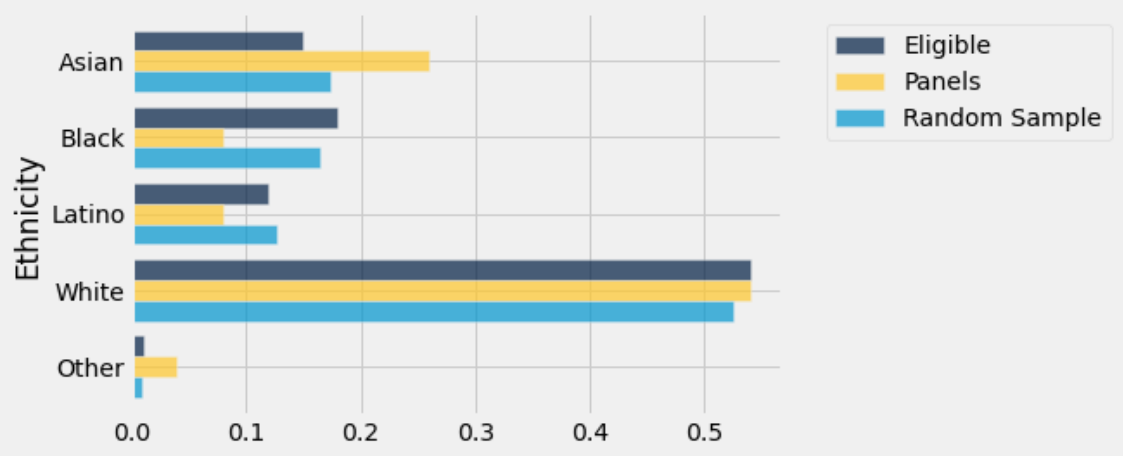
[12]:
```
eligible = jury.column('Eligible')
```

[13]:
```
sample_distribution = sample_proportions(1453, eligible)
panels_and_sample = jury.with_column('Random Sample', sample_distribution)
```

[14]:
```
panels_and_sample
```

[14]:
```
Ethnicity | Eligible | Panels | Random Sample
Asian     | 0.15     | 0.26   | 0.174123
Black     | 0.18     | 0.08   | 0.164487
Latino    | 0.12     | 0.08   | 0.126635
White     | 0.54     | 0.54   | 0.52512
Other     | 0.01     | 0.04   | 0.00963524
```

[15]:
```
panels_and_sample.barh('Ethnicity')
```



[16]:
```
total_variation_distance(panels_and_sample.column('Random Sample'), eligible)
```

[16]:
```
0.03075705437026844
```

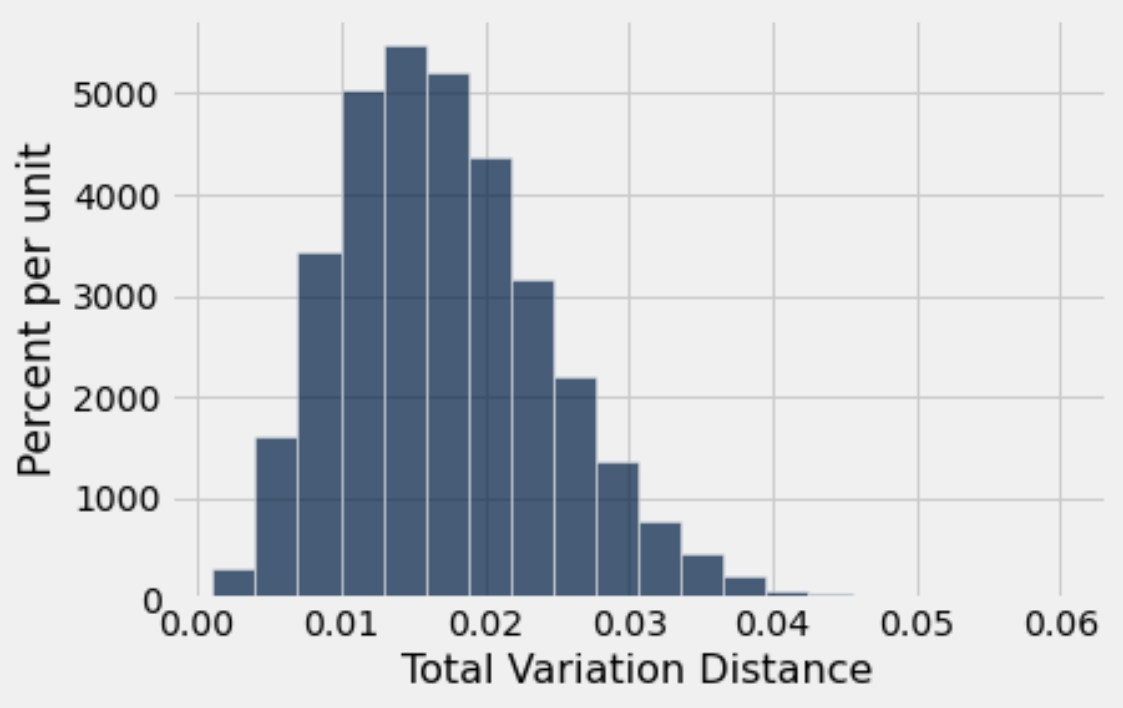
[17]:
```
sample_distribution = sample_proportions(1453, eligible)
total_variation_distance(sample_distribution, eligible)
```

[17]:
```
0.006262904335856854
```

[18]:
```
tvds = make_array()

for i in np.arange(10000):
    sample_distribution = sample_proportions(1453, eligible)
    new_tvd = total_variation_distance(sample_distribution, eligible)
    tvds = np.append(tvds, new_tvd)
```

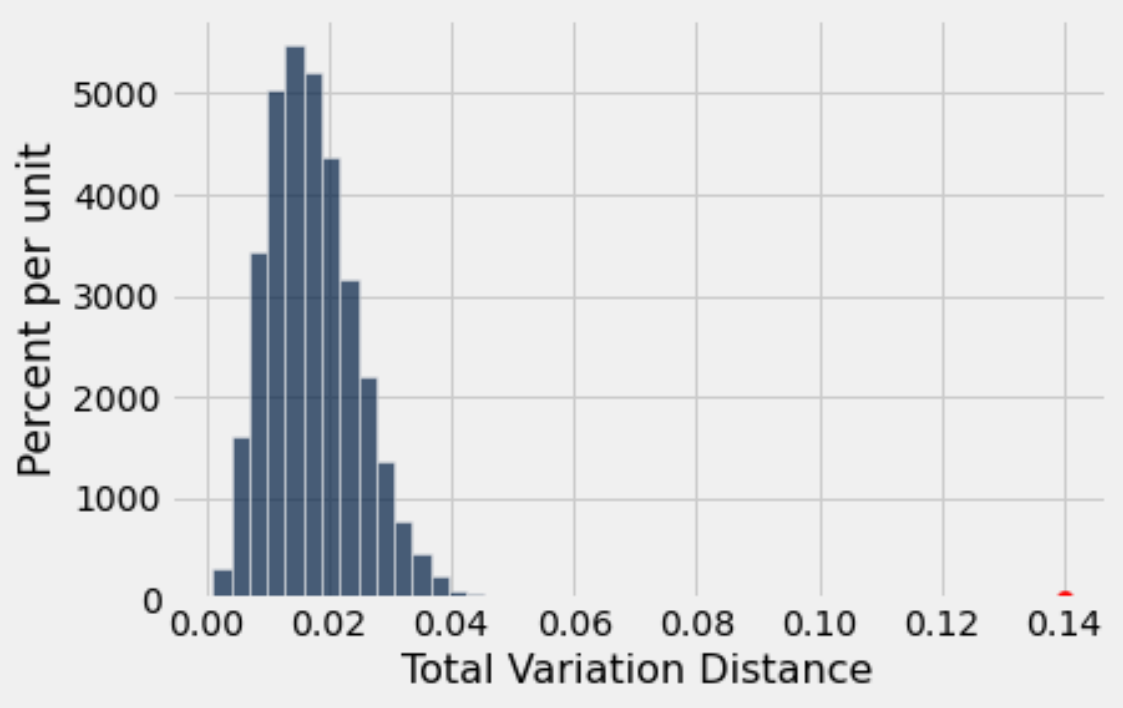
[19]:
```
Table().with_column('Total Variation Distance', tvds).hist(bins = 20)
```



[20]:
```
observed_tvd = total_variation_distance(jury.column('Panels'), eligible)
observed_tvd
```

[20]: 0.14

[21]:
```
Table().with_column('Total Variation Distance', tvds).hist(bins = 20)
plots.scatter(observed_tvd, 0, color = 'red', s=40);
```


5000
4000
3000
2000
1000
0
Percent per unit
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Total Variation Distance

[ ]:

[ ]: